Univerzitet Crne Gore
Prirodno-matematički fakultet

Kosta Pavlović

UMETANJE VODENIH ŽIGOVA U DIGITALNE AUDIO SIGNALE KORIŠĆENJEM DUBOKIH NEURONSKIH MREŽA

– DOKTORSKA DISERTACIJA –

Podgorica, 2024.

University of Montenegro
Faculty of Natural Sciences and Mathematics

Kosta Pavlović

DIGITAL AUDIO WATERMARKING USING DEEP NEURAL NETWORKS

– PHD THESIS –

Podgorica, 2024

PODACI I INFORMACIJE O DOKTORANDU

Ime i prezime: Kosta Pavlović
Datum i mjesto rođenja: 8. maj 1994.
Naziv završenog postdiplomskog studijskog programa i godina završetka studija: Računarske nauke, 2018. godine

PODACI I INFORMACIJE O MENTORU

Ime i prezime: Prof. dr Igor Đurović
Titula: doktor nauka elektrotehnike
Zvanje: redovni profesor
Naziv univerziteta i organizacione jedinice : Univerzitet Crne Gore, Elektrotehnički Fakultet

Komisija za ocjenu:

1. Prof. dr Milenko Mosurović, redovni profesor na PMF-u Univerziteta Crne Gore
2. Prof. dr Igor Đurović, redovni profesor na ETF-u Univerziteta Crne Gore
3. Prof. dr Goran Kvaščev, vanredni profesor na ETF-u Univerziteta u Beogradu
4. Prof. dr Vesna Popović-Bugarin, redovni profesor na ETF-u Univerziteta Crne Gore
5. Doc. dr Igor Jovančević, docent na PMF-u Univerziteta Crne Gore

Komisija za odbranu:

1. Prof. dr Milenko Mosurović, redovni profesor na PMF-u Univerziteta Crne Gore
2. Prof. dr Igor Đurović, redovni profesor na ETF-u Univerziteta Crne Gore
3. Prof. dr Goran Kvaščev, vanredni profesor na ETF-u Univerziteta u Beogradu
4. Prof. dr Vesna Popović-Bugarin, redovni profesor na ETF-u Univerziteta Crne Gore
5. Doc. dr Igor Jovančević, docent na PMF-u Univerziteta Crne Gore

Datum odbrane: 6. jun 2024.

PODACI O DOKTORSKOJ DISERTACIJI

Naziv doktorskih studija: Računarske nauke, Prirodno-matematički fakultet, Univerzitet Crne Gore

Naslov disertacije: Umetanje vodenih žigova u digitalne audio signale korišćenjem dubokih neuronskih mreža

Rezime: U ovoj disertaciji predložen je novi obrazac za kreiranje sistema vodenog žiga za digitalne audio signale, koji se zasniva na dubokom učenju. Rezultati istraživanja ukazuju da sistemi implementirani prema ovom obrascu ostvaruju visoku otpornost na različite audio efekte, bez značajnog narušavanja kvaliteta signala prilikom dodavanja vodenog žiga.

Ključne riječi: sistemi vodenog žiga; zaštita autorskih prava i autentičnosti; duboko učenje; konvolucione neuronske mreže; audio efekti

Naučna oblast: Vještačka inteligencija, Obrada signala

Uža naučna oblast: Duboko učenje, Sistemi vodenog žiga

UDK: 004.89

INFORMATION ON THE PHD THESIS

Name of the doctoral program: Computer Science, Faculty of Natural Sciences and Mathematics, University of Montenegro

Thesis title: Digital audio watermarking using deep neural networks

Summary: This dissertation proposes a novel framework for creating audio watermarking systems based on deep learning. Research results indicate that systems implemented according to this framework achieve high resilience to various audio effects, without significantly compromising signal quality during watermark embedding.

Keywords: watermarking systems; copyright and authenticity protection; deep learning; convolutional neural networks; audio effects **Scientific field:** Artificial intelligence, Signal processing

Scientific subfield: Deep learning, Digital watermarking

UDC: 004.89

Predgovor

Digitalni podaci predstavljaju jedan od ključnih proizvoda stvaralaštva i izražavanja u modernom svijetu. Skladištenje i distribuiranje digitalnog sadržaja dominantno se odvija putem Interneta. Time se povećavaju mogućnosti za zloupotrebama, bilo neovlaštenim korišćenjem ili distribucijom, bilo stvaranjem plagijata ili drugim malicioznim radnjama. Ovaj problem postaje još alarmantniji saznanjem da je većina proizvedenog sadržaja privatnog karaktera. Postavljaju se ključna pitanja: kako zaštititi svoje digitalne podatke, kako dokazati vlasništvo nad digitalnim podacima, itd.

Nažalost, čini se da još uvijek ne postoji efikasno rješenje kako bi se pomenutim zloupotrebama ozbiljno suprotstavilo. Ogromna nadolazeća potreba za rješavanjem ovih problema, kao i očigledan prostor za poboljšanje trenutnih rješenja su me inspirisali da dublje istražim ovu temu i pokušam dati svoj doprinos. Vjerujem da je ova tema od izuzetnog značaja i da će vremenom još više dobijati na važnosti.

Digitalni votermarking je tehnologija koja je godinama razvijana za rješavanje ovih problema modernog društva. Prepoznao sam priliku da se u daljem razvoju votermarking tehnologije upotrijebe tehnike mašinskog i dubokog učenja. Ove metode doživjele su ubrzan napredak posljednjih godina, ali još uvijek nisu primjenjivane u vodećim votermarking tehnikama, otvarajući prostor za istraživanja i inovacije u tom pogledu. Dodatno, oblast mašinskog učenja je predmet mog najvećeg profesionalnog interesovanja još od studentskih dana. Bilo je posebno zadovoljstvo primijeniti ga za rješavanje jednog ovako važnog problema.

U pisanju ove teze trudio sam se da prikažem odabranu temu ne samo s tehničkog aspekta, već i da istražim i prezentujem njenu širu društvenu i globalnu važnost, pogotovo u uvodnim poglavljima. Tekst je napisan s ciljem da bude sveobuhvatan i samodovoljan, odnosno da pruži dovoljno informacija i konteksta, bez potrebe za obiljem dodatne literature. Ukoliko čitaoci ovu disertaciju budu smatrali korisnom i razumljivom, smatraću da je postignut glavni cilj njenog pisanja. Iskreno se nadam da će ovaj tekst poslužiti kao temeljna literatura nekome ko želi upoznati ili istraživati oblast votermarkinga digitalnih audio signala i uopšte zaštitu digitalnog sadržaja. Smatram da bi to bio najljepši način da se opravda trud uložen u ovo istraživanje i pisanje ovog teksta.

Zahvalnosti

Najveću zahvalnost za izradu ove disertacije dugujem kolegi Slavku Kovačeviću, čije konstruktivne kritike, savjeti i komentari su bili ključni za dublje razumijevanje izazova koji su se pojavljivali tokom istraživanja, te su na kraju dali neizmjeran doprinos njegovom uspjehu.

Duboku zahvalnost izražavam svom mentoru, prof. dr Igoru Đuroviću. Njegovo iskustvo i sposobnost prepoznavanja ključnih aspekata istraživanja, kao i usmjeravanja u pogledu njegovog sprovođenja bili su od neosporivog značaja u svim fazama izrade ove disertacije. Zahvaljujući njegovom mentorstvu, stekao sam dragocjeno znanje koje će oblikovati moju buduću karijeru.

Želim iskreno zahvaliti i profesoru Milenku Mosuroviću na njegovoj posvećenosti u čitanju moje disertacije i pažljivom ispravljanju grešaka. Hvala mu na vremenu i trudu koji je uložio kako bi doprinio kvalitetu ovog teksta.

Zahvaljujem Skupštini Crne Gore na dragocjenoj pomoći obezbjeđivanjem reprezentativnog korpusa podataka koji je korišten u okviru istraživanja i profesoru Savu Tomoviću koji je pomogao da se ova saradnja ostvari. Ovi podaci činili su osnovu svih eksperimenata i bili su od suštinskog značaja za uspjeh ovog rada.

Takođe, želim izraziti zahvalnost Inovaciono preduzetničkom centru Tehnopolis na bezuslovnom ustupanju servera s moćnom grafičkom karticom. Ova tehnološka podrška omogućila je sprovođenje eksperimenata, obučavanje i optimizaciju modela dubokog učenja na način koji bi inače bio nedostižan.

Hvala na podršci, razumijevanju i strpljenju vjerenici Ani, roditeljima, sestri, prijateljima. Pisanje disertacije trajalo je prilično duže nego što smo svi očekivali. Nadam se da će kvalitet teksta opravdati utrošeno vrijeme i sva odricanja. Vaša podrška bila je neophodna. Hvala vam što ste uvijek bili uz mene.

Sažetak

Ova disertacija predstavlja novi pristup u oblasti sistema vodenog žiga za digitalne audio signale. Predloženi sistem se ističe kao prvi koji sve ključne operacije umetanja i detekcije vodenih žigova obavlja dubokim neuronskim mrežama. Dizajnirana je procedura zajedničkog obučavanja predloženog skupa neuronskih mreža s ciljem postizanja visoke otpornosti na različite napade, uključujući desinhronizaciju, uz očuvanje kvaliteta signala. Ovom procedurom obučena su dva modela sistema vodenog žiga čije su performanse ispitane na skupu govornih signala. Rezultati pokazuju visoku efikasnost oba modela i neprimjetnost umetnutih vodenih žigova. Sistem ostvaruje gotovo savršenu detekciju pri dejstvu audio efekata. Većina efekata izaziva najviše oko 3% pogrešno detektovanih bitova vodenog žiga. Vrijednosti mjera PESQ i SNR za kvalitet votermarkovanog signala iznad 3.6 i 23 dB, respektivno, ukazuju da skoro da nema čujnih promjena u signalu nakon dodavanja vodenog žiga. Sveobuhvatna analiza performansi sistema pokazuje da je predloženi sistem po većini kriterijuma performansi na nivou najboljih tehnika iz literature, a da ih u određenim aspektima prevazilazi.

Abstract

This dissertation introduces a novel approach in the field of audio watermarking. The proposed system stands out as the first to perform all key watermark embedding and detection operations with deep neural networks. A joint training procedure for the proposed set of neural networks has been designed to achieve high resistance to various attacks, including desynchronization, while preserving signal quality. Two models of the watermarking system were trained using this procedure and their performance was evaluated on a dataset of speech signals. The results demonstrate the high efficiency of both models and the imperceptibility of the embedded watermarks. The system achieves almost perfect detection under typical audio effects. The majority of effects cause at most around 3% incorrectly detected watermark bits. PESQ and SNR values for the quality of the watermarked signal above 3.6 and 23 dB, respectively, indicate that there are almost no perceptible changes in the signal after embedding the watermark. A comprehensive performance analysis of the system shows that, in most performance criteria, the proposed system is comparable to the best techniques from the literature and surpasses them in certain aspects.

Sadržaj

1 Uvod

Digitalizacija savremenih načina komunikacije donijela je poboljšanja u obradi multimedijalnih podataka, njihovom skladištenju i prikupljanju i učinila ih je znatno pristupačnijim kranjim korisnicima. Uporedo sa povećanjem dostupnosti multimedijalnih podataka napredovali su i softveri za njihovu obradu i postajali su sve dostupniji i lakši za korišćenje. Međutim, povećao se i broj zloupotreba ovih softvera, najviše u pogledu povreda autorskih prava.

Intelektualnom svojinom smatra se bilo kakva duhovna tvorevina u nauci ili umjetnosti, bilo u pisanom, govornom ili nekom drugom obliku. Zbog njene nematerijalne prirode, intelektualnu svojinu je značajno teže zaštititi od materijalne svojine. Autori na raspolaganju nemaju mnogo sredstava kojima bi spriječili druge da eksploatišu njihovu intelektualnu tvorevinu. Industrija zabave, ali i mnoge druge industrije koje produkuju multimedijalne, odnosno nematerijalne sadržaje gube velike svote novca zbog povreda autorskih prava. Korisnicima interneta postalo je izuzetno jednostavno napraviti veoma kvalitetne kopije tuđeg materijala i nelegalno ih distribuirati, ponekad besplatno, a ponekad i za sopstvenu korist. Veoma brzo postalo je jasno da će neovlašćeno dijeljenje i konzumiranje digitalnog sadržaja uzeti maha. Jedan od prvih poznatih sudskih sporova koji je istakao problem ilegalne distribucije digitalnog sadržaja desio se kada je američki bend Metalika podigao tužbu protiv Napstera zbog povrede autorskih prava. Napster je bila platforma za dijeljenje audio sadržaja u MP3 formatu, koja je omogućavala korisnicima da besplatno razmjenjuju muzičke datoteke.

Internet piraterija podataka čini se kao nezaustavljiv trend, a u javnosti još uvijek ne postoji jasan mehanizam kojim bi se ona zaustavila. Iz ovih razloga, zaštita intelektualne svojine, pogotovo u savremenom, digitalnom svijetu, sve više dobija na značaju. Stoga je veoma važno da se savremene tehnologije stave u službu zaštite autorskih prava i da zajedno sa odgovarajućim zakonima zaštite intelektualna dobra stvorena od strane pojedinaca ili grupa i da im se omogući da uživaju u plodovima svog rada.

1.1 Metode zaštite multimedijalnih podataka

Zaštita multimedijalnih podataka se u nekoj mjeri može ostvariti primjenom kriptografskih tehnika. Sadržaju šifrovanih podataka se ne može pristupiti bez posjedovanja ključa za dešifrovanje. Ipak, kriptografija ima ograničena kada je u pitanju zaštita intelektualne svojine. Ovakvi sistemi podložni su takozvanim napadima

„čovjeka u sredini" (*engl. man-in-the-middle attack*). Zlonamjerni korisnik može presresti komunikaciju između dvije stranke i na taj način pribaviti podatke i ključ za dešifrovanje. Osim toga, zlonamjerni korisnik može, kao konzument, kupiti multimedijalne podatke i tako dobiti pristup njima na sasvim legitiman način. Nakon što su podaci dešifrovani, u domenu kriptografije ne postoji način na koji bi se spriječila njihova modifikacija, neovlašćena distribucija, i druga zloupotreba. Stoga je očigledna potreba za primjenom drugih tehnika kako bi se zaustavili ovi trendovi. Umetanje vodenog žiga je proces kojim se digitalni signali označavaju nizom bitova. Taj niz bitova naziva se vodeni žig (*engl. watermark*). Na ovaj način se informacija o vlasništvu umeće u sami sadržaj i mnogo ju je teže ukloniti ili falsifikovati.

Vodeni žig predstavlja identifikacioni kod kojim autor označava digitalne podatke ili potvrđuje da su oni njegova svojina. On može biti nasumično izgenerisan, ali i skrojen po želji, u vidu slike, poput logoa autora, ili proizvoljno odabranog teksta, a njegovo prisustvo u signalu može biti obznanjeno ili tajno. Digitalni signali u koje se ugrađuje vodeni žig takođe mogu biti različitog tipa. Ugrađivanje vodenih žigova se može vršiti nad slikom i videom, audio i govornim signalima, slobodnim tekstom, programskim kodom, različitim vrstama dokumenata, itd. Svi ovi signali se jednim imenom nazivaju signali nosioci.

Vodeni žig se u digitalne signale ugrađuje tako da ga je, kada je to potrebno, moguće rekonstruisati kako bi se potvrdila autentičnost signala. Rekonstrukcija mora biti moguća čak i kada je signal izmijenjen, bilo namjernim, bilo nenamjernim manipulacijama. Osim toga, algoritam za umetanje vodenog žiga u digitalni signal mora biti osmišljen tako da se time ne ugrozi informacija koju signal nosi. Poželjno je čak da vodeni žig bude u potpunosti neprimjetan. Kako su ova dva zadatka međusobno suprotstavljena, prilikom kreiranja algoritma potrebno je napraviti kompromis između njih. U nastavku disertacije će se, zbog jednostavnosti i razumljivosti, tuđica votermarking koristiti kao sinonim za procese umetanja i detekcije vodenih žigova.

Kriptografske i votermarking tehnike ne moraju nužno biti suprotstavljene, već se mogu primjenjivati uporedo kako bi se povećala sigurnost cjelokupnog sistema. Vodeni žig se može šifrovati prije umetanja. Na drugoj strani, nakon rekonstrukcije šifrovanog žiga, može se izvršiti njegovo dešifrovanje kako bi se dobio originalni vodeni žig. Takođe, prilikom kreiranja sistema vodenog žiga poželjno je pridržavati se Kerkhofsovog principa za kriptografske sisteme. Prema ovom principu, kriptografski sistem treba da bude dizajniran tako da njegova sigurnost ne počiva na tajnosti algoritama šifrovanja i dešifrovanja. Ukoliko se i za votermarking sisteme primijeni isto pravilo, algoritmi umetanja i detekcije vodenih žigova trebali bi biti javno dostupni. Na ovaj način bi se njihova uspješnost mogla podrobnije ispitati i mogli bi

se otkriti svi nedostaci prije uvođenja sistema u upotrebu. Ipak, često je neophodno da se neke komponente votermarking sistema drže tajnim kako se ne bi mogle zloupotrebljavati.

Informacija o vlasništvu može se smjestiti i u zaglavljima fajlova sa digitalnim signalom. Može se umetnuti kao bar-kod ili QR kod, ako su signali nosioci slike, ili kao tekstualna zabilješka vlasništva (npr. „© ime vlasnika") na slikama ili dokumentima. Međutim, označavanje vlasništva pomoću vodenih žigova ima nekoliko očiglednih prednosti u odnosu na ove metode. Bar i QR kodovi, kao i tekstualne zabilješke, su očigledni i mogu se ukloniti jednostavnim operacijama. Lako ih je falsifikovati. Pored toga, prekrivaju jedan dio signala, narušavajući time i kvalitet i informaciju koju taj signal nosi. Sa druge strane, vodeni žigovi su skriveni u digitalnom signalu i zbog toga ih je teže ukloniti. Vodeni žigovi neće biti uklonjeni ni prilikom kopiranja ili konverzija fajlova u druge formate, što bi se moglo desiti sa zaglavljima fajlova. Nedostatak votermarkinga u odnosu na druge tehnike je u nešto složenijoj proceduri dodavanja ovih informacija o vlasništvu u sami signal. Takođe, dodavanjem informacija o vlasništvu u zaglavlje fajla sami signal ostaje nepromijenjen, dok vodeni žigovi, iako minimalno, ipak degradiraju kvalitet signala. Tekstualna zabilježba o autorskim pravima je ušla u mnoge zakone širom svijeta kao glavno sredstvo za oglašavanje autorskih prava pa je, zbog svog pravnog značaja, uprkos pomenutim nedostacima, i dalje u široj upotrebi od vodenih žigova.

Kod audio signala potreba za umetanjem vodenih žigova još više dolazi do izražaja. Audio signali ne mogu sadržati bar i QR kodove, niti tekstualne zabilježbe o polaganju autorskih prava. Alternativa je dodavanje posebne audio sekvence na početak ili kraj signala u kojoj bi se konstatovalo ko polaže autorska prava na taj audio snimak. Iz očiglednih razloga, ovaj metod nije u praktičnoj upotrebi. Takvu sekvencu bilo bi lako ukloniti, a i samo njeno postojanje bi negativno uticalo na utisak slušalaca tog audio zapisa.

Digitalni audio signali, odnosno govorni signali, su upravo u fokusu ove disertacije i sve razmatrane i predložene tehnike testirane su nad njima. Ljudski govor je poseban tip audio signala i, još uvijek, predstavlja najvažniji oblik komunikacije. Govorom se prenosi ogromna količina veoma vrijednih informacija. Pravo na slobodu mišljenja i izražavanja je jedno od osnovnih ljudskih prava po članu 19 Univerzalne deklaracije o ljudskim pravima[1] i treba preduzeti sve raspoložive mjere kako bi se očuvao njegov integritet.

Principi i tehnike umetanja vodenih žigova u govorne signale u najvećoj mjeri su identični kao i za druge vrste audio signala. Govor se u pojedinim svojstvima

[1] Rezolucija Generalne skupštine Ujedinjenih nacija A/RES/217

razlikuje od ostalih vrsta audio signala. Pojedine tehnike koriste ove specifičnosti kako bi na bolji način umetale vodene žigove u govorne signale. Govor se odlikuje drugačijom vremenskom i spektralnom strukturom. Specifičan je u pogledu tonaliteta, raspodjele energije, osnovne frekvencije i širine frekvencijskog opsega, ali i načina na koji ga ljudsko uho percipira. Pred votermarking sistemima za govor postavljaju se i nešto drugačiji zahtjevi u odnosu na ostale sisteme. Za govorne signale važna je razumljivost, dok je kvalitet sporedan, pa se, prilikom dizajniranja sistema, veća pažnja može posvetiti sigurnosti i robustnosti vodenih žigova ili povećanju broja bitova koji se ugrađuju u signale. Dodatna prednost govornih signala je u činjenici da ih je značajno lakše i jeftinije generisati i prikupljati nego slike i druge vrste audio signala. Zbog toga je primjena govornih signala u svim oblastima računarskih nauka u stalnom porastu.

1.2 Istorijat vodenih žigova

Oblast votermarkinga je veoma blisko povezana s oblašću steganografije, odnosno skrivenog pisanja. Steganografske tehnike koriste se za skrivenu komunikaciju. Odabrana poruka skriva se u digitalni medij i prenosi se komunikacionim kanalom. Među pomenutim dvijema oblastima postoje i značajne temeljne razlike. Osnovna razlika je da je u steganografiji postojanje poruke tajno. Poruka se skriva tako da niko osim pošiljaoca i primaoca nije svjestan njenog postojanja. Takođe, prilikom umetanja vodenog žiga mora se postarati da se očuva kvalitet signala na najbolji mogući način, dok je kod stenografije najvažnije osigurati tajnost poruke, a kvalitet signala je sporedan. Uprkos pomenutim razlikama, tehnike koje se koriste za skriveno pisanje i umetanje vodenih žigova mogu biti veoma slične, čak i identične. I vodeni žig je potrebno na što bolji način sakriti u signalu nosiocu. Stoga, kako je oblast steganografije znatno starija, neke od steganografskih tehnika preuzete su i prilagođene potrebama votermarking sistema.

Ljudi su tehnike skrivenog pisanja razvijali od davnina. Vjerovatno su i praistorijski ljudi imali potrebu i posjedovali vještine skrivenog pisanja, odnosno steganografije. Prve zabilježene primjere upotrebe steganografije dao je Herodot oko 440. god. pne. u svom djelu „Istorija". U jednom primjeru je Histaeus, vladar grčkog grada Mileta koji je tada bio pod persijskom kontrolom, zbog svoje lojalnosti i uspješnosti u ratovima pozvan da se kao prijatelj i savjetnik pridruži kralju Dariju I u glavnom gradu Persije, Susi. Vlast u Miletu povjerena je Histaeusovom zetu Aristagori. Histaeus nije bio zadovoljan ovim spletom okolnosti i želio da povrati vlast u Miletu. Namjera mu je bila da podstakne pobunu Grka protiv Persijanaca, a da onda bude poslat od strane Darija I da tu pobunu uguši i opet preuzme vlast.

On je poslao poruku svom nasljedniku Aristagori istetoviranu na glavi svog najvjernijeg roba. Nakon tetoviranja, sačekao je da robu poraste kosa kako bi se poruka prikrila i poslao ga je u Milet. Kada je rob došao kod Aristagore, obrijao je svoju glavu i pokazao mu poruku koja je Aristagoru podstakla da pokrene pobunu protiv Persijanaca. Darije I poslao je Histaeusa da uguši pobunu, ali on u tome nije imao uspjeha, jer pobunjeni Milećani nisu željeli njegov povratak na vlast i protjerali su ga na Lezbos. Histaeus se nakon toga pridružio Grcima u borbi protiv Persije i bio je uhvaćen i pogubljen u jednoj od bitaka.

U drugom primjeru navodi se kako je Demaratus, prognani kralj Sparte, upozorio Grke o predstojećem napadu Persije. Demaratus je, nakon abdikacije, dobio mjesto na dvoru Kserksa I. Tu je svjedočio izgradnji najveće flote koju je svijet do tada vidio i koja je spremana za iznenadni napad na Grčku. Demaratus, koji je očigledno i dalje osjećao ljubav prema domovini, odlučio je da pošalje poruku upozorenja. Premda se plašio da bi poruka mogla biti presretnuta, morao ju je sakriti. Poruku je urezao na drvenom kalupu, a zatim je prekrio voskom tako da izgleda kao prazna voštana ploča za pisanje i poslao je u Grčku. Voštana ploča sa porukom nije otkrivena od strane Persijanaca i uspješno je stigla na odredište. Ovo je omogućilo Grcima da se na vrijeme pripreme za napad Persije i odbrane svoju državu od invazije.

Očekivano, steganografske tehnike najširu primjenu našle su u vojnoj komunikaciji, jer tamo postoji najveća potreba za sigurnim prenosom tajnih poruka. Da bi osigurala komunikaciju, vojska je razvila tehniku širenja spektra (*engl. spread spectrum*) koja je zasnovana na širenju uskopojasnog signala na mnogo veći opseg frekvencija. Komunikaciju koja se prenosi na ovaj način mnogo je teže presretnuti. Takođe, širenjem spektra postiže se veća otpornost na zagušivanja komunikacionih kanala, bilo da su ona namjerna ili slučajna, kao i na druge vrste ometanja. Jednu od prvih primjena tehnike širenja spektra u vojnoj komunikaciji osmilili su američka glumica Hedi Lamar i njen prijatelj, kompozitor i pijanista, Džordž Anthil. Tokom Drugog svetskog rata prvi put su u upotrebu uvedena torpeda sa radio sistemom za navođenje. Međutim, ti signali, koji su kontrolisali kretanje, su veoma lako mogli biti ometani i torpedo bi mogao biti skrenut sa kursa. Ideja koju su Lamar i Anthil patentirali bila je da se slanjem različitih djelova signala na različitim frekvencijama može spriječiti praćenje i ometanje radio sistema za navigaciju torpeda. Anthil je predloženi sistem otjelotvorio u obliku samosvirajućeg klavira sa radio signalima. Ova tehnika je, zbog svoje efikasnosti, sigurnosti i pouzdanosti, preuzeta i u digitalnom votermarkingu i daljim unapređivanjem postala je jedna od najistaknutijih tehnika u ovoj oblasti.

Prve tehnike umetanja vodenih žigova počele su se koristiti još u XIII vijeku. Žigovi su ugrađivani na papirima pomoću tankih žičanih obrazaca postavljenih ka-

lupima za proizvodnju papira. Proizvođači su ih koristili za dekoraciju, ali i za sopstvenu promociju. U XVIII vijeku su ove tehnike unaprijeđene i počele su služiti kao mjera protiv falsifikovanja novca i dokumenata. U tom periodu ova tehnika je i dobila svoj naziv. Pretpostavlja se da je razlog bio efekat koji je njena primjena ostavljala na papir, a koji je bio poput vodenih mrlja. Sredinom XX vijeka javljaju se prvi znaci multimedijalnog votermarkinga. Identifikacioni kodovi, zapisani Morzeovom azbukom, ubacivani su u muzičke signale povremenom primjenom zareznog (*engl. notch*) filtra centriranog oko 1kHz. Odsustvo energije na ovoj frekvenciji bilo je naznaka da je filtar primijenjen u određenom intervalu, i u zavisnosti od dužine tog intervala očitavana je tačka (.) ili crtica (-) u identifikacionom kodu. Razvojem računara krajem XX vijeka i njihovom sve većom primjenom, ove tehnike preseljene su u digitalni domen, a osmišljene su i njihove nove primjene.

1.3 Primjene vodenih žigova u digitalnim signalima

Zaštita autorskih prava predstavlja najveću motivaciju u istraživanju i pronalaženju novih ideja za digitalne votermarking sisteme, iako postoje i druge, ništa manje značajne, primjene. Savremeni sistemi vodenog žiga mogu pružiti veliku pomoć u zaštiti autorskih prava. Mogu se kreirati programi koji bi na internetu tražili multimedijalne sadržaje sa vodenim žigom autora i time otkrivali i upozoravali na potencijalne povrede njihovih autorskih prava. Autori digitalnog sadržaja čija su prava ugrožena mogli bi dokazati svoje vlasništvo nad podacima u sudskom postupku detekcijom sopstvenog vodenog žiga u njima. Ovo se ne bi moglo postići korišćenjem bar i QR kodova i tekstualnih zabilješki jer je njih lako falsifikovati. Ovakav sistem mogao bi poslužiti i građanima koji poštuju zakon da pomoću njega pronađu vlasnika kako bi tražili dozvolu za korišćenje multimedijalnih sadržaja. Ukoliko bi se različit vodeni žig umetao u multimedijalni sadržaj prilikom svake njegove legalne prodaje, u slučaju da dođe do nelegalne distribucije, detekcijom vodenog žiga mogle bi se identifikovati odgovorne osobe. Sistem poput ovog mogle bi koristiti i kompanije za sprečavanje curenja povjerljivih dokumenata, snimaka i fotografija.

Potpuna prevencija neovlašćenog kopiranja i distribucije multimedijalnih podataka pomoću votermarking sistema čini se veoma teško dostižnom. Da bi se to postiglo, softveri za reprodukciju multimedijalnih sadržaja, poput audio i video plejera, programa za pregled i obradu slika ili dokumenata, morali bi biti kreirani tako da na detekciju neautorizovanog sadržaja reaguju tako što zaustave reprodukciju. Da bi proizvođači ovih softvera počeli da ugrađuju komponente za detekciju vodenih žigova u svoje proizvode, oni bi na to morali biti obavezani odgovarajućom pravnom regulativom.

Pored predmeta u vezi sa zaštitom autorskih prava, votermarking sistemi mogli bi se primijeniti i u ostalim sudskim postupcima, za provjeru autentičnosti dokaza. Na osnovu člana 156 Zakonika o krivičnom postupku Crne Gore[2], audio i video snimci mogu se koristiti kao dokazni materijal na sudu, pa je potvrda njihove autentičnosti od izuzetne važnosti. Kriptografija ovaj problem rješava pomoću digitalnog potpisa. Digitalni potpis predstavlja šifrovani sažetak podataka koji treba autentifikovati. Kreira se tehnikom asimetrične kriptografije koja koristi privatni ključ za šifrovanje i javni ključ za dešifrovanje. Najprije se, pomoću odabrane heš funkcije, izračunava sažetak, obično od 128 ili 256 bitova. Dobijeni sažetak se zatim šifruje tajnim ključem. Autentičnost digitalnog signala se provjerava tako što se za priloženi signal izračuna sažetak korišćenjem identične heš funkcije. Dobijeni rezultat se upoređuje sa rezultatom dešifrovanja signala javnim ključem. Ako se rezultati prethodne dvije operacije poklapaju, signal se smatra autentičnim. Ukoliko neko vrši manipulacije nad digitalnim signalom ovi rezultati se ne mogu poklopiti. Digitalni potpis se ne može falsifikovati, jer je ključ za šifrovanje poznat samo njegovom vlasniku. Mana digitalnih potpisa je što su oni zapravo metapodaci koji su odvojeni od samih signala. Oni se mogu izgubiti u prenosu ili prilikom konverzija fajlova u druge formate koji ne predviđaju prostor za ove podatke. Bolje rješenje bi bilo umetnuti potpis u sami signal tehnikama votermarkinga. Međutim, procedura umetanja potpisa bi izmijenila signal, pa bi se prilikom autentifikacije dobila drugačija vrijednost heš funkcije i proces provjere bio bi neuspješan. Da bi se ovo prevazišlo signal se mora podijeliti na dva dijela. Na osnovu jednog dijela signala bi se izračunavao digitalni potpis, a drugi dio bi se koristio za umetanje potpisa. Problem koji ostaje je to da je veliki broj algoritama za generisanje digitalnog potpisa zaštićen patentima, što ograničava njihovu primjenu.

Sistemi vodenog žiga primjenu nalaze i u drugim zadacima digitalne forenzike, ne samo u provjeri autentičnosti signala. Ukoliko je signal podlegao izmjenama i vodeni žig se može iskoristiti kako bi se saznalo koji djelovi signala su izmijenjeni. Takođe, može se provjeriti koliko su značajne izmjene koja su se desile, kao i da li su te izmjene legitimne. Nekada je granica između legitimnih i nelegitimnih operacija nejasna, pa je poželjno identifikovati svaki od efekata koji je primijenjen nad signalom. Na kraju se može pokušati i rekonstrukcija originala ukoliko se identifikovani efekti mogu invertovati. Digitalni forenzičari mogu nekada odgovoriti na ova pitanja različitim tehnikama za detekciju anomalija u signalima. Prednost korišćenja vodenih žigova u ove svrhe, u odnosu na druge tehnike digitalne forenzike, je u tome što vodeni žig prolazi kroz sve modifikacije zajedno sa signalom, što može pomoći u zaključivanju koje su se promjene nad signalom desile. Vodeni žig takođe može je-

[2] „Službeni list Crne Gore" br. 57/2009, 49/2010, 35/2015, 145/2021

dinstveno identifikovati uređaj koji je proizveo digitalni signal nosilac u slučajevima kada je potrebno provjeriti porijeklo digitalnog sadržaja.

Votermarking sistemi mogu se koristiti i u vojnoj komunikaciji gdje se mora provjeriti autentičnost svake izdate naredbe. Zapravo, umetanje vodenih žigova može poboljšati sve sisteme u kojima je verifikacija sagovornika prilikom komunikacije od iznimnog značaja. Takvi sistemi mogu se primijeniti i u kontroli vazdušnog saobraćaja, prilikom komunikacije između pilota i kontrolora leta, kao i u internet telefoniji (VoIP).

Jednu veliku opasnost u današnjem svijetu predstavlja i pojava dipfejkova (*engl. deepfakes*). Dipfejkovi su vještački digitalni mediji (najčešće audio ili video), kreirani dubokim neuronskim mrežama. Kako ih moćne tehnike mašinskog učenja i vještačke inteligencije pomoću kojih su stvoreni čine naizgled vjerodostojnim, mogu se koristiti za stvaranje lažnih vijesti i manipulisanje javnošću. Širenje ovog sadržaja može imati nemjerljiv uticaj kako na pojedinca, tako i na čitavo društvo. Provjera autentičnosti takvih audio i video zapisa, koja bi se vršila pomoću sistema vodenog žiga, mogla bi spriječiti širenje dezinformacija i povredu ugleda institucije ili ličnosti. Povećana sofisticiranost ovih napada na integritet informacija iziskuje osmišljavanje novih metoda autentifikacije kako bi se navedene prijetnje otklonile.

Sve je više servisa koji zahtijevaju verifikaciju identiteta korisnika da bi im se moglo pristupiti. Identifikacija često podrazumijeva dostavljanje ličnih podataka koje korisnici ostavljaju na raspolaganju vlasnicima pomenutog servisa. Ti podaci mogu i često bivaju zloupotrijebljeni, najčešće u svrhe agresivnih marketinških kampanja i promocija. Još su veće opasnosti sa servisima preko kojih korisnici vrše osjetljive operacije poput bankarskih transakcija. Ukoliko bi se identifikacija vršila putem vodenih žigova, osjetljivi, privatni podaci korisnika bi mogli biti sačuvani.

Kada pojedinci ili kompanije žele reklamirati svoje proizvode i usluge, oni obično zakupe vrijeme za emitovanje na TV i radio stanicama. U prošlosti se dešavalo da vrijeme za emitovanje reklama bude pretrpano i da se pojedine reklame ne prikažu ugovoreni broj puta ili u ugovorenom trajanju. Oglašivači žele da budu sigurni da će plaćeni broj i vrijeme emitovanja biti zadovoljeni. Prvobitna ideja je angažovanje posmatrača koji bi pratili, mjerili i bilježili ono što čuju i vide. Jasno je da je ovakav način rješavanja ovog problema veoma skup i podložan greškama. Automatizovani sistem za monitoring emitovanja reklamnog sadržaja bio bi mnogo bolje rješenje. Takav sistem pratio bi emitovani sadržaj i tražio preklapanja emitovanih signala i referentnih signala u bazi podataka. Ukoliko se desi preklapanje, ono bi bilo zabilježeno. Međutim, korišćenje signala kao ključa za pretraživanje u bazi nije praktično. Jedan signal sadrži veliku količinu informacija i pravljenje indeksa za pretraživanje

baze na osnovu njih bilo bi nepodesno. Poređenje signala nije trivijalan zadatak. Veoma je moguće da do potpunog preklapanja signala i ne bi moglo doći, jer se signal tokom emitovanja izmijeni, najčešće degradira, pa bi se morala osmisliti tehnika kojom bi se pronašao najsličniji signal. To još dodatno komplikuje ovaj sistem. Ugrađivanjem vodenih žigova u emitovane signale mogli bi se prevazići neki od ovih problema. Umjesto čitavih signala poredili bi se samo vodeni žigovi što bi unaprijedilo efikasnost i robustnost ovih sistema. Ovakvi sistemi imali bi prednost u odnosu na sisteme koji su trenutno u upotrebi zato što ne bi zahtijevali nikakve promjene u procesu rada emitera. Trenutni sistemi za monitoring reklamnog sadržaja zahtijevaju od emitera da ubacuju identifikacione informacije u zaglavlje emitovanih digitalnih signala i da obezbijede njihov siguran prenos do prijemnika. Emiter je u ovom slučaju obavezan da uvede značajne izmjene u standardni način poslovanja zbog novih zahtjeva koji moraju biti zadovoljeni. Vodeni žigovi su sastavni dio signala i u potpunosti su kompatibilni sa trenutnim načinima emitovanja signala sa TV i radio stanica, pa predstavljaju dobro alternativno rješenje za monitoring emitovanja.

Umetanje i detekcija vodenih žigova ne mora služiti isključivo kao sigurnosni sloj u aplikacijama za vršenje provjera autentičnosti i prevenciju zloupotreba. Ove tehnike mogu se upotrijebiti i da poboljšaju korisničko iskustvo kod određenih servisa. Svjedoci smo jasne tendencije ljudi da dok na računaru ili televizoru gledaju film ili seriju uporedo koriste drugi uređaj, najčešće mobilni telefon, kako bi na internetu pronalazili dodatne informacije o tom sadržaju, glumcima, prethodnim epizodama ili referencama iz popularne kulture. Dodatno, gledaoci pristupaju i društvenim mrežama kako bi na njima ostavljali svoje utiske i komentarisali ih sa drugima. Korišćenje ovih sekundarnih, tzv. (*engl. second screen*) aplikacija je veoma popularan trend, pa su provajderi zabavnog sadržaja počeli proizvoditi sopstvene *second screen* aplikacije kako korisnici ne bi morali da koriste više aplikacija odjednom. Votermarking sistemi bi mogli doprinijeti većoj udobnosti prilikom korišćenja ovih servisa. Vodeni žigovi bili bi umetnuti u audio ili video zapise koje bi *second screen* aplikacija detekovala. Vodeni žig predstavljao bi ključ po kojem bi se dalje mogle tražiti i prikazati sve relevantne informacije, uključujući i trenutne aktuelnosti vezane za taj sadržaj.

1.4 Struktura disertacije

Ostatak disertacije organizovan je na sljedeći način. U drugom poglavlju su objašnjeni osnovni koncepti u oblasti votermarkinga. Takođe, predstavljeni su i zahtjevi koji se moraju ispuniti prilikom kreiranja votermarking sistema i izazovi koji se na

tom putu javljaju. Mjerila kojima se ocjenjuje uspješnost ovih sistema u ispunjavanju postavljenih zahtjeva predstavljena su u trećem poglavlju. Četvrto poglavlje sadrži detaljan pregled tradicionalnih tehnika za ugrađivanje vodenih žigova u digitalne audio signale. U petom poglavlju su opisane neuronske mreže i njima srodni koncepti, kao glavno oruđe koje se koristi za izvođenje votermarking šema u ovom radu. U šestom poglavlju predložene su nove tehnike za votermarking digitalnih audio signala zasnovane na dubokom učenju. Korpus podataka koji je korišten u okviru ove studije opisan je u sedmom poglavlju. U osmom poglavlju izloženi su i diskutovani postignuti rezultati u poređenju sa drugim relevantnim pristupima. Izvedeni zaključci i budući pravci istraživanja dati su u poslednjem, devetom poglavlju.

2 Osnove digitalnog votermarkinga

Uopštena šema sistema vodenog žiga data je na Slici 1. Sistem sadrži dvije osnovne komponente:

- umetač,
- detektor.

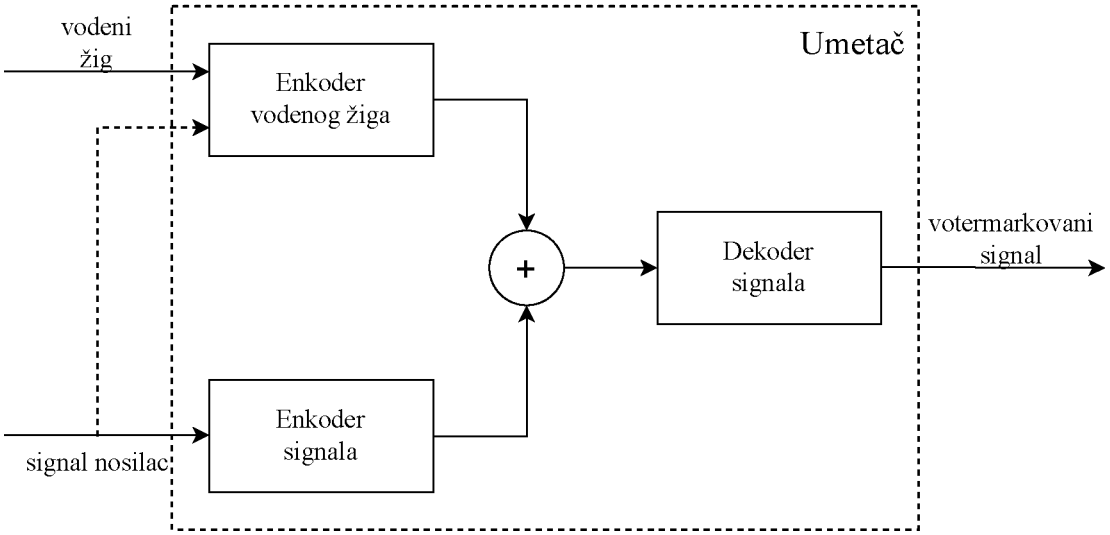
Umetač ugrađuje bitove vodenog žiga u signal nosilac i proizvodi votermarkovani signal. Detektor prepoznaje signale sa vodenim žigom i ekstrahuje ugrađene bitove iz njih.

Pred votermarking sistemom su postavljeni određeni zahtjevi. U zavisnosti od toga kako sistem odgovori na te zahtjeve ocjenjuje se njegov kvalitet. Umetanje žigova treba izvršiti tako da su oni što manje primjetni u signalu. Prilikom detekcije treba praviti što manje grešaka. Takođe, poželjno je da je sistem u mogućnosti da ugradi što više bitova vodenog žiga u signal. Sve operacije treba izvršiti uz što manje trošenje računarskih resursa.

Slika 1 sistem vodenog žiga predstavlja kao model sistema za komunikaciju u kojem se odabrana poruka (vodeni žig) prenosi od pošiljaoca (umetača) do primaoca (detektora). Na putu između umetača i detektora nalazi se komunikacioni kanal unutar kojeg se signal može u nekoj mjeri izmijeniti. U većini aplikacija od sistema vodenog žiga se očekuje da budu robustni, odnosno otporni na ovakve smetnje. Pored promjena nastalih u toku prenosa signala, promjene u signalu mogu biti izazvane i na druge načine. Nekada su one neophodne i opravdane, a nekada su posljedica malicioznog ili nestručnog rukovanja signalom. Sistem mora biti u stanju da neometano nastavi funkcionisanje i u slučaju takvih smetnji ili barem mora prepoznati da je došlo do oštećenja signala.

Sve ovo ukazuje na složenost kreiranja kvalitetnog votermarking sistema. Nemoguće je na idealan način ispuniti sve zahtjeve, pa je potrebno napraviti kompromise. Prioritet zahtjeva diktiran je namjenom samog sistema.

Slika 1: Uopštena šema sistema vodenog žiga.



Slika 2: Blok šema umetača.

U nastavku ovog poglavlja detaljnije su opisane dvije komponente votermarking sistema i njihovi različiti tipovi. Dati su opšti modeli po kojima se kreiraju šeme za umetanje i detekciju vodenih žigova. U trećem dijelu poglavlja opisane su odlike votermarking sistema na osnovu kojih se vrši njihovo poređenje.

2.1 Umetanje vodenih žigova

Umetač je dio votermarking sistema koji vrši skrivanje vodenih žigova u signalima nosiocima. On ima dva ulaza. Prvi je signal koji se votermarkuje x, a drugi je niz w od L_w bitova koji predstavlja vodeni žig. Izlaz iz umetača je signal sa umetnutim vodenim žigom. Pojednostavljena blok šema umetača data je na Slici 2.

Proces umetanja vodenih žigova u digitalne signale sastoji se iz više faza. Odabrani vodeni žig može podleći transformacijama prije nego se efektivno doda u signal nosilac. Nekada se vodeni žig šifruje tajnim ključem prije umetanja kako bi se povećala njegova sigurnost. Po potrebi se mogu koristiti i druge operacije koje vodeni žig transformišu u obrazac pogodan za umetanje u digitalni signal. U radu [1], predloženo je da se vodeni žig kodira u sekvencu pseudošuma (*engl. pseudo-noise sequence* - PN) i da se kao takav ugrađuje u signal. Takođe, kako je vodeni žig značajno manjih dimenzija od signala nosioca, skaliranje je operacija koja se jako često vrši nad vodenim žigom. Ova operacija se najčešće izvodi višestrukim repliciranjem vodenog žiga, kako bi se njegove dimenzije poklopile sa dimenzijama reprezentacije signala nosioca. Komponenta umetača u kojoj se vrše ove operacije nad vodenim žigom naziva se enkoder vodenog žiga. Na izlazu iz ove komponente imamo kodirani i preprocesirani vodeni žig v, spreman za ugrađivanje u signal.

Signal nosilac takođe prolazi kroz nekoliko koraka obrade prije samog umetanja vodenog žiga kroz komponentu enkodera signala (Slika 2). Osnovna operacija u ovoj komponenti je prebacivanje ulaznog signala u domen u kojem će se izvršiti umetanje vodenog žiga. Nekada se ovaj korak može preskočiti i umetanje žiga se može izvršiti u izvornom domenu signala, što bi za audio signale bio vremenski domen. Tada bi opšta formula za umetanje jednog bita vodenog žiga izgledala ovako:

$$y(n) = x(n) + \alpha v(n), \tag{1}$$

gdje je x izvorni oblik signala nosioca, v skalirani i kodirani vodeni žig, y predstavlja votermarkovani signal, a n je indeks odbirka signala. Simbol n korišten je i u ostatku disertacije da označi odbirak signala u vremenskom domenu. Parametrom α se kontroliše snaga vodenog žiga i pravi se kompromis između neprimjetnosti vodenog žiga, sa jedne, i njegove robustnosti, sa druge strane. U ovoj jednakosti α je konstanta, ali se u različitim pristupima ovaj parametar prilagođava sadržaju signala sofisticiranim tehnikama kako bi se pravio balans između robustnosti i neprimjetnosti vodenog žiga [1, 2]. Takođe, operacija sabiranja odbiraka signala nosioca sa vodenim žigom iz prethodne jednakosti predstavlja najjednostavniji način votermarkovanja digitalnog signala. Ovako jednostavnom operacijom, očekivano, željene performanse sistema neće biti dostignute. Ona je ovdje data da posluži kao uopšteni primjer kojim bi se unificirale sve funkcije za umetanje vodenih žigova, ali ona se ne koristi u praksi. Konstrukcija kompleksnijih i domišljatijih operacija za umetanje vodenih žigova glavna je tema svih naučnih radova u ovoj oblasti.

Umetanje vodenih žigova u izvornom domenu signala je efikasno. Međutim, u literaturi se ipak češće pribjegava korišćenju transformacionih domena. Prebacivanjem signala u neki drugi domen može se značajno olakšati kreiranje algoritama za umetanje i detekciju vodenih žigova. U transformacionom domenu karakteristike signala se jasnije izdvajaju, pa su šeme umetanja i detekcije vodenih žigova jednostavnije. Takođe je lakše zaključiti koje će modifikacije manje uticati na ljudsku percepciju signala i na taj način poboljšavati kvalitet votermarkovanog signala. Za očekivati je i da robustnost bude superiorna, jer se modifikacije nad signalom u manjoj mjeri odražavaju na reprezentacije signala u transformacionim domenima. U nastavku disertacije, odbirci signala u transformacionom domenu će biti indeksirani sa k, dok će za odbirke u vremenskom domenu biti zadržana oznaka n. Ova notacija je uvedena kako bi se jasno razlikovala ova dva fundamentalna pristupa ugrađivanju vodenih žigova.

Furijeova transformacija je vrsta matematičke operacije koja omogućava upravo ovakvo mapiranje. Način izračunavanja koeficijenata Furijeove transformacije za digitalne signale, kao i način rekonstrukcije signala iz ovih koeficijenata opisani su u

Prilogu A.

Uopštena formula za umetanje jednog bita vodenog žiga u transformacionom domenu gotovo je identična formuli (1) za vremenski domen:

$$Y(k) = X(k) + \alpha v(k). \tag{2}$$

X je reprezentacija ulaznog signala u transformacionom domenu, a k je indeks odbirka. Votermarkovani signal na izlazu, označen sa Y, će također biti u istom domenu. Stoga je neophodno da se u posljednjem koraku procedure umetanja signal vrati u izvorni domen i kao takav emituje. Na Slici 2 je komponenta koja obavlja ovaj zadatak označena kao dekoder signala. Pored diskretne Furijeove transformacije, za ugrađivanje bitova vodenog žiga intenzivno se koriste i koeficijenti diskretne kosinusne transformacije (DCT) [3, 4] ili diskretne vejvlet (*engl. wavelet*) transformacije (DWT) [5]. Nad ovim koeficijentima se mogu vršiti dalje operacije u cilju izvlačenja robustnijih karakteristika signala, pa se sa tim vrijednostima vrši umetanje vodenih žigova. Nekada se iz koeficijenata transformacije izračunavaju amplituda i faza signala, pa se votermarking procedure sprovode nad jednim od tih kanala.

U jednakostima (1) i (2) korišteni su različiti indeksi n i k za odbirke signala u vremenskom, odnosno transformacionom domenu. Iste oznake indeksa zadržane su i u nastavku disertacije kako bi prilikom definisanja votermarking šema bilo jasno u kojem se domenu primjenjuju.

Šeme za umetanje date u jednakostma (1) i (2) karakterišu se kao slijepe ili neinformisane. Razlog tome je što se vodeni žig kodira nezavisno od signala nosioca u koji se ugrađuje, odnosno, svojstva signala nosioca ne utiču na proces dodavanja vodenog žiga, već samo na krajnji rezultat. Kako je signal nosilac poznat prilikom umetanja, sasvim je legitimno koristiti ga u sklopu algoritma za umetanje. Prije kodiranja vodenog žiga mogu se ispitati svojstva signala nosioca i vodeni žig se može kodirati tako da se što bolje ugradi u dati signal. Na ovaj način mogu se kreirati mnogo kvalitetniji votermarking sistemi. Vodeni žig se može kodirati tako da bude robustniji, a može se i provjeriti kako dodavanje vodenog žiga utiče na kvalitet signala i u skladu sa tim izvršiti modifikacije vodenog žiga kako bi se kvalitet poboljšao. Umetač koji kodira vodeni žig prema sadržaju signala nosioca naziva se informisanim. Šeme informisanog i neinformisanog umetača objedinjene su na Slici 2. Skica informisanog umetača se, u odnosu na neinformisani, razlikuje samo u jednom detalju, a to je dodatni ulaz za enkoder vodenog žiga koji predstavlja signal nosilac. Ovaj ulaz iscrtan je isprekidanom linijom na slici.

Uopštena šema umetanja (1) se u slučaju informisanog umetača modifikuje na sljedeći način:

$$y(n) = x(n) + \alpha\Phi\left(x(n), v(n)\right). \tag{3}$$

Šema u transformacionom domenu definiše se analogno. Funkcijom Φ se generiše obrazac za umetanje koji se dobija na osnovu vodenog žiga i signala nosioca. Ova funkcija može biti veoma složena i često podrazumijeva korišćenje perceptivnih modela kako bi se procijenio uticaj dodavanja vodenog žiga na kvalitet signala. U informisanim pristupima se najčešće svaki vodeni žig predstavi sa nekoliko različitih vektora. Zatim se prilikom ugrađivanja vodenog žiga odgovarajući vektor bira na jedan od dva moguća načina. Prvim pristupom se maksimizuje robustnost dok se kvalitet signala održava iznad definisane granice, a drugim se maksimizuje kvalitet signala uz održavanje robustnosti na zadatoj granici.

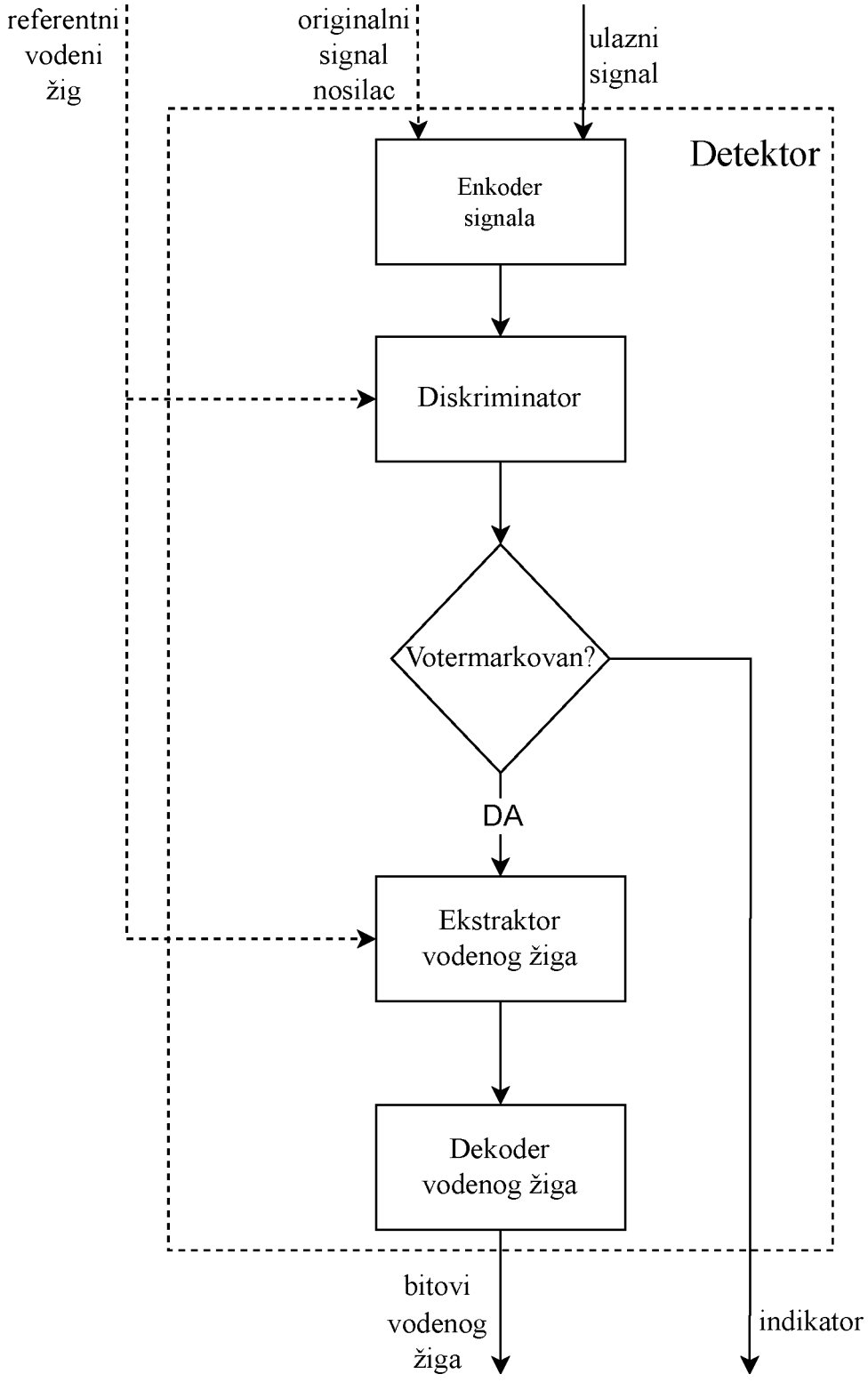
Ustaljeni pristup u savremenim tehnikama za umetanje vodenog žiga je podjela signala na disjunktne intervale i skrivanje vodenog žiga u intervalu odbiraka signala, odnosno koeficijenata odabrane transformacije, umjesto u pojedinačnom odbirku ili koeficijentu, kako je definisano prethodnim jednakostima. Procedura umetanja obično uključuje izračunavanjem neke karakteristične vrijednosti za čitav interval, koja se zatim koristi za ugrađivanje vodenog žiga. Postupak izračunavanja ove vrijednosti razlikuje se od tehnike do tehnike. Uopšteno, može se definisati kao funkcija više promjenljivih Ψ koja preslikava prostor odbiraka $\mathbb{R}^{L_I}$, gdje L_I predstavlja dužinu intervala, odnosno broj odbiraka u intervalu, u skup realnih brojeva $\mathbb{R}$. Opisana vrsta procedure umetanja može se definisati sljedećom jednakošću:

$$y(n) = x(n) + \alpha\Phi\left(\Psi\left(x(n_1), x(n_2), ..., x(n_{L_I})\right), v(n)\right). \tag{4}$$

Ovim načinom dodavanja vodeni žig će se manje osjetiti prilikom reprodukcije signala. Takođe, detekcija vodenih žigova je manje osjetljiva na modifikacije signala. Pojedinačni odbirci se mogu lako poremetiti, dok je čitav interval i odnose među odbircima u intervalu teže poremetiti operacijama obrade signala. Gubi se jedino na broju bitova vodenog žiga koji se mogu ugraditi. Primjer ovakvog načina umetanja vodenih žigova imamo u radovima [6, 7], gdje se vrši dekompozicija na singularne vrijednosti (*engl. singular value decomposition* - SVD) koeficijenata odabrane transformacije, pa se te vrijednosti koriste za ugrađivanje bitova vodenog žiga.

2.2 Detekcija vodenih žigova

Detektor je dio votermarking sistema koji ima dva zadatka. Prvi je da detektuje da li je u datom signalu z prisutan vodeni žig. Ovaj zadatak, odnosno razlikovanje votermarkovanih i nevotermarkovanih signala izvršava komponenta koju nazivamo diskriminator. Ukoliko diskriminator prepozna da je signal votermarkovan, potrebno je ekstrahovati umetnute bitove vodenog žiga v odgovarajućom komponentom. Posljednja komponenta u detektoru je dekoder vodenog žiga koji vraća vodeni žig iz



Slika 3: Blok šema detektora.

kodiranog oblika v u izvorni oblik w, prije transformacija koje su prethodile umetanju. Blok šema detektora data je na Slici 3. Prva komponenta detektora je enkoder signala koji transformiše signal u domen u kojem je izvršeno umetanje vodenog žiga. Očekivano je da se i detekcija vrši u istom domenu kao i umetanje i to je princip kojeg se pridržavaju svi poznati pristupi iz literature.

Detektor ima dva izlaza. Jedan izlaz detektora je indikator koji označava da li je u signalu detektovan vodeni žig, a na drugom izlazu su ekstrahovani bitovi vodenog žiga. Dva izlaza detektora se mogu spojiti u jedan ukoliko se iz skupa svih vodenih žigova izdvoji jedan koji se ne ugrađuje u signale, već služi kao indikator da signal nije votermarkovan, kada se nađe na izlazu detektora. U idealnom slučaju, izlaz iz detektora ima 2^{L_w} mogućih vrijednosti, jer je to broj različitih žigova od L_w bitova. Međutim, kreiranje ovakvog sistema obično se negativno odražava na ostale performanse, pa se votermarking sistemi dizajniraju tako da mogu ugrađivati manji broj različitih vodenih žigova. U literaturi je predložen veliki broj sistema koji imaju mogućnost ugrađivanja samo jednog žiga [8, 9]. Jedini zadatak detektora u takvom sistemu je da odredi da li ulazni signal sadrži vodeni žig, a ukoliko je signal votermarkovan, bitovi su jednoznačno određeni.

U zavisnosti od broja ulaza, razlikujemo dvije vrste detektora. Slijepi (neinformisani) detektori na ulaz dobijaju samo jedan signal koji može biti votermarkovan ili ne. Sa druge strane, ukoliko je detektoru na ulazu dostupan i originalni signal nosilac, tu vrstu detektora nazivamo informisanim detektorom. Šema sa Slike 3 objedinjuje obje vrste detektora. Opcioni ulazi predstavljeni su isprekidanim linijama. Na skici se može primijetiti da se detektoru na ulaz može dovesti i referentni vodeni žig. Većina detektora u literaturi podrazumijeva da prilikom detekcije ima na raspolaganju referentni vodeni žig. Detekcija se vrši tako što u skupu svih vodenih žigova pronađe onaj koji je najvjerovatnije ugrađen u signal. Iako ovi detektori dobijaju dodatne informacije na ulazu, pored onih neophodnih, oni se smatraju neinformisanim.

Obje vrste detektora koriste se u praksi. Očekivano, informisani detektori imaju mnogo bolje performanse od slijepih. Dovoljno je samo oduzeti originalni od votermarkovanog signala i detekciju vršiti samo na dobijenoj razlici, što značajno olakšava problem detekcije. Nekada je informisani detektor dovoljno snabdjeti samo dijelom informacija o signalu nosiocu, npr. dijelom koeficijenata odabrane transformacije, kako bi se poništio uticaj signala nosioca na vodeni žig. Informisani detektori mogu se upotrebljavati u sistemima u kojima samo ograničena grupa korisnika može vršiti detekciju. U sistemima gdje veliki broj korisnika može ili treba vršiti detekciju, ovakvi detektori gube smisao.

Slijepi detektori imaju mnogo širu primjenu. U nekim situacijama ne možemo očekivati da imamo dostupan originalni signal ili neki njegov dio prilikom detekcije vodenih žigova, čak i kada je broj korisnika s pravom detekcije ograničen. Nekada zbog praktičnih razloga, a nekada originalni signali i ne postoje. Na primjer, u slučaju programa koji pretražuje internet kako bi otkrio povrede autorskih prava, ako bi se koristio informisani detektor, on bi morao prilikom svake provjere određenog signala vršiti detekciju u paru sa svim originalnim signalima iz baze. Ipak, informisani detektor bi mogao biti od pomoći u ovom slučaju. Nakon što slijepi detektor otkrije potencijalno nelegalno distribuiranu kopiju, vlasnik bi mogao pronaći i dostaviti nevotermarkovani original, kako bi se provjera izvršila i sa informisanim detektorom i time minimizovala mogućnost greške. Kod nekih aplikacija votermarkinga, kao što su provjera falsifikata, odnosno detekcija dipfejkova, originalni snimak neće postojati, pa je u tim slučajevima korišćenje informisanog detektora nemoguće. Nedostatak korišćenja informisanih detektora je i u potrebi za dodatnim prostorom koji bi se morao namijeniti za skladištenje originalnih snimaka. Originalni snimci iziskivali bi i dodatno vrijeme za prenos, kada je to potrebno, ili bi nužno bilo povećanje propusnog opsega mreže. Ovi problemi bili bi naročito izraženi u primjenama votermarkinga u komunikaciji, internet telefoniji ili monitoringu emitovanja radio i TV stanica.

Većina detektora u tradicionalnim votermarking sistemima zasnovana je na korelaciji. Detekcija se vrši izračunavanjem korelacije ulaznog signala i referentnog vodenog žiga i njenim upoređivanjem sa predefinisanim pragom. Ukoliko za neki vodeni žig vrijednost korelacije sa signalom pređe definisani prag, smatra se da je signal votermarkovan upravo tim žigom. Ulazni signal se često nizom transformacija, kao u proceduri umetanja, prebacuje u domen pogodniji za detekciju žiga. Korelacija se zatim izračunava sa dobijenom reprezentacijom signala. Na primjer, prvi korak informisanih detektora je oduzimanje originalnog signala nosioca i ulaznog signala detektora kako bi se dobio vektor koji će imati veću korelaciju sa odgovarajućim vodenim žigom.

U nekim sistemima se prilikom opisivanja algoritma za detekciju ne navodi eksplicitno nijedna mjera korelacije. Veliki broj savremenih pristupa za detekciju vodenih žigova zasnovan je na traženju karakterističnih mjesta u signalu i izračunavanju vrijednosti koje će ukazati da je u tom dijelu signala skriven bit vodenog žiga. Međutim, i kod ovakvih sistema često se može pronaći veza opisane šeme za detekciju i korelacije ili bar analiza njihovih performansi mora uključivati ispitivanje efekta korelacije.

Ispitivanje korelacije može se vršiti i u cilju detekcije samo jednog bita vodenog žiga. Ovo sistem oslobađa od značajnih praktičnih ograničenja koja bi u suprotnom bila nametnuta. Ukoliko bi se računanje korelacije vršilo za sve moguće vodene žigove, to bi značilo da sistem treba da izvrši iscrpnu pretragu čitavog prostora

vodenih žigova, što bi bilo praktično neizvodljivo. Ekstrakcijom jednog po jednog bita se zaobilaze ovi problemi, jer je broj mogućih vrijednosti za bitove 2, dok ukupan broj vodenih žigova može biti izuzetno veliki.

Vrijednost korelacije može se računati na više načina. Najjednostavnija mjera je linearna korelacija:

$$r_L(z, v) = \frac{1}{N} z \cdot v. \tag{5}$$

Ova mjera, u osnovi, predstavlja skalarni proizvod ulaznog signala z i referentnog vodenog žiga v ($z \cdot v$). Linearna korelacija se na sličan način može definisati i u transformacionom domenu. Svaki detektor vodenog žiga koji izračunava neku linearnu funkciju odbiraka ulaznog signala i upoređuje dobijenu vrijednost sa datim pragom smatra se korelacionim.

Mana linearne korelacije je što su njene vrijednosti visoko zavisne od opsega u kojem su vrijednosti odbiraka signala. Stoga je šema za detekciju, zasnovana na linearnoj korelaciji, neotporna na efekte poput skaliranja amplitude signala, odnosno smanjivanja i pojačavanja jačine zvuka ili osvjetljenja piksela slike. Ovaj problem može se prevazići normalizacijom, odnosno svođenjem signala i vodenog žiga na jedinične vektore prije računanja skalarnog proizvoda. Jednakost za računanje normalizovane korelacije data je u vremenskom domenu. Ova mjera ima analognu definiciju u transformacionom domenu.

$$r_N(z, v) = \frac{z \cdot v}{\|z\|\|v\|}. \tag{6}$$

Kao testna statistika u šemama za detekciju vodenih žigova može se koristiti i koeficijent korelacije. Izračunavanje ove mjere sastoji se u umanjivanju svakog odbirka vektora za njegovu srednju vrijednost prije računanja skalarnog proizvoda.

$$r_C(z, v) = \frac{(z - \overline{z}) \cdot (v - \overline{v})}{\|(z - \overline{z})\|\|(v - \overline{v})\|}, \tag{7}$$

gdje $\overline{z}$ i $\overline{v}$ predstavljaju srednje vrijednosti vektora z i v, respektivno.

Vrijednost korelacije vodenog žiga v i signala z, izračunata na osnovu odabrane jednakosti (5), (6) ili (7), upoređuje se sa definisanim pragom τ i ukoliko vrijednost prelazi dati prag, procjenjuje se da je signal z označen vodenim žigom v. Konačno, ekstrahovani vodeni žig se dobija dekodiranjem u izvorni domen. Pseudokod za opisanu proceduru dat je Algoritmom 1.

Algoritam 1 Pseudokod za proceduru detekcije vodenog žiga. Parametar z predstavlja ulazni signal detektora, x je opcioni parametar koji predstavlja originalni signal nosilac, $\mathcal{V}$ je skup vodenih žigova, a τ je prag za detekciju.

```
1: procedure DETEKCIJA(z, x, V, τ)
2:     for v ∈ V do
3:         z ← KODIRANJESIGNALA(z, x)
4:         r ← KORELACIJA(z, v)
5:         if r > τ then return DEKODIRANJEVODENOGŽIGA(v)
6:     return ∅
```

2.3 Odlike sistema vodenog žiga

2.3.1 Robustnost vodenog žiga

Nakon što napusti umetač, a prije nego dođe do detektora, votermarkovani signal, označen sa y, može podleći različitim izmjenama. Na ulaz detektora tada dolazi signal z koji predstavlja signal y izmijenjen nepoznatom funkcijom $\mathcal{E}$ ($z = \mathcal{E}(y)$). Takođe, nevotermarkovani signal x može biti izmijenjen istim skupom efekata i doveden na ulaz detektora. Neke od ovih izmjena su posljedica legitimnih operacija koje se uobičajeno vrše nad signalom u različitim sistemima, dok su druge rezultat nestručnog ili zlonamjernog rukovanja signalom. Često je potrebno da votermarking sistem bude otporan na odgovarajući skup ovih operacija. Izbor tog skupa zavisi od namjene sistema. U pojedinim aplikacijama, poput potvrđivanja autentičnosti digitalnih signala, otpornost detektora na bilo kakve modifikacije votermarkovanih signala je nije poželjna. Detektori u tim slučajevima treba da ekstrahuju bitove samo iz nepromijenjenih signala, dok se u slučaju i najmanje modifikacije signal mora označiti kao nevotermarkovan. Ovakva vrsta vodenog žiga se naziva krhkim (*engl. fragile*) i posvećena mu je značajna pažnja u literaturi [10–12]. Postoje i polukrhki (*engl. semi-fragile*) [13] vodeni žigovi koji su otporni na manje modifikacije votermarkovanog signala, dok se značajnijim modifikacijama vodeni žig poništava. Ipak, većina aplikacija zahtijeva otpornost na efekte za koje se smatra da se mogu desiti između vremena ugradnje i detekcije. Nasuprot otpornosti na izmjene votermarkovanog signala, pogrešno prepoznavanje vodenog žiga u nevotermarkovanim signalima nepoželjno je u svim votermarking sistemima.

Ukoliko performanse votermarking sistema ne opadaju značajno pod dejstvom audio efekata, on se smatra robustnim. Vodeni žig ne mora biti otporan na operacije koje u potpunosti izobliče signal ili ga učine beskorisnim. Ukoliko je signal neupotrebljiv i vodeni žig tada gubi smisao.

2.3.1.1 Tipovi izobličenja audio signala

Zbog obilja efekata koji se mogu primijeniti nad signalom, robustnost predstavlja najveći izazov sistemima vodenog žiga. Robustnost na neki efekat je gotovo nemoguće postići bez kompromisa u kvalitetu signala, broju bitova koji se mogu ugraditi ili robustnosti na neku drugu operaciju. Stoga je poželjno ignorisati one efekte na koje nije nužno da sistem bude otporan i fokusirati se na one za koje postoji realna šansa da se dese i ugroze funkcionisanje sistema. U nastavku su opisani uobičajeni efekti koji mogu izazvati distorzije u gotovo svim votermarkovanim signalima. Otpornost na efekte iz ove grupe je jedan od ključnih preduslova za postizanje robustnosti sistema vodenog žiga.

2.3.1.1.1 Šum

Kada se votermarkovani signal prenosi komunikacionim kanalom za očekivati je da može biti pogođen šumom. Ukoliko ovaj šum ne degradira drastično kvalitet signala, votermarking sistem treba da bude otporan na njega. Šum se najčešće predstavlja po sljedećem aditivnom modelu:

$$z(n) = y(n) + \epsilon(n), \tag{8}$$

gdje y predstavlja ulazni signal, ϵ šum, a z je zašumljeni signal. Dodavanje manje količine šuma u signal neće značajno poremetiti korelaciju vodenog žiga i signala, pa su i tradicionalni korelacioni detektori otporni na ovu vrstu degradiranja signala.

2.3.1.1.2 Skaliranje amplitude

Audio votermarking sistemi u većini aplikacija treba da budu otporni i na efekte poput promjene jačine zvuka, odnosno skaliranje amplitude. Ukoliko bi napadač, ili čak sami korisnik, ovako jednostavnom operacijom onemogućio detekciju vodenog žiga, votermarking sistem izgubio bi svoju svrhu. Ovaj efekat se može predstaviti sljedećom jednakošću:

$$z(n) = ay(n), \tag{9}$$

gdje je a konstanta kojom se kontroliše skaliranje amplitude signala, dok su y i z originalni i rezultujući signal, respektivno.

2.3.1.1.3 Kompresija

Ukoliko se nalazi u okviru kompleksnijeg sistema, sistem vodenog žiga mora biti kompatibilan sa svim njegovim komponentama. Umetanje vodenog žiga ne smije

onemogućiti primjenu nekih drugih, sasvim validnih, operacija nad signalom. Na primjer, za potrebe skladištenja, signal se obično mora komprimovati, uglavnom sa gubicima. Ovim procesom se signal predstavlja s manjom količinom informacija od originalne. Time se obično gube podaci koji nisu od presudnog značaja za percepciju signala. Algoritmi za kompresiju i umetanje vodenih žigova su fundamentalno suprotstavljeni. Idealan algoritam kompresije treba da obriše sve redudantne podatke u reprezentaciji signala. Vodeni žigovi su jedna vrsta redudantnih podataka i oni bi u tom slučaju bili obrisani. Da bi vodeni žig opstao u signalu nakon ovakve kompresije, on bi morao unijeti perceptivne promjene u signal, što bi se najvjerovatnije negativno odrazilo na njegov kvalitet.

Algoritmi kompresije sastoje se iz više koraka od kojih neki dovode do gubitka informacija, a neki ne. Na početku algoritma se obično izračunavaju koeficijenti odabrane transformacije, poput DCT ili DWT. Ovo je proces koji ne izaziva gubitke. Takođe, primjenom entropijskih kodova podaci se predstavljaju sa najmanjim potrebnim brojem bitova, ali bez gubitaka. Operacije prilikom koje može doći do gubitka informacija u kompresiji su kvantizacija i filtriranje signala. To su dvije operacije na koje votermarking sistem treba da bude otporan. Kvantizacija se može predstaviti sljedećom jednakošću:

$$z(n) = q \left\lfloor \frac{y(n)}{q} + 0.5 \right\rfloor , \tag{10}$$

gdje je z kvantizovani signal, y signal prije kvantizacije, a parametar q je faktor kvantizacije.

Kvantizacijom se svi odbirci signala zaokružuju na najbliži cjelobrojni umnožak faktora kvantizacije q. Promjene odbiraka signala, izazvane kvantizacijom, mogu se, uz određene pretpostavke, posmatrati kao šum. Ako je vrijednost odbirka jednaka cjelobrojnom umnošku faktora q, kvantizacija ga neće izmijeniti. U suprotnom, originalna vrijednost odbirka biće uvećana ili umanjena najviše za $q/2$. Pod pretpostavkom da je svaka vrijednost odbirka između dva uzastopna cjelobrojna umnoška kvantizacionog faktora jednako vjerovatna, zaključuje se da kvantizacioni šum ima uniformnu raspodjelu $\mathcal{U}(-q/2, q/2)$. Slijedi da se kvantizacija može simulirati aditivnim modelom šuma, sa odgovarajućom funkcijom raspodjele. Ovaj model kvantizacije može se smatrati vjerodostojnim ukoliko je faktor kvantizacije q dovoljno mala vrijednost. Neophodan uslov je da važi da je: $2\pi/q > 2\omega_{\max}$ [14], gdje je $\omega_{\max}$ maksimalna frekvencija prisutna u signalu.

2.3.1.1.4 Digitalni filtri

Druga česta vrsta obrade koja se može vršiti i u svrhe kompresije signala je njegovo filtriranje. Digitalni filtar je matematička operacija koja se primjenjuje nad digitalnim signalom sa ciljem da se promijeni njegov spektralni sadržaj. Ova operacija izvodi se konvolucijom signala y i odabranog filtra h:

$$z = y * h. \tag{11}$$

Funkcija h naziva se impulsni odziv filtra. To je funkcija vremena koja predstavlja izlaz filtra kada mu se na ulaz dovede jedinična delta funkcija.

Pored impulsnog odziva, filtar se može definisati i diferencnom jednačinom, funkcijom prenosa, skupom nula i polova ili frekvencijskim odzivom. Frekvencijski odziv filtra $H(j\omega)$ je funkcija ugaone frekvencije ω koja opisuje kakav će biti spektralni sadržaj izlaza iz filtra. Ova funkcija za datu frekvenciju određuje kakva će biti amplituda i koliki će biti fazni pomjeraj izlaznog signala u odnosu na ulazni signal na toj frekvenciji. Na osnovu frekvencijskog odziva filtra mogu se odrediti njegova amplitudna i fazna karakteristika. Amplitudna karakteristika filtra $|H(j\omega)|$ predstavlja nivo odnosa izlazne i ulazne amplitude pri različitim frekvencijama. Faznom karakteristikom se opisuje pomjeraj u fazi izlaznog signala u odnosu na ulazni na određenoj frekvenciji. Prilikom dizajniranja filtra veća pažnja posvećuje se amplitudnoj karakteristici, kojom se ilustruje kako filter pojačava ili prigušuje različite frekvencijske komponente signala.

Frekvencijski i impulsni odziv su povezani na taj način što je frekvencijski odziv zapravo Furijeova transformacija impulsnog odziva. Izlaz filtra može se dobiti konvolucijom ulaznog signala sa impulsnim odzivom filtra u vremenskom domenu ili množenjem sa frekvencijskim odzivom u frekvencijskom domenu.

Primjenom filtra pojedini djelovi signala se ublažavaju ili dodatno naglašavaju, u zavisnosti od njegove namjene. S odgovarajućim filtrom se oni mogu oslabiti do gotovo potpunog isčezavanja. Tada se ti djelovi signala mogu u potpunosti ignorisati ili predstaviti veoma malim brojem bitova, što se može iskoristiti u kompresiji. Na primjer, za kompresiju govornih signala mogu se eliminisati sve frekvencijske komponente iznad 8 kHz, jer na tim frekvencijama definitivno neće biti sadržaja.

Filtriranje nije korisno samo u svrsi kompresije signala. Ova operacija koristi se u mnogim drugim vrstama obrade signala. Eliminacija šuma iz signala često se vrši njegovim propuštanjem kroz odgovarajući filtar. Vodeni žig se može posmatrati kao jedna vrsta šuma koja se dodaje u signal. Ukoliko se nalazi u djelovima signala koje filtar značajno mijenja, može biti obrisan. Zato je potrebno da se postigne otpornost

na odgovarajuću grupu filtara za koje se pretpostavlja ili zna da će se primjenjivati nad signalima. Da bi se ovo postiglo vodeni žig treba umetnuti tamo gdje filtar ne unosi velike promjene ili kreirati detektor koji se može prilagoditi promjenama koje filtar izazove.

Reprodukcija audio signala u nekom ambijentu se takođe može aproksimirati pomoću filtra. Sistem koji je otporan na ovaj efekat može se koristiti da detektuje vodene žigove u audio signalima u realnim okruženjima, u kojima se reprodukuje zvučni sadržaj. Ova sposobnost daje dodatnu prednost u borbi za zaštitu autorskih prava i mnogim drugim primjenama.

2.3.1.2 Tehnike za postizanje robustnosti

Postoji nekoliko strategija za neutralisanje prethodno pomenutih efekata. Neke su opšte i mogu se primijeniti na različite efekte, a neke su razvijene za prevazilaženje konkretnih efekata. Ove strategije ne isključuju jedna drugu i više njih se može primjenjivati u okviru istog sistema.

Strategija koja se koristi u gotovo svim pristupima u literaturi je izbjegavanje ugrađivanja vodenih žigova u visokofrekventnim komponentama signala. Na ovaj način postiže se otpornost na niskopropusno filtriranje i druge vrste operacija koje degradiraju visokofrekventne koeficijente. To je veoma važno jer ljudsko uho nije osjetljivo na visoke frekvencije, pa one mogu biti neopaženo obrisane, a sa njima i vodeni žig, ukoliko je tu ugrađen. Stoga bi korišćenje ovih frekvencija za umetanje vodenih žigova učinilo sistem nepouzdanim. Osim toga, većina efekata je dizajnirana tako da očuva dijelove signala koji su značajni za njegovu percepciju, pa je korišćenje tih djelova pogodnije za umetanje vodenih žigova u cilju dostizanja robustnosti. Međutim, neprimjetnost vodenog žiga je u ovom slučaju mnogo teže dostići, jer se modifikacijama perceptivno važnih koeficijenata degradira kvalitet signala. U zavisnosti od namjene sistema potrebno je prioritizovati kvalitet ili robustnost, ili napraviti kompromis. Na primjer, moguće je za umetanje žigova koristiti koeficijente srednje važnosti u pogledu percepcije [7, 15]. Na ovaj način se modifikuju djelovi signala koji neće ugroziti kvalitet, a postoji dobra šansa i da ih uobičajene operacije nad signalom ne poremete.

Veliki broj predloženih tehnika vrši redudantno umetanje kojim se jedan bit vodenog žiga umeće na više mjesta u signalu. Prilikom detekcije se, ako je jedan dio signala degradiran, vodeni žig može izvaditi sa drugih lokacija. Za donošenje odluke o rezultatu detekcije može se izvršiti nezavisno ekstrahovanje bitova iz svake sekcije signala u kojoj je on umetnut, a zatim je moguće primijeniti princip većinskog glasanja ili na proizvoljni način ukombinovati rezultate.

Redudantno umetanje je posebno uspješna strategija u transformacionim domenima. Bilo koji efekat će izazvati promjenu gotovo svih odbiraka signala u izvornom domenu. Nasuprot tome, neki koeficijenti transformacije se neće izmijeniti uopšte ili će se izmijeniti neznatno, pa je moguća detekcija vodenog žiga iz ovih koeficijenata. Ovi koeficijenti se mogu identifikovati analitički ili empirijski, testiranjem na dovoljnom broju signala. Širenje spektra je jedna od tehnika redudantnog umetanja u frekvencijskom domenu. Umetanjem bitova vodenog žiga na različitim frekvencijama u signalu postiže se otpornost na različite vrste efekata, prevashodno na šum i filtriranja signala. Takođe, ova tehnika uvodi minimalne promjene u koeficijente, pa se time i kvalitet signala održava na valjanom nivou.

Prethodno opisane strategije fokusirane su na dostizanje robustnosti kroz izmjene u proceduri umetanja vodenih žigova. Sasvim drugačija strategija je da se u detektoru pokušaju prevazići efekti nastali nakon umetanja vodenih žigova. Ukoliko je poznato koji efekat se desio, u nekim situacijama može se rekonstruisati originalni signal. Na primjer, ukoliko je u audio signal uvedeno kašnjenje, ono se može izbrisati. Identifikacija grupe efekata koji su primijenjeni na signalom je veoma složen zadatak. Nekada je potrebno izvršiti iscrpnu pretagu testiranjem svih potencijalnih efekata i njihovih varijanti. Nakon definisanja relevantne grupe napada i opsega za njihove parametre, ispituje se svaka kombinacija vrijednosti parametara koja ne izaziva preveliko degradiranje kvaliteta signala. Za svaku od ovih kombinacija primjenjuju se inverzni efekti i dobijeni signal se predaje detektoru. Na primjer, ako se provjerava da li je audio signalu produženo trajanje, potrebno je vršiti skraćivanje signala u malim koracima, reda veličine nekoliko milisekundi, sve dok eventualno ne dođe do uspješne detekcije. Pored toga što ovo može biti izuzetno zahtjevan zadatak s aspekta računarske složenosti, testiranjem velikog broja signala povećava se šansa da detektor signal u kojem nije ugrađen vodeni žig označi kao votermarkovan. Čitav ovaj proces je nesumnjivo lakše izvesti sa informisanim detektorom, jer se upoređivanjem originala sa procesiranim signalom može zaključiti koji efekti su primijenjeni i kako ih invertovati.

2.3.2 Sigurnost vodenog žiga

Sigurnost je još rigoroznija kategorija za votermarking sisteme. Ona predstavlja sposobnost sistema da prevaziđe sve akcije koje vrše maliciozni korisnici sa ciljem da spriječe votermarking sistem u služenju svojoj namjeni. Značaj sigurnosti vodenih žigova zavisi od aplikacije. U nekim aplikacijama sigurnost je sporedna, jer se vodeni žigovi dodaju kako bi obogatili sadržaj, pa niko nema interes da ometa njihovo korišćenje. Primjer takve upotrebe vodenih žigova su *second screen* aplikacije, po-

menute u Sekciji 1.3. Nasuprot tome, votermarking sistemi koji se koriste u vojnoj komunikaciji moraju dostići najveći mogući stepen sigurnosti.

S obzirom na njihovu malicioznu prirodu, akcije koje su usmjerene ka opstrukciji votermarking sistema smatraju se napadima. Napadi se mogu, ali ne moraju, vršiti modifikacijama votermarkovanog signala. Robustnost je neophodan, ali i ne dovoljan uslov da bi se vodeni žig smatrao sigurnim. Siguran vodeni žig mora biti otporan i na operacije dizajnirane s posebnom namjerom da ga ugroze, a ne samo na uobičajene operacije obrade signala. Međutim, ukoliko vodeni žig nije otporan na standardne operacije nad signalom, on se ne može smatrati sigurnim, jer napadač može koristiti i te operacije ukoliko njima ispunjava svoj cilj. Dakle, prilikom dizajniranja robustnog sistema mora se voditi računa samo o regularnim operacijama nad signalom, dok se za postizanje sigurnosti razmatraju svi mogući potezi koje napadač može preduzeti kako bi poremetio korišćenje sistema. Sigurnost ne počiva samo na tehnikama koje se upotrebljavaju, već i u načinu korišćenja votermarking sistema. Iluzorno je ulagati napore u kreiranje sigurnog sistema ukoliko njegovim nespretnim korišćenjem dođe do neželjenih ishoda.

Napadi na vodeni žig se vrše kako bi se izvela neka od sljedeće tri procedure: neovlašćeno brisanje, neovlašćena detekcija ili neovlašćeno umetanje vodenog žiga. Važnost zaštite od svakog od ova tri napada zavisi od primjene sistema. Napadač nekada i ne mora sprovesti proceduru do kraja. Za ispunjenje njegovog cilja nekada je dovoljno da sprovede samo njen dio. Svaka od ove tri vrste napada iziskuje posebnu analizu i dizajniranje posebnih metode za sprečavanje.

Neovlašćeno brisanje je vrsta napada na votermarking sisteme koja se izvodi kada napadač ima namjeru da različitim operacijama nad votermarkovanim signalom dovede do toga da detektor ne može pronaći vodeni žig u njemu. Ukoliko je ovaj uslov zadovoljen, vodeni žig se smatra obrisanim. Pritom je poželjno da rezultujući signal bude što sličniji originalnom kako bi bio upotrebljiv, mada on ne mora biti očuvan u istoj mjeri kao i nakon umetanja. Brisanjem vodenog žiga iz signala se najozbiljnije ugrožava zaštita autorskih prava. Maliciozni korisnik, koji ima mogućnost da obriše vodene žigove, može napraviti nevotermarkovane kopije signala i nelegalno ih distribuirati i sticati dobit.

U nekim primjenama votermarkinga, poput one u internet telefoniji ili *second screen* aplikacija, šema za detekciju je javno dostupna i bilo ko može detektovati vodene žigove u signalima. Nasuprot tome, u mnogim drugim aplikacijama, najčešće iz razloga privatnosti, treba ograničiti prava detekcije i tu operaciju dozvoliti samo jednoj grupi korisnika. Najočigledniji oblik neautorizovane detekcije dešava se kada napadač pročita vodeni žig iz votermarkovanog signala. Tada bi, ukoliko je šema za

umetanje javno dostupna, maliciozni korisnik mogao koristiti nečiji vodeni žig da generiše sadržaj pod tuđim imenom i time sebi pribavlja korist ili narušava ugled osobe čiji je vodeni žig otuđen. Pored ovoga, blaže oblike neutorizovane detekcije imamo kada napadač može prepoznati da su signali označeni različitim vodenim žigovima, ali ne može otkriti sadržaj žigova, ili kada napadač prepoznaje da je neki signal votermarkovan, ali takođe ne može pročitati vodeni žig, niti ga razlikovati od drugih vodenih žigova. Sve ove situacije, iako različite destruktivnosti, mogu se negativno odraziti na sigurnost sistema i njegovu primjenljivost.

Neovlašćeno umetanje vodenog žiga prodrazumijeva proces prilikom kojeg napadač u signal ugrađuje svoj vodeni žig ili žig nekog drugog lica. Kreiranjem sopstvenog vodenog žiga i njegovim neovlašćenim umetanjem u signal, napadač može taj signal označiti kao njegovo vlasništvo i ubirati korist od toga. Dobavljanjem tuđeg vodenog žiga i njegovim nedozvoljenim umetanjem u odabrani signal napadač može proizvoljni multimedijalni sadržaj lažno predstaviti kao vlasništvo lica kome vodeni žig pripada. Napadač ovo može koristiti prilikom kreiranja dipfejkova i činiti različite povrede integriteta targetirane ličnosti.

Sistemi vodenog žiga čija je šema detekcije informisana su posebno osjetljivi na ovu vrstu napada. Jednostavnim oduzimanjem svog vodenog žiga od reprezentacije votermarkovanog signala napadač može doći do sopstvene verzije originalnog signala nosioca. Razlika između votermarkovanog signala i napadačeve verzije originalnog signala, koja se koristi u informisanoj detekciji, ima veliku korelaciju sa vodenim žigom napadača. U toj situaciji napadač može tvrditi vlasništvo nad votermarkovanim signalom jednako kao i njegov stvarni vlasnik, jer se njegov vodeni žig može detektovati u njemu.

Osnovni preduslov za stvaranje sigurnog sistema vodenog žiga je kreiranje procedure za umetanje koja nije invertibilna. Ukoliko je za proceduru umetanja nekog sistema moguće izvesti inverznu funkciju u realnom vremenu, onda se taj sistem može ugroziti na različite načine. Sama inverzna funkcija bi za proizvoljni votermarkovani signal generisala njegov original i umetnuti vodeni žig i time dovela do neautorizovane detekcije. Najbolji pristup za tretiranje ovog problema je upotreba informisanog umetača koji generiše kodiranu verziju vodenog žiga na osnovu signala nosioca. Na taj način se vodeni žig ne može rekonstruisati bez posjedovanja originalnog signala.

Čak i kada ne može u potpunosti invertovati operacije koje se vrše u sistemu vodenog žiga, napadač može prikupiti saznanja o samom vodenom žigu ili postupcima za njegovo umetanje i detekciju. On takođe može doći u posjed određenog broja primjeraka votermarkovanih signala. Ove informacije mogu pomoći napadaču

da projektuje efikasnije napade i ozbiljnije remeti korišćenje votermarking sistema. Na primjer, ako napadač pribavi nekoliko signala označenih istim vodenim žigom, njihovim upoređivanjem on može doći do saznanja o načinu na koji je taj vodeni žig umetnut i to znanje iskoristiti za neovlašćeno umetanje drugih vodenih žigova. Ukoliko napadač sakupi primjerke istog signala označenog različitim vodenim žigovima, na osnovu ovih primjeraka može se dobiti aproksimacija originalnog signala, odnosno obrisati vodeni žig. Ako napadač na raspolaganju ima parove votermarkovanih i originalnih signala, tada on može analizirati razlike između ovih primjeraka i doći do saznanja o slabostima sistema koja će mu pomoći da dizajnira efektivnije napade. Na primjer, napadač može primijetiti da se vodeni žig u signalima nalazi na visokim frekvencijama i eliminisati ga na jednostavan način primjenom niskopropusnog filtra. Posjedovanje detektora svakako može pomoći napadaču. Uzastopnim testiranjem svojih napada nad raspoloživim detektorom, on može lakše saznati da li neke operacije dovode do brisanja vodenog žiga iz datog signala, ali i naučiti nešto o samom procesu detekcije.

Ukoliko ne posjeduje nikakvo znanje o algoritmima za umetanje i detekciju i nema na raspolaganju alate poput detektora vodenog žiga, napadač se mora osloniti na opšte znanje o slabostima votermarking sistema. On može primijeniti operacije za koje je poznato da mogu eliminisati vodeni žig iz signala, kao što su dodavanje ili eliminacija šuma, odnosno filtriranje, kompresija signala, itd. U prvom redu su tu svakako efekti desinhronizacije. Oni su se pokazali kao najveća prepreka u dizajniranju votermarking sistema. U literaturi još uvijek ne postoji tehnika za koju se može neosporno tvrditi da je u potpunosti otporna na ove efekte. Napadač može vršiti permutacije ili brisanja odbiraka signala dok god se time osjetno ne remeti kvalitet. Većina šema za detekciju se oslanja na poravnanje signala nosioca i vodenog žiga. Vodeni žig se obično ugrađuje na fiksnim pozicijama u signalu. Ovi napadi remete to poravnanje i to je glavni razlog njihove problematičnosti. Štaviše, ovi napadi su nelinearne operacije, a teško je na adekvatan način rekonstruisati signal izmijenjen nelinearnim operacijama pomoću linearnih operacija, kojim se služe uobičajeni algoritmi za detekciju vodenog žiga. Efekti desinhronizacije za audio signale uvode vremensko skaliranje, kašnjenje ili pomjeranje visine tona. Za slike su to geometrijske transformacije, poput rotacije, translacije i skaliranja.

Neke tehnike [4, 16] se od efekata desinhronizacije brane ugrađivanjem posebnih sinhronizujućih kodova u signale. Dodavanjem ovih kodova u signal, a zatim i njihovom ekstrakcijom iz istog može se zaključiti pod dejstvom kojeg efekta desinhronizacije je dati signal. Dejstvo identifikovanog efekta ili grupa efekata se zatim invertuje, pa se obavlja detekcija vodenog žiga. Sinhronizujući kodovi se, umjesto po predefinisanoj šemi, mogu ugrađivati u istaknutim djelovima signala ili neposredno

prije ili poslije njih. Primjer ovih istaknutih tačaka su pikovi signala. Na ovaj način postiže se otpornost na pojedine efekte desinhronizacije, jer se lokacija vodenog žiga traži relativno u odnosu na te istaknute tačke, a ne na fiksnoj poziciji. U svemu ovome pomaže ukoliko prilikom detekcije na raspolaganju imamo i originalni signal, odnosno ukoliko koristimo informisani detektor.

Međutim, ovaj pristup ima brojne nedostatke. Prvi je što se ugrađivanjem dodatnih informacija u signal smanjuje broj bitova vodenog žiga koji se može ugraditi bez ugrožavanja kvaliteta signala. Takođe, ekstrakcija sinhronizujućeg koda postaje još jedna slaba tačka sistema. Različiti efekti mogu uticati i na ovaj kod i onemogućiti njegovu uspješnu ekstrakciju. Prema tome, sinhronizujući kodovi mogu poslužiti da ukažu da je došlo do desinhronizacije i spriječe pogrešnu detekciju. Ipak, oni ne predstavljaju sredstvo za dostizanje robustnosti u punom smislu. Ovaj neizbježni gubitak informacija u slučaju napada može nas spriječiti da otkrijemo da li su signali označeni vodenim žigom. Nemogućnost detektovanja vodenog žiga u tim situacijama, učinila bi aplikacije votermarking sistema, poput zaštite autorskih prava i mnogih drugih, neizvodljivim. Promjene koje bi unijeli ovi napadi mogle bi biti suptilne i neprimjetne ljudskom slušnom sistemu što bi omogućilo nesmetanu distribuciju i konzumiranje tog audio sadržaja.

Sinhronizujući kodovi imaju i sigurnosne nedostatke. Obično se isti sinhronizujući kodovi umeću u različite audio signale. Ovim se olakšava njihova detekcija. Međutim, takođe se omogućava napadaču da lakše otkrije ove kodove na osnovu skupa votermarkovanih signala i umanji sposobnost sistema da se suprotstavi efektima desinhronizacije. Takođe, ako napadač zna koje istaknute tačke se koriste za ugrađivanje sinhronizujućih kodova, on može namjerno targetirati te djelove signala prilikom dizajniranja napada i obrisati sinhronizujuće kodove iz signala. Najveća efikasnost u borbi protiv efekata desinhronizacije može se očekivati od sistema koji su u stanju da detektuju i invertuju ove efekte. Pored toga, tehnika redudantnog umetanja bitova vodenog žiga može pomoći u prevazilaženju napada kao što su brisanje i premutacija odbiraka. Širenjem vodenog žiga po čitavom spektru signala sistem se uspijeva zaštiti od nekih napada kojima se briše vodeni žig.

Za procedure umetanja i detekcije vodenih žigova postoje adekvatne analogije u kriptografiji. Umetanje se može poistovjetiti sa šifrovanjem, a detekcija sa dešifrovanjem. Stoga se problemi neovlašćenog umetanja i neovlašćene detekcije mogu rješavati primjenom kriptografskih tehnika. Od neovlašćenog detektovanja sistem se može braniti šifrovanjem vodenih žigova. Vodeni žig se može šifrovati tajnim ključem prije umetanja, a zatim i dešifrovati nakon detekcije. Tada i ukoliko napadač dođe do bitova koji su umetnuti u signal, on ih neće moći dešifrovati bez posjedovanja tajnog ključa i time otkriti originalni vodeni žig. Slaba tačka ovog pri-

stupa je to što je potrebno obezbijediti siguran način za razmjenu ključa između umetača i detektora kako ga napadač ne bi mogao saznati. Takođe, ovim metodom se ne može spriječiti neovlašćena detekcija šifrovane poruke, a nekada je i samo saznanje o prisustvu vodenog žiga u signalu korisna informacija napadaču koji želi da ga obriše.

U odbrani od neovlašćenog umetanja mogu se koristiti asimetrični kriptografski algoritmi. Tako se otklanja potreba za sigurnom razmjenom ključeva. Vlasnik digitalnog signala šifruje vodeni žig svojim privatnim ključem, a prilikom detekcije se koristi javni ključ za dešifrovanje. Dok god mu je privatni ključ nepoznat, napadač nije u stanju da ugradi u signal svoj vodeni žig, jer ga ne može ispravno šifrovati. Međutim, ovom metodom zaštite se ne može spriječiti umetanje tuđeg vodenog žiga u odabrani signal. Ukoliko napadač dođe u posjed tuđeg vodenog žiga šifrovanog privatnim ključem, on ga može u takvom, šifrovanom, obliku prenijeti i ugraditi u drugi signal. Zbog toga je potrebno vodeni žig proširiti digitalnim potpisom signala nosioca kako bi ugrađeni bitovi zavisili od signala nosioca, te ne bi mogli neprimjetno biti dodati u drugi signal. Na strani detekcije se može provjeriti da li detektovani bitovi odgovaraju signalu u koji su ugrađeni, izračunavanjem digitalnog potpisa votermarkovanog signala i njegovim upoređivanjem sa ekstrahovanim. Ova procedura mora biti pažljivo dizajnirana jer ugrađivanje vodenih žigova mijenja signal, a sa tim potencijalno i njegov digitalni potpis. Stoga je neophodno digitalni potpis izračunavati na osnovu djelova signala koji neće biti izmijenjeni umetanjem vodenih žigova. Nedostatak koji ostaje je što se i blagim modifikacijama signala u tim djelovima može drastično izmijeniti njegov digitalni potpis i onemogućiti detekcija.

2.3.3 Očuvanje kvaliteta signala

Tehnike za umetanje vodenih žigova u audio signale dizajniraju se s težnjom da promjene koje ta tehnika unese u signal nosilac budu što manje primjetne, odnosno čujne. Veoma je teško dostići potpunu neprimjetnost vodenih žigova. Kvalitet signala je potrebno održati na visokom nivou da se ne bi ugrozila vrijednost informacije koju on nosi ili negativno uticalo na profit koji on treba da proizvede zbog lošeg iskustva slušalaca, odnosno gledalaca. Očuvanje kvaliteta signala, nakon umetanja vodenog žiga, je sve teže postići usljed razvoja visoko kvalitetnih uređaja za reprodukciju zvuka, kojima se sve jasnije mogu primijetiti nesavršenosti u audio signalu.

Važno je napomenuti da cilj votermarking tehnika nije da maksimizuju kvalitet signala. Postoje posebne metode koje su namijenjene otklanjanju šumova i maksimizovanju kvaliteta signala, predložene u radovima [17–23]. Votermarking tehnikama se nastoji očuvati postojeći kvalitet signala, odnosno učiniti da se votermarkovani

signal što manje razlikuje od originalnog. Tako će vodeni žig biti sakriven na najbolji mogući način. Ako je originalni signal osjetno lošijeg kvaliteta od votermarkovanog, to može razotkriti vodeni žig. Ukoliko se može garantovati da niko nema pristup originalnom signalu, tada je dopustivo poboljšavati kvalitet koliko je to moguće, jer napadač ne može izvršiti poređenje dva signala. Nekada se može osloniti i na to da će efekti kojima votermarkovani signal podliježe dodatno sakriti vodeni žig. Na primjer, postojanje šuma može maskirati šum nastao dodavanjem vodenog žiga i učiniti ga teže primjetnim.

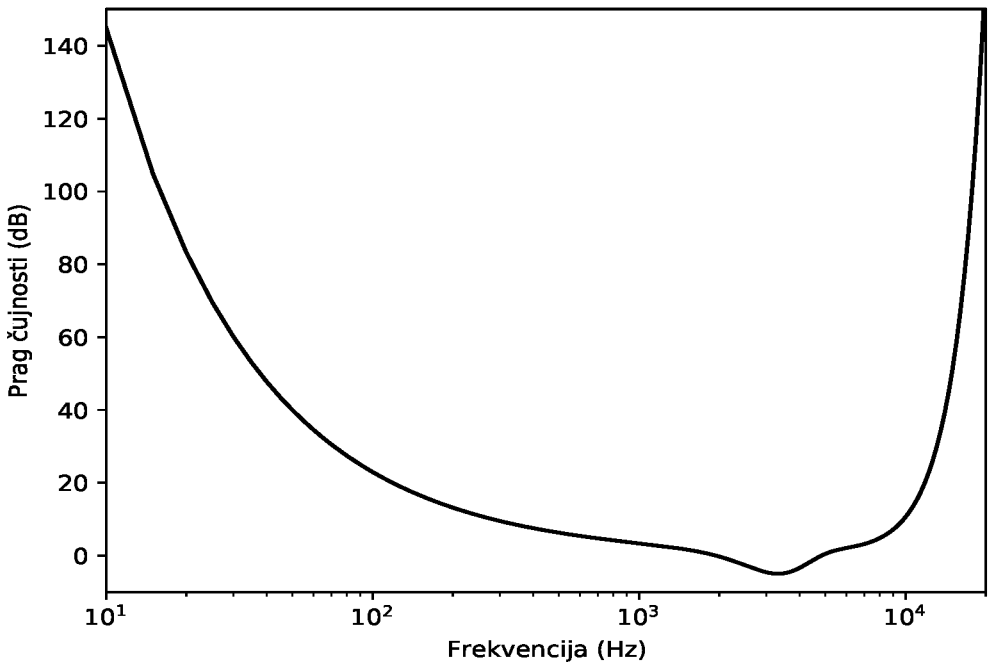
U kreiranju votermarking sistema istraživači su se služili svojstvima ljudskog slušnog i vizuelnog sistema kako bi na što bolji način sakrili vodene žigove u digitalne signale. Dizajnirani su različiti perceptivni modeli koji su integrisani u votermarking sisteme kako bi pospješili očuvanje kvaliteta signala tokom umetanja vodenih žigova. Perceptivni modeli numeričkim metodama pokušavaju simulirati ljudsku percepciju.

Dizajniranje modela ljudskog sluha i vida je veoma složeno, jer je teško aproksimirati ljudsku procjenu očuvanja kvaliteta signala. Ljudske procjene su suštinski subjektivne i mogu značajno varirati od osobe do osobe. Čak se procjene jedne osobe mogu mijenjati vremenom i zavisiti od različitih faktora, poput okruženja, raspoloženja, starosti, ostalih činilaca psihofizičkog stanja, itd. Pored toga, odgovor ljudskih perceptivnih sistema na različite stimuluse je prilično neujednačen i zavisi od nekoliko svojstava stimulusa. Na primjer, reakcija ljudskog slušnog sistema na određeni ton zavisi od njegove frekvencije, intenziteta, kao i od drugih tonova na susjednim frekvencijama i njihove jačine. Na Slici 4 iscrtan je minimalni prag čujnosti u zavisnosti od frekvencije f, dobijen na sljedeći način [24]:

$$p\check{c}(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 0.001\left(\frac{f}{1000}\right)^{4}. \qquad (12)$$

Vrijednost praga mjeri se u decibelima (dB) i predstavlja minimalni nivo zvuka potreban da prosječno ljudsko uho detektuje izolovani ton na frekvenciji f. Izolovani ton je sinusoida koja je određena svojom frekvencijom, amplitudom i fazom. Slika 4 pokazuje da je ljudsko uho najosjetljivije na tonove na frekvenciji oko 3 kHz, dok osjetljivost polako opada na nižim frekvencijama, a naglo opada za frekvencije iznad 10 kHz. Odgovori na različite frekvencije su veoma raznoliki, a pritom ovaj grafik ne oslikava cjelokupan model ljudskog slušnog sistema.

Prag čujnosti u jednakosti (12) izračunat je za referentni zvuk od 20μPaskala, koji odgovara zvuku intenziteta $0.98 \cdot 10^{-12}$ W/m^2, u okruženju sa standardnim atmosferskim pritiskom od 101.325 kPa, na temperaturi od 25 °C. Ovaj zvuk aproksimira najtiši zvuk koje mladi čovjek sa neoštećenim sluhom može detektovati na datoj frekvenciji. Prag čujnosti za zvuke većeg intenziteta se razlikuje, iako odnos



Slika 4: Prag čujnosti za ton od 20μPaskala u zavisnosti od frekvencije.

u zavisnosti od frekvencije ostaje sličan. Različite studije [24, 25] su pokazale da je ljudsko uho osjetljivije na promjene u zvucima sa većim intenzitetom, nego na promjene u zvucima slabijeg intenziteta. Ova zavisnost nije linearna jer se na nižim frekvencijama percepcija glasnosti mnogo brže mijenja, dok je na visočijim frekvencijama ton potrebno značajno pojačati kako bi ga ljudsko uho ocijenilo glasnijim.

Iako su frekvencija i intenzitet najvažnije karakteristike u percepciji audio signala, da bi perceptivni model bio kompletan, potrebno je uvrstiti sve tonove u zvučnom signalu i na odgovarajući način modelovati njihove međusobne uticaje. Ton određenog intenziteta, na određenoj frekvenciji može biti čujan, ako se posmatra izolovano, a takođe i potpuno neprimjetan ukoliko postoji ton na bliskoj frekvenciji koji je mnogo jačeg intenziteta. Na kraju se perceptivne procjene za svaki od tonova u signalu moraju ukombinovati na odgovarajući način.

Pomoću perceptivnih modela mogu se kreirati sofisticiranije šeme za umetanje vodenih žigova. Oni mogu biti sastavni dio enkodera vodenog žiga sa Slike 2 i određivati potrebne transformacije vodenog žiga koje će dovesti do njegovog uspješnijeg skrivanja u signalu. Na primjer, perceptivni modeli se mogu iskoristiti za odabir vrijednosti parametra α u jednakostima (1) i (2), kojim se kontroliše snaga vodenog žiga. Pomoću ovih modela je moguće definisati šeme umetanja koje sa različitim intenzitetom ugrađuju vodeni žig u različitim djelovima signala u zavisnosti

od njihovog sadržaja. Prigušivanjem jednih djelova vodenog žiga, a pojačavanjem drugih se on može bolje sakriti u signalu. Vodeni žig se može sa većim intenzitetom sakriti u djelovima signala u kojima ga je teže primijetiti. Primjer takvih djelova signala su audio segmenti u kojima imamo prisutvo različitih tonova. U djelovima signala u kojima bi vodeni žig bio jasno uočljiv, poput intervala tišine, on se ugrađuje sa manjim intenzitetom.

Perceptivni modeli definisani u Furijeovom ili vejvlet domenu [26] određuju koliko je ljudska percepcija osjetljiva u kom dijelu spektra. Na osnovu modela se može za svaki opseg frekvencija izračunati s kojim intenzitetom se u tom dijelu spektra može izvršiti umetanje vodenog žiga. Jednakost (2) tada dobija sljedeći oblik:

$$Y(k) = X(k) + \alpha(k)v(k), \tag{13}$$

gdje je α sada vektor težinskih koeficijenata kojim se kontroliše intenzitet dodavanja vodenog žiga u signal.

Primjena perceptivnog modela ne može uticati na detekciju vodenih žigova, ukoliko se i prilikom detekcije upotrijebi isti perceptivni model. Perceptivni model detektora može otkriti s kojim intenzitetom je vodeni žig ugrađen u kom dijelu signala, sve dok votermarkovani signal ima slične perceptivna svojstva kao i original. Ovo bi trebalo da važi, jer se votermarking šema kreira tako da što manje izmijeni signal. Za informisane detektore ovo ograničenje ne postoji, jer oni mogu perceptivni model primijeniti na originalnom signalu. Na ovaj način se promjene koje je ovaj model unio mogu invertovati, a zatim se detekcija može obaviti kao da perceptivni model nije ni korišten.

Detekcija naizgled može biti otežana ukoliko detektor ne pretpostavlja korišćenje perceptivnog modela tokom umetanja. Tada on mora biti robustan na efekte izazvane perceptivnim modelom. Efekti opisani u Sekciji 2.3.1 remete detekciju vodenog žiga. Međutim, svrha perceptivnog modela je da omogući da se vodeni žig ugrađuje s maksimalnim mogućim intenzitetom koji ne narušava kvalitet signala. Stoga bi korišćenje optimalnog perceptivnog modela trebalo da za fiksirani kvalitet signala rezultuje maksimalnim vrijednostima korelacije ugrađenog i odgovarajućeg referentnog vodenog žiga, a nikako da ometa detekciju.

2.3.4 Kapacitet

Vodeni žigovi se ugrađuju u digitalne signale prevashodno u cilju obilježavanja vlasništva. Za ispunjenje ovog zadatka veličina, odnosno broj bitova, vodenog žiga naizgled nije od presudne važnosti. Vlasništvo se može označiti i samo jednim bitom,

dok god taj bit jedinstveno određuje vlasnika i dok god ga detektor može prepoznati u signalu. Ipak, poželjno je da vodeni žigovi sadrže što više bitova. Veći broj bitova implicira i veći broj različitih vodenih žigova. Ovo za posljedicu ima da više vlasnika digitalnog sadržaja može koristiti isti votermarking sistem za označavanje svojine, što u krajnjem obezbjeđuje i više korisnika cjelokupnog sistema.

Mogućnost ugrađivanja vodenih žigova s većim brojem bitova važna je i s aspekata robustnosti i sigurnosti. Postoje tehnike koje koriste redudantne bitove kako bi mogle detektovati ili korigovati greške u vodenim žigovima i na taj način pospješile robustnost sistema. Takođe, u Sekciji 2.3.2 navedene su tehnike koje proširuju vodeni žig dodatnim informacijama kako bi se sistem zaštitio od nedozvoljenih operacija.

Broj bitova koje je sistem u stanju da ugradi u fiksiranom broju odbiraka signala naziva se kapacitet votermarking sistema. Povećanje kapaciteta nije jednostavan poduhvat. Ugrađivanje većeg broja bitova obično izuzetno negativno utiče na kvalitet signala. Bitovi se ne mogu sakriti tako da budu neprimjetni. Ukoliko bi se kvalitet održao na istom nivou kao i ranije, sistem bi vjerovatno neke od redudantnih lokacija morao iskoristiti za skrivanje novih bitova. Na ovaj način gubi se na robustnosti. Stoga je kapacitet, pored robustnosti i očuvanja kvaliteta, jedna od odlika votermarking sistema za koju se moraju praviti kompromisi kako bi se postigle optimalne performanse za datu aplikaicju.

2.3.5 Računska složenost

Digitalni votermarking sistemi su softveri, odnosno računarski programi, pa im se može razmatrati računska složenost. Računska složenost se odnosi na vrijeme potrebno sistemu da izvrši definisane zadatke i na količinu memorije koju on tom prilikom koristi. Kao i za sve računarske sisteme, poželjno je da votermarking sistemi budu što efikasniji.

Kao i za većinu ostalih kriterijuma, primjena diktira očekivanja i u pogledu računarske složenosti. U nekim aplikacijama votermarking sistem mora funkcionisati u realnom vremenu da bi se mogao koristiti, dok u drugim aplikacijama sistem može biti vrijedan čak iako mu je potrebno duže vrijeme za obavljanje zadataka. Na primjer, tokom monitoringa TV i radio stanica i umetač i detektor moraju biti maksimalno efikasni jer se sadržaj gotovo neprekidno generiše. Umetač ne smije usporavati emitovanje sadržaja, a detektor mora biti u stanju da prati emitovanje. Suprotno tome, prilikom provjere autentičnosti vrijeme nije ograničavajući faktor, niti prilikom ugrađivanja vodenih žigova, niti prilikom njihove detekcije. Ovi sistemi upotrebljavaju se u sudskim sporovima za dokazivanje vlasništva koji se pokreću

relativno rijetko i pritom je ishod detekcije mnogo važniji od vremena potrebnog da se ona sprovede.

Prostorna složenost predstavlja maksimalni memorijski prostor koji algoritam zauzima u jednom trenutku svog izvršavanja. Kako je memorija s vremenom postala lako dostupna, prostorna ograničenja nisu stroga kao vremenska.

3 Mjerila performansi

Ranije je pomenuto da se od sistema za umetanje vodenog žiga u digitalne audio signale zahtijeva da ispune više zadataka, pa stoga postoji i veći broj metoda kojima se procjenjuje koliko je uspješno riješen svaki od ovih izazova. Važnost svakog pojedinačnog mjerila kvaliteta zavisi od planirane namjene sistema vodenog žiga. Čak i interpretacija mjerila može varirati u zavisnosti od primjene. Mjerila su u međusobnoj sprezi, poboljšanjem jednog, narušavaju se ostala, pa je praktično nemoguće ostvariti idealne performanse po svim aspektima. Votermarking sistem kreira se sa ciljem da postigne najbolji kompromis performansi za odabranu aplikaciju. Uglavnom se postizanje otpornosti na napade, odnosno robustnost, i očuvanje kvaliteta audio signala smatraju osnovnim zadacima koje ovaj sistem treba da riješi na što bolji način. Prilikom poređenja različitih tehnika vodenog žiga, upravo ovim mjerilima daje se prioritet. Međutim, postoje i drugi kriterijumi po kojima se može mjeriti kvalitet sistema vodenog žiga i na osnovu kojih se mogu porediti različiti pristupi. Dizajniranje novih kriterijuma i mjera je predmet aktivnog istraživanja [27], pa može se očekivati da nova mjerila budu definisana u budućnosti. Međutim, standardizovan skup mjerila performansi bio bi svrsishodan. Iako još uvijek autori mogu birati mjerila za ocjenu performansi po sopstvenom nahođenju, u savremenoj literaturi se nametnuo jedan skup mjera kojima se najčešće procjenjuju i međusobno porede različite tehnike za umetanje vodenog žiga u audio signale. Taj skup najzastupljenijih mjera korišten je i u ovom radu. Sve ove mjere navedene su i opisane u nastavku poglavlja. Najprije su opisana mjerila za očuvanje kvaliteta signala, a zatim mjerila za detekciju. Pored procjenjivanja robustnosti i kvaliteta audio signala, mjereni su i kapacitet sistema, kao i prostorna i vremenska složenost, odnosno njihove praktične alternative.

3.1 Očuvanje kvaliteta audio signala

Da bi se moglo ocijeniti koliko dobro posmatrana tehnika skriva vodeni žig i izvršilo njeno poređenje sa drugim tehnikama potrebno je numerički kvantifikovati kvalitet signala nakon procedure umetanja. Kvalitet je jedna od najvažnijih odlika signala. Njegovo mjerenje je neophodno u gotovo svim scenarijima gdje signal podliježe obradi. Uspješnost i primjenljivost različitih tehnika za obradu signala veoma često zavisi i od toga koliko su signali degradirani nakon obrade. Posljedica toga je i nastanak mnoštva mjerila za ocjenu kvaliteta signala. Značajan broj ovih mjerila može se primjenjivati i kod sistema vodenog žiga, što nam daje mnogo veći izbor nego kod ostalih kriterijuma za ocjenu performansi ovih sistema. Međutim, sistemi

vodenog žiga nemaju za cilj da maksimizuju kvalitet signala. Od sistema vodenog žiga prevashodno se traži da votermarkovani signali budu što sličniji originalima, što ne mora nužno značiti da su i najboljeg mogućeg kvaliteta. Stoga su mjerila koja će biti pomenuta u ovoj sekciji orijentisana na poređenje originalnog i votermarkovanog signala i estimaciju njihove perceptivne udaljenosti, a ne na nezavisno ocjenjivanje kvaliteta pojedinačnog signala. Odlučeno je da se procjene kvaliteta signala nazivaju mjerilima a ne metrikama, jer ne ispunjavaju sva svojstva metrika. Neka od mjerila mogu uzimati negativne vrijednosti, neka nisu simetrična, dok neka ne zadovoljavaju nejednakost trougla.

Najbolje mjerilo kvaliteta audio signala je ljudsko uho. Važno je napomenuti da je ljudski slušni sistem značajno osjetljiviji od vizuelnog sistema, pa je samim tim i mnogo teže postići neprimjetnost vodenih žigova u audio signalima nego u slici ili videu. Slušni sistem je, zbog njegove specifičnosti i složenosti, gotovo nemoguće na potpuno vjerodostojan način reprodukovati računskim metodama, pa se do najboljih ocjena kvaliteta dolazi subjektivnim testovima. U tim testovima formira se grupa, poželjno obučenih, slušalaca, koji upoređuju originalne i obrađene signale i ocjenjuju njihov kvalitet na osnovu predefinisane skale. Najčešće se koristi MOS skala [28] po kojoj se signali ocjenjuju brojevima od 1 do 5. Međutim, ovakve testove je veoma teško izvesti u praksi. Formiranje relevantne grupe ocjenjivača iziskuje vrijeme i novac, a i za sprovođenje testova potrebno je više vremena, jer subjekti moraju preslušati i ocijeniti sve audio snimke iz korpusa. Takođe, zbog njihove subjektivne prirode, rezultate ovih testova teško je ponoviti. Iz ovih razloga se ipak pribjegava korišćenju objektivnih mjera kvaliteta [29].

Objektivne mjere stepen očuvanja kvaliteta procjenjuju numeričkim poređenjem originalnog i obrađenog signala. Vrijednosti za neke mjere izračunavaju se u vremenskom domenu, ali postoje i mjere koje su definisane u frekvencijskom domenu signala. Dizajnirane su s namjerom da na najbolji mogući način aproksimiraju subjektivne ocjene kvaliteta, odnosno ljudski slušni sistem. Zapravo, ove mjere ne moraju aproksimirati neku od subjektivnih skala za kvalitet, već je dovoljno da njihove vrijednosti budu u istom relativnom odnosu kao i vrijednosti u subjektivnim testovima. Korišćenjem objektivnih mjera kvaliteta dobija se na brzini i ekonomičnosti. Takođe je i poređenje različitih tehnika jednostavnije i nepristrasno.

Najčešće korištena mjera za ocjenu kvaliteta bilo koje vrste signala je odnos signal-šum (*engl. signal-to-noise ratio* - SNR). SNR mjeri snagu željenog signala u odnosu na šum, odnosno nepoželjni signal. Kada se vrši umetanje vodenog žiga teži se minimizovanju razlike između originalnog signala x i votermarkovanog signala y,

pa se SNR računa na sljedeći način:

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1}\left(x(n)-y(n)\right)^2}\right). \tag{14}$$

Međutim, ova metrika ima nekoliko ozbiljnih nedostataka kada se primjenjuje na audio signalima. Prvi nedostak je što jednako vrednuje sve segmente signala, i one u kojima ima sadržaja i one u kojima je tišina. Razlike signala u segmentima sa tišinom obično su mnogo manje, pa će u tim intervalima SNR imati izrazito negativne vrijednosti što će značajno poremetiti cjelokupnu procjenu. Ovaj nedostatak može se donekle ublažiti postavljanjem praga za jačinu signala i eliminacijom svih djelova signala ispod te vrijednosti [30]. Drugi način za ublažavanje uticaja tihih segmenata na vrijednost SNR mjere je pomjeranje logaritamske funkcije za 1, pa SNR dobija sljedeći oblik (*engl. shifted signal-to-noise ratio* - shSNR):

$$\mathrm{shSNR} = 10\log_{10}\left(1+\frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1}\left(x(n)-y(n)\right)^2}\right). \tag{15}$$

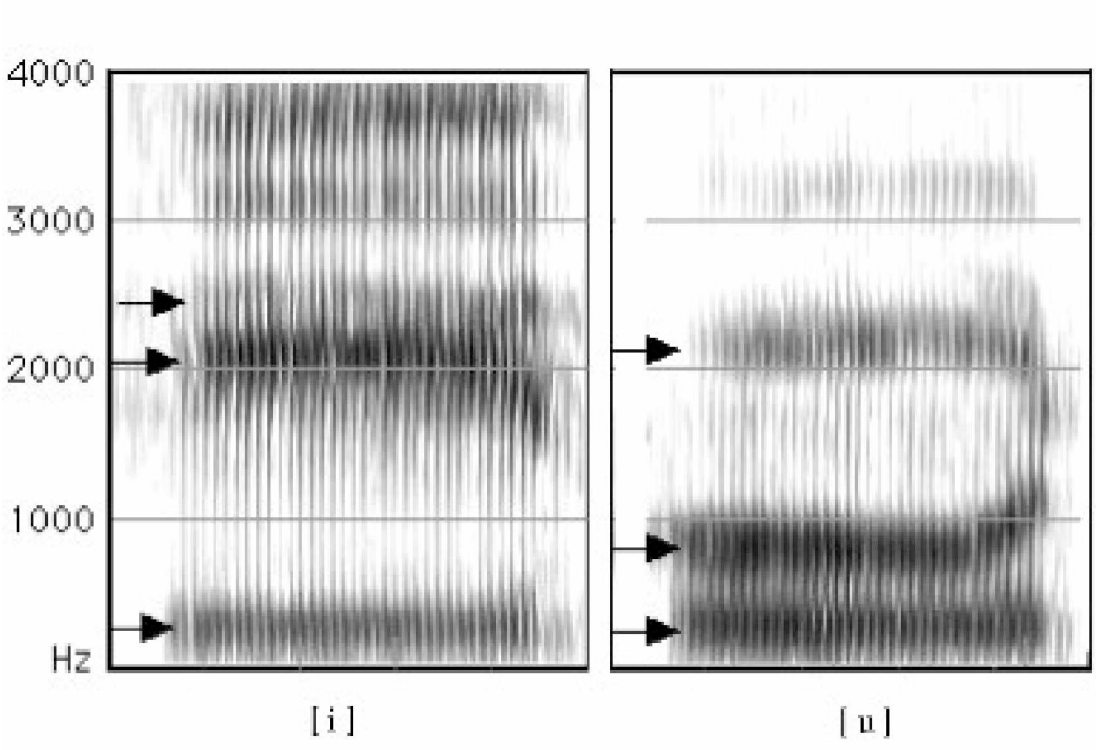
SNR mjera takođe sve frekvencijske komponente signala tretira ekvivalentno, iako je poznata činjenica da ljudski sluh nije podjednako osjetljiv na sve frekvencije. Da bi se prevazišao ovaj problem, SNR mjera definisana je u frekvencijskom domenu i proširena je sa težinskim koeficijentima (*engl. frequency weighted signal-to-noise ratio* - fwSNR) [31]:

$$\mathrm{fwSNR} = 10\cdot\frac{\sum_{k=0}^{K-1} T(k)\log_{10}\frac{X(k)^2}{\left(X(k)-Y(k)\right)^2}}{\sum_{k=0}^{K-1} T(k)}. \tag{16}$$

Koeficijenti $T(k)$ su unaprijed zadati. Svaki od njih određuje značaj odgovarajućeg pojasa frekvencija na kvalitet zvuka. Na osnovu različitih studija, predložena je nekolicina šema za dodjeljivanje vrijednosti ovim koeficijentima. Jedna od tih šema predložena je u [32].

Potencijalno najveći nedostatak SNR mjere je što zahtijeva da signali koji se porede budu u potpunosti vremenski usklađeni. Ukoliko bi signal pomjeren za nekoliko milisekundi uporedili sa njegovom originalnom verzijom, vrijednosti SNR mjere bile bi izrazito velike. Očigledno je da ovo nije u skladu sa načinom na koji ljudsko uho percipira zvuk, jer je po subjektivnim mjerilima uticaj kašnjenja na ocjenu kvaliteta zvuka zanemarljiv. Votermarking sistemi koji uvode kašnjenje bi korišćenjem ovog mjerila bili neopravdano diskriminisani, iako bi te promjene mogle biti potpuno nečujne.

Prednost SNR mjere i ostalih mjera zasnovanih na njoj je što su veoma jednostavne za realizaciju. Takođe, SNR se nalazi u osnovi većine mjera razumljivosti [33].



Slika 5: Spektrogrami samoglasnika „i" i „u" u engleskom jeziku [34]. Strelicama su označeni frekvencijski pikovi u oba spektrograma.

Razumljivost govora je, pored kvaliteta, još jedan važan atribut ove vrste signala. Stoga je važno i ovu mjeru uključiti u evaluaciju sistema koji procesiranju govorne signale, uključujući i sisteme vodenog žiga. Međutim, zbog pomenutih ograničenja neophodno je bilo predložiti mjere koje vjerodostojnije simuliraju procjenjivanje kvaliteta od strane ljudskog slušnog sistema. Uzima se u obzir da frekvencijska rezolucija ljudskog slušnog sistema nije uniformna, odnosno da ljudsko uho ne reaguje na svaku frekvenciju na identičan način, kao i da percipirana glasnost signala nelinearno zavisi od njegovog intenziteta i varira sa frekvencijom, kao što je objašnjeno u Sekciji 2.3.3.

Studijama ljudske percepcije zvuka pokazalo se da se najveća razlika u zvukovima osjeti kada se oni razlikuju po frekvencijskim pikovima (vrhovima) u spektru. Ove razlike naročito su naglašene pri izgovaranju pojedinih samoglasnika. Na Slici 5 su prikazani spektrogrami zvučnih signala nastalih pri izgovaranju samoglasnika „i" i „u", respektivno. Spektrogram je grafički prikaz spektralnog sadržaja signala u vremenu. Na spektrogramu se vrijeme prikazuje duž horizontalne ose, frekvencije na vertikalnoj osi, a intenzitet boje ili nijansa predstavlja jačinu frekvencija u datom vremenskom trenutku. Na spektrogramima sa Slike 5 jasno se uočavaju razlike u frekvencijskim vrhovima dva zvučna signala. U radu [35] predložena je mjera, mo-

tivisana ovim saznanjima, koja je zasnovana na spektralnom nagibu (*engl. spectral slope*). Ovom mjerom se više naglašava razlika u lokacijama na kojima se nalaze spektralni vrhovi, a ignorišu se njihov intenzitet i ostale odlike. Izračunavanje započinje određivanjem spektralnog nagiba za svaki pojas frekvencija:

$$S_x(k) = X(k+1) - X(k). \tag{17}$$

Ista vrijednost računa se i za koeficijente procesiranog, odnosno u ovom slučaju votermarkovanog signala Y. Na kraju se razlike u spektralnim nagibima signala x i y sumiraju, sa težinama, kako bi se ocijenila njihova sličnost:

$$SS = \sum_k T(k)(S_x(k) - S_y(k))^2. \tag{18}$$

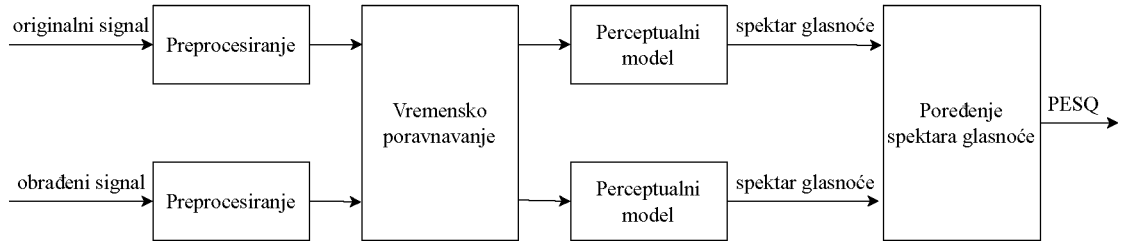
Izračunavanje ponderisanog spektralnog nagiba slično je izračunavanju ponderisanih SNR mjera. Međutim, za razliku od tih mjera, ponderisani spektralni nagib naglašava razlike u lokacijama spektralnih vrhova, a ostale spektralne detalje zanemaruje. SNR mjere, kao što je već navedeno, potpuno ekvivalentno tretiraju sve razlike u signalima.

Mjera ponderisanog spektralnog nagiba je prva u kojoj su uključena svojstva ljudskog slušnog sistema. Njenim daljim unapređivanjem nastale su naprednije mjere za ocjenu preceptualne sličnosti audio signala. U ovim mjerama uvrštene su i druge osobenosti ljudskog sluha, poput bolje sposobnosti da diskriminiše niže od viših frekvencija, nelinearnog variranja percipirane glasnoće s frekvencijom, itd. Prilikom izračunavanja ovih mjerila ljudski slušni sistem modeluje se kao niz transformacija zvučnih signala. Ovim transformacijama se originalni i obrađeni audio signal prenose u drugi domen u kojem se računa distanca između njih. Barkova mjera spektralne distorzije (*engl. Bark spectral distortion measure* - BSD) [36] predstavlja jednu od najistaknutijih perceptivnih mjera razlike u audio signalima. Ona se izračunava nad tzv. spektrima glasnoće audio signala $\mathcal{L}_x$ i $\mathcal{L}_y$:

$$BSD = \sum_k T(k)(\mathcal{L}_x(k) - \mathcal{L}_y(k))^2. \tag{19}$$

Težinski koeficijenti $T(k)$ uzimaju vrijednosti 0 ili 1, odnosno određuju da li će se djelovi spektra glasnoće uključiti u izračunavanje vrijednosti BSD mjere ili ne. Smatra se da, ukoliko je na nekom mjestu razlika između spektra glasnoće originalnog i spektra glasnoće obrađenog signala ispod datog praga, tada ova razlika neće biti čujna i ne mora se uvrštavati u ocjenu. Dakle, pravilo za određivanje vrijednosti koeficijenata $T(k)$ definisano je sljedećom jednakošću:

$$T(k) = \begin{cases} 0, & |\mathcal{L}_x(k) - \mathcal{L}_y(k)| < \mathcal{L}_\tau \\ 1, & |\mathcal{L}_x(k) - \mathcal{L}_y(k)| \geq \mathcal{L}_\tau. \end{cases} \tag{20}$$



Slika 6: Blok šema za izračunavanje vrijednosti PESQ mjere.

Vrijednost praga $\mathcal{L}_\tau$ može se računati procedurom opisanom u [37].

Prethodno pomenute mjere, iako uključuju perceptivne modele, i dalje nisu pogodne za korišćenje u situacijama kada je u procesiranom signalu uvedeno kašnjenje ili kada su mu obrisani ili dodati neki odbirci, tj. kada originalni i obrađeni signal nisu vremenski poravnati. Tokom 1990-ih godina predloženo je nekoliko mjera sa fokusom na prevazilaženju ovog nedostatka. Ovdje izdvajamo PESQ[38] (*engl. perceptual evaluation of speech quality*) mjeru za ocjenu kvaliteta govornih signala. PESQ mjera je odabrana od strane Telekomunikacionog sektora za standardizaciju u okviru Međunarodne unije za telekomunikacije - MTU (*engl. International Telecommunications Union*) kao preporučena mjera za ocjenu kvaliteta govornih i time je zamijenila do tada aktuelnu PSQM[39] (*engl. perceptual speech quality measure*). U opštijem slučaju, za ocjenu kvaliteta audio signala koristi se srodna PEAQ[40] (*engl. perceptual evaluation of audio quality*) mjera.

Blok šema PESQ prikazana je na Slici 6. Originalni i obrađeni signal prolaze kroz nekoliko koraka obrade prije samog poređenja. Prvi korak je skaliranje intenziteta originalnog i obrađenog signala tako da se oni dovedu na standardni nivo za slušanje. Nakon toga se vrši gruba procjena mimoilaženja dva signala u vremenu, korelacijom njihovih anvelopa. Procijenjena kašnjenja se zatim koriste kako bi se signal podijelio na više segmenata. Nad ovim segmentima se potom traži poravnanje sa većom rezolucijom kako bi se odredilo tačno kašnjenje i kako bi se signali mogli što preciznije uskladiti u vremenu. Usklađeni signali prolaze kroz psihoakustični model, gdje se, nizom transformacija, izračunavaju spektri glasnoće, slično kao kod BSD mjere. Na kraju se porede ovi spektri kako bi se izračunala distanca signala. Rezultat ove procedure je vrijednost iz opsega od -0.5 do 4.5.

PESQ se od BSD mjere razlikuje po tome što ne tretira ekvivalentno pozitivne i negativne razlike između spektara glasnoće $\mathcal{L}_x$ i $\mathcal{L}_y$. U slučaju kada je $\mathcal{L}_x(k) < \mathcal{L}_y(k)$, to indicira da je u spektar obrađenog signala dodata neka komponenta ili šum. Suprotno tome, ukoliko je $\mathcal{L}_x(k) > \mathcal{L}_y(k)$, smatra se da je frekvencijska komponenta uklonjena iz originalnog signala ili je značajno ublažena. Dodavanje nove komponente se lakše primjećuje od brisanja ili ublažavanja komponente zbog efekta ma-

skiranja, pa je potrebno na različite načine vrednovati ove dvije pojave prilikom upoređivanja dva signala kako bi se dobila što preciznija ocjena. Ovo svojstvo PESQ mjeru čini nesimetričnom.

Ovdje je dat opšti pregled koraka koje je potrebno sprovesti kako bi se izračunala vrijednost PESQ mjere. Njihov detaljan opis zahtijevao bi uvođenje novih koncepata iz obrade signala i psihoakustike koji nisu relevantni za nastavak ove teze. Svi detalji mogu se pronaći u navedenoj referenci koja sadrži sveobuhvatan pregled principa obrade signala i psihoakustike koji su u osnovi ove mjere.

3.2 Stopa grešaka detektora

Komponenta votermarking sistema za detekciju vodenih žigova i ekstrakciju njihovih bitova može napraviti različite vrste grešaka. U cilju analize kvaliteta cjelokupnog sistema, potrebno je procijeniti vjerovatnoću tih grešaka i ispitati da li one ispunjavaju postavljene zahtjeve, odnosno da li su ispod definisanih granica. Iako ovdje govorimo o greškama detektora, uloga umetača je podjednako važna u njihovom minimizovanju. Šemu za umetanje treba dizajnirati tako da detekcija žigova bude što efikasnija i robustnija.

Nekada se može desiti da detektor u signalu koji nije votemarkovan greškom detektuje vodeni žig. Stopa ove vrste greške (*engl. false positive rate* - FPR) posmatra se kao posebna mjera performansi votermarking sistema. Predstavlja se kao vjerovatnoća da će diskriminator u proizvoljnom nevotermarkovanom signalu pronaći vodeni žig. Empirijski se izračunava kao procenat nevotermarkovanih signala u kojima je pogrešno detektovan vodeni žig u odnosu na ukupan broj testiranih nevotermarkovanih signala:

$$\mathrm{FPR} = \frac{|\{z|\textsc{Detekcija}(z, \mathcal{V}, \tau) \neq \emptyset \wedge z \in \mathcal{W}'\}|}{|\{z|z \in \mathcal{W}'\}|} \cdot 100\%, \tag{21}$$

gdje je $\mathcal{W}'$ skup nevotermarkovanih signala, $\mathcal{V}$ i τ ranije definisani skup vodenih žigova, odnosno prag za detekciju, a Detekcija funkcija iz Algoritma 1.

Analogno se može posmatrati i situacija u kojoj detektor u votermarkovanom signalu ne pronalazi vodeni žig (*engl. false negative rate* - FNR):

$$\mathrm{FNR} = \frac{|\{z|\textsc{Detekcija}(z, \mathcal{V}, \tau) = \emptyset \wedge z \in \mathcal{W}\}|}{|\{z|z \in \mathcal{W}\}|} \cdot 100\%, \tag{22}$$

a $\mathcal{W}$ je skup votermarkovanih signala.

Prethodno pomenute vrste grešaka generiše diskriminator komponenta unutar detektora. Ukoliko diskriminator komponenta uspješno detektuje postojanje vode-

nog žiga u signalu, moguće je da se u procesu ekstrakcije žiga napravi greška na jednom ili više bitova.

Neophodne ili targetirane visine stopa ovih grešaka zavise od primjene votermarking sistema. U sistemima u kojima se detektor rijetko koristi, poput dokazivanja autentičnosti, dopustivo je da ove stope budu nešto veće. Ipak, u velikom broju aplikacija se detektor intenzivno koristi. Intenzivno korišćenje detektora imamo u sistemima koji služe za detekciju potencijalnih povreda autorskih prava, monitoringu TV i radio stanica, internet komunikaciji, itd. Tada je neophodno da stopa pogrešno detektovanih vodenih žigova i stopa pogrešno ekstrahovanih vodenih žigova budu infitezimalne. U suprotnom bi se, zbog količine podataka koji prolaze ovim sistemima, nagomilao veliki broj grešaka, što bi izazvalo ozbiljne probleme.

U nekim aplikacijama ove greške su od različite važnosti. Na primjer, u sistemu koji se koristi za potvrdu vlasništva nemogućnost detektovanja vodenog žiga u signalu može izazvati ozbiljne posljedice za vlasnika multimedijalnog sadržaja, jer on neće biti u mogućnosti da dokaže svoje vlasništvo nad podacima. Detektovanje pogrešnog vodenog žiga u signalu imalo bi istu posljedicu. Kako ovaj sistem na ulazu detekcije ne dobija nevotermarkovane signale, detektovanje vodenog žiga u nevotermarkovanom signalu se neće nikada desiti. Sličnu situaciju imamo i kod sistema koji kontrolišu nelegalnu distribuciju autorskog sadržaja. U tim sistemima neprepoznavanje vodenog žiga u votermarkovanom signalu izaziva mnogo ozbiljnije posljedice nego kada se desi lažni alarm, odnosno kada sistem prepozna vodeni žig unutar tuđeg ili nevotermarkovanog signala. U prvom slučaju autor ne biva obaviješten o nedozvoljenom korišćenju njegovog djela i time mu se nanosi šteta. U drugom slučaju se jednostavnom provjerom signala može zaključiti da ipak nije došlo do povrede autorskih prava. Zarad boljeg korisničkog iskustva poželjno je i da se ovaj tip greške što rjeđe dešava, ali su one manje nezgodne. Nasuprot tome, ukoliko bi votermarking sistem falsifikovani signal (dipfejk) označio kao autentičan, odnosno prepoznao vodeni žig u njemu, takva greška bi mogla izazvati ozbiljne negativne posljedice i diskreditovati ličnosti kojima bi se vlasništvo nad signalom pripisalo.

Greške diskriminatora su povezane. Promjenom vrijednosti praga detekcije τ smanjuje se stopa jednog tipa greške, a druga se povećava. Postavljanjem praga za detekciju na minimum, možemo garantovati da se nikada neće desiti greška da detektoru promakne votermarkovani signal. Međutim, na taj način će detektor i veliki broj nevotermarkovanih signala označiti kao votermarkovane. Iscrtavanjem ROC krive (*engl. receiver operating characteristic curve*) u zavisnosti od vrijednosti parametra τ može se zaključiti koja vrijednost praga detekcije vodi do optimalnih performansi.

Greške u ekstrakciji bitova vodenog žiga možemo donekle nadomjestiti primjenom kodova za detekciju i korekciju grešaka. Vodeni žig se kodira u jednom od ovih kodova i kao takav se ugrađuje u signal. Postoji veliki broj ovakvih kodova [41], a izbor odgovarajućeg zavisi od očekivanog tipa grešaka i računarskih ograničenja. Ukoliko sistem podliježe velikom broju efekata desinhronizacije koji mogu obrisati ili dodati uzastopne grupe odbiraka signala, tada se očekuje da će detektor praviti lavinske greške. Lavinske greške su greške koje se javljaju na uzastopnim nizovima bitova. Za ublažavanje ove vrste grešaka, najpogodniji su ciklični kodovi. U slučaju da su greške na nasumičnim mjestima, podesnije je koristiti blok kodove. Ovim pristupom se može otkloniti jedan udio grešaka, jer kodovi za detekciju i korekciju grešaka imaju ograničenja u pogledu broja grešaka koje mogu detektovati i korigovati. Ukoliko broj grešaka prevazilazi mogućnosti koda, one se neće moći razriješiti. Negativna strana korišćenja kodova je što oni zahtijevaju prenos redudantnih bitova. Votermarking sistem mora i ove bitove sakriti u signalu. Na taj način se smanjuje količina informacija, odnosno broj vodenih žigova koje sistem može ugraditi u signal u jedinici vremena, pa se i u ovom pogledu moraju praviti kompromisi.

Greške u ekstrakciji bitova se najčešće mjere procentom, odnosno stopom, pogrešno detektovanih bitova vodenog žiga u posmatranom vremenskom intervalu (*engl. bit error rate* - BER). Ova vrijednost računa se jednostavnim poređenjem bitova originalnog vodenog žiga $w = (w(1), w(2), ..., w(L_w))$, sa detektovanim vodenim žigom $\hat{w} = (\hat{w}(1), \hat{w}(2), ..., \hat{w}(L_w))$

$$\text{BER} = \frac{|\{w(i) \neq \hat{w}(i) | i = 1, 2, ..., L_w\}|}{L_w} \cdot 100\% \tag{23}$$

Osim BER mjere, u literaturi se za mjerenje performansi detektora sporadično koristi i normalizovana korelacija, definisana u (6). Što je prosječna korelacija votermarkovanog signala i odgovarajućeg vodenog žiga veća, to se detektor može smatrati uspješnijim, a ujedno i robustnijim.

Stopa grešaka detektora se očekivano povećava ako signal podlegne različitim obradama ili napadima od strane zlonamjernih korisnika. U takvim situacijama ispituju se robustnost i sigurnost votermarking sistema. Detekciona statistike mogu se koristiti kao mjerila performansi i u ovim situacijama. Vrijednosti iz jednakosti (21), (22) i (23) izračunavaju se na potpuno identičan način, s tim što se kao testni skupovi uzimaju skupovi obrađenih, odnosno napadnutih votermarkovanih i nevotermarkovanih signala. Obično je najbolje svaki efekat posmatrati nezavisno i procjenjivati robustnost stopama grešaka detektora kada je signal pod dejstvom posmatranog efekta. Takođe je moguće procjenjivati robustnost i za proizvoljne kombinacije efekata. Dodatni način za estimaciju robustnosti i sigurnosti sistema vodenog žiga je

ispitivanje minimalnog intenziteta audio efekata potrebnih da se onemogući uspješna detekcija.

3.3 Kapacitet audio votermarking sistema

Kapacitet predstavlja broj bitova vodenog žiga koji se mogu ugraditi u posmatrani signal u jedinici vremena. Njegove vrijednosti date su u bitovima po sekundi (bps). Većina tehnika vodenog žiga dizajnirane su tako da teoretski mogu ugraditi proizvoljan broj bitova u signale. Ipak, s povećanjem broja bitova vodenog žiga drastično se degradiraju robustnost i kvalitet signala, pa je potrebno napraviti kompromise. Kako robustnost i kvalitet signala imaju prednost u odnosu na kapacitet, obično se kapacitet fiksira na određenu vrijednost, pa se procjenjuju robustnost i kvalitet signala za tu vrijednost kapaciteta. Na ovaj način dobija se najpravednije poređenje različitih tehnika.

Nekada se votermarking tehnikama vrši umetanje sekundarnih bitova u signal nosilac. Ovi bitovi mogu se koristiti za neutralisanje efekata desinhronizacije, korekciju grešaka u detektovanim bitovima vodenog žiga i druge prakse kojima se smanjuje stopa grešaka detektora. Međutim, oni se ne uvrštavaju u kapacitet posmatranog sistema.

3.4 Vremenska i prostorna složenost

Vremenska složenost se procjenjuje brojem operacija koje algoritam treba da izvrši u zavisnosti od dimenzije ulaza. Najčešće se zapisuje u asimptotskoj $\mathcal{O}$ notaciji koja opisuje gornju granicu broja operacija potrebnog za izvršavanje algoritma. Za funkciju $f(n)$ kažemo da asimptotski ne raste brže od funkcije $g(n)$, što označavamo sa $f(n) = \mathcal{O}(g(n))$, ako $\exists n_0 \in \mathbb{N} \wedge \exists c > 0 \in \mathbb{R}$ takvo da $\forall n > n_0 \ f(n) \leq cg(n)$. U ovom kontekstu $f(n)$ je broj koraka algoritma za dimenziju ulaza n, a $\mathcal{O}(g(n))$ je procjena njegove vremenske složenosti. Važna svojstva $\mathcal{O}$ notacije:

1. $\mathcal{O}(cg(n)) = \mathcal{O}(g)$ za neko $c \in \mathbb{R}\ c > 0$
2. $\mathcal{O}(g_1(n)) + O(g_2(n)) + ... + O(g_k(n)) = \mathcal{O}(\max\{g_1(n), g_2(n), ..., g_k(n)\})$
3. $\mathcal{O}(g(n))\mathcal{O}(h(n)) = \mathcal{O}(g(n)h(n))$

Pored vremenske složenosti, koja je teorijski koncept, obično se, empirijskim putem, izračunava i vrijeme izvršavanja algoritma na odgovarajućoj hardverskoj platformi, sa datom dimenzijom ulaza. Ova informacija je mnogo važnija u praksi

prilikom procjenjivanja da li se isplati posmatranu tehniku umetanja vodenih žigova uvoditi u upotrebu. Prostorna složenost se takođe može teorijski procjenjivati O notacijom, kao funkcija koja predstavlja potrebnu količinu memorije u zavisnosti od dimenzije ulaza. Empirijski se procjenjuje očitavanjem memorije koju algoritam zazume prilikom izvršavanja.

4 Tradicionalne votermarking tehnike

Prve tehnike za umetanje vodenog žiga u digitalne signale pojavile su se tokom devedesetih godina prošlog vijeka. Prvi poznati rad u ovoj oblasti je [42]. Prekretnica za dalji razvoj i intenziviranje istraživanja bio je rad [43], u kojem je dat predlog korišćenja tehnika zasnovanih na širenju spektra za ugrađivanje vodenih žigova u slike koje su se mogle uopštiti za audio, video i druge multimedijalne podatke. U narednom periodu razvijeno je mnoštvo tehnika za ugrađivanje vodenih žigova u različite tipove digitalnih signala, ali prevashodno za slike [44–50]. Votermarking digitalnih slika je najistraživaniji dio ove naučne oblasti. Neke ideje pozajmljene su odavde i prilagođene, a zatim i primijenjene na drugim vrstama digitalnih signala.

Zbog specifičnosti ljudskog slušnog sistema i za votermarking audio signala razvijeno je mnoštvo originalnih tehnika. Raznovrsnost osmišljenih tehnika ukazuje i na složenost problema koji se moraju riješiti prilikom dizajniranja sistema za umetanje vodenih žigova u audio signale. Pristupi se mogu kategorizovati na različite načine. Ključni način za podjelu je na osnovu vrste domena u kojem se ugrađuje vodeni žig. U tom pogledu razlikujemo tehnike za umetanje vodenog žiga u vremenskom domenu i tehnike za umetanje vodenog žiga u transformacionom domenu. Transformacioni domen se odnosi na domen neke od ortogonalnih tranformacija signala, poput DFT, DCT ili DWT. Metode za ugrađivanje vodenih žigova na ova dva načina predstavljene su generalizovanim jednakostima (1) i (2) u Sekciji 2.1. U svim kredibilnim pristupima u literaturi detekcija vodenih žigova vrši se u istom domenu kao i umetanje, pa je podjela po domenu za detekciju suvišna. Pored podjele na osnovu domena, votermarking pristupi se mogu razlikovati i po drugim osnovama. U Sekciji 2 pomenuta je podjela tehnika za umetanje i detekciju na informisane i neinformisane. Međutim, ova podjela ima sve manji praktični značaj jer većina modernih tehnika vrši informisano umetanje i neinformisanu detekciju vodenih žigova. Votermarking pristupi se dijele i na osnovu toga da li se za umetanje vodenog žiga koristi aditivni ili multiplikativni model. Jednakosti (1) i (2) predstavljaju aditivni model ugrađivanja vodenog žiga. Multiplikativni modeli mogu se predstaviti pomoću aditivnih. Razlika se svodi na to da li šema umetanja zavisi od signala nosioca ili ne. Kako je većina votermarking šema za umetanje informisana, ni ova podjela nije od suštinske važnosti.

Metode mašinskog učenja odavno se koriste u sistemima vodenog žiga. Prvobitno su primjenjivane u fazi detekcije [51, 52] tako što je šema sa precizno propisanim skupom pravila zamjenjena detektorom koji je algoritmom mašinskog učenja obučavan da prepoznaje bitove vodenog žiga ugrađene u signalu. Kasnije su ovi prin-

cipi prenijeti i na komponentu za umetanje. Tehnika [53] koristi mašinsko učenje za optimizaciju vrijednosti parametara procedure umetanja.

Nakon revitalizacije oblasti dubokog učenja, istraživačka zajednica intenzivno se posvetila primjeni tih tehnika u votermarkingu slika, što je rezultovalo brojnim publikacijama. U procesu razvoja ovih sistema pojavili su se različiti izazovi. Jedan od početnih radova na ovu temu [54] morao se ograničiti na neslijepu detekciju. Autori [55] zadržali su se na obučavanju neuronske mreže za detekciju vodenih žigova. Votermarking sistemi koji koriste neuronske mreže za izvođenje svih koraka obrade, od početka do kraja, koncipirani su u kasnijim radovima [56–61]. Međutim, ovi pristupi nisu direktno primjenljivi na zvuk iz više razloga. Audio signali generalno imaju manju redundantnost u poređenju sa slikama, što ostavlja manje prostora za ugrađivanje dodatnih podataka. Čak i sitne promjene u zvuku mogu se lako opaziti, pa je stoga potrebno posvetiti dodatnu pažnju očuvanju kvaliteta. Osjetljivost na promjene otežava i postizanje robustnosti, jer se vodeni žigovi ne mogu ugraditi sa velikim intenzitetom. Zbog ovih fundamentalnih razlika, tehnike vodenog žiga za audio signale moraju biti posebno dizajnirane kako bi se riješili izazovi specifični za ovu vrstu signala. Pretpostavljamo da su pobrojani izazovi uticali na to da još uvijek nije konstruisana šema koja je u potpunosti zasnovana na dubokom učenju, a da ispunjava sve zahtjeve audio votermarking sistema.

Zbog sveprisutne upotrebe modela dubokog učenja, javila se potreba za zaštitom autorskih prava njihovih tvoraca. Ideja votermarkinga dubokih modela ispostavila se kao izazovan i zanimljiv koncept. Obrađena je u člancima [62–66].

U nastavku poglavlja dat je detaljan pregled tehnika za ugrađivanje vodenih žigova u digitalne audio signale. Tehnike su podijeljene u dvije grupe, prema vrsti domena u kojem se vrši umetanje vodenog žiga. U svakoj od grupa izdvojene su i opisane votermarking metode karakteristične po tehnikama obrade signala koje se u njima koriste i navedeni su najreleventniji naučni radovi vezani za njih. Prilikom opisivanja određene tehnike, jednakosti (1) i (2) su, na odgovarajući način, modifikovane tako da preciznije opisuju posmatranu tehniku.

4.1 Tehnike za ugrađivanje vodenog žiga u vremenskom domenu

Umetanje vodenog žiga u vremenskom domenu vrši se operacijama direktno nad odbircima audio signala. Tehnike dizajnirane u ovom domenu su obično efikasne i jednostavne za realizaciju. Najjednostavniji način za skrivanje bitova vodenog žiga je u najmanje značajnim bitovima odbiraka signala (*engl. least significant bit encoding*

- LSB) [67, 68]. Ova šema sprovodi se tako što se izvrši zamjena najmanje važnih bitova u odbircima signala sa bitovima vodenog žiga. Detekcija je jednostavna, jer je detektoru poznato sa kojih pozicija treba da pročita skrivene bitove. Ove tehnike imaju veliki kapacitet, odnosno mogu umetnuti veliki broj bitova vodenog žiga u jedinici vremena. Međutim, njihov veliki nedostatak je slaba robustnost i sigurnost. Pojava čak i blagog šuma u signalu ili bilo kojeg drugog efekta u najvećoj mjeri remeti upravo najmanje važne bitove odbiraka i time onemogućava ispravnu detekciju vodenog žiga. Robustnost se može poboljšati korišćenjem većeg broja bitova odbirka za umetanje jednog bita vodenog žiga, ali ovo negativno utiče na očuvanje kvaliteta signala. Dodatno, ukoliko je napadaču poznato da su biti vodenog žiga skriveni u najmanje važnim bitovima odbiraka, on lako može otkriti ili uništiti poruku. Da bi se poboljšala sigurnost, biti vodenog žiga mogu se šifrovati. Međutim, ovo ne može spriječiti napadača da proizvoljno izmijeni najmanje značajne bitove signala i time obriše vodeni žig. Napadač, takođe, može veoma jednostavno izvršiti neautorizovano umetanje, kopiranjem najmanje značajnih bitova iz votermarkovanog signala u neki drugi. Iz pomenutih razloga, tehnike zasnovane isključivo na ovom principu nisu često u praktičnoj upotrebi. Mogu se koristiti za krhke vodene žigove ili u kombinaciji sa drugim tehnikama kojima se ublažavaju njihovi nedostaci.

Šema za umetanje vodenog žiga u jednoj od najistaknutijih tehnika u vremenskom domenu [69] može se predstaviti sljedećom jednakošću:

$$y(n) = x(n) + \alpha \left(|x(n)|v(n)\right) * h_L, \tag{24}$$

gdje h_L predstavlja niskopropusni filtar, $v(n) \in \{-1, 1\}$. Spektralna gustina srednje snage vodenog žiga se, primjenom niskopropusnog filtra, suzbija ispod spektralne gustine srednje snage signala nosioca, čime se minimizuje uticaj vodenog žiga na rezultujući signal, odnosno pospješuje očuvanje kvaliteta signala. Parametar α se bira tako da ograniči srednju snagu vodenog žiga ispod praga čujnosti. Šema za detekciju je informisana i zasnovana na traženju korelacionih vrhova između ulaznog signala i vodenog žiga.

Tehnikom umetanja, predloženom u radu [1] vodeni žig je predstavljen skupom fazno modulisanih sinusoida, odnosno tonova. Modulacija je određena elementima pseudo-slučajne sekvence. Bit vodenog žiga dodaje se na sljedeći način:

$$y(n) = x(n) + \sum_i \alpha(i)\mathrm{PN}\left(n, i\right) \sin\left(\frac{2\pi n(f_0 + i)}{N}\right). \tag{25}$$

PN je pseudo-slučajna sekvenca, čiji način generisanja zavisi od bita vodenog žiga $v(n)$. Ovo je jedna od prvih tehnika u kojoj je predloženo da se vrijednost parametra α prilagođava pomoću psihoakustičnog modela i uzima različite vrijednosti

u zavisnosti od frekvencijskih komponenti u kojima se ugrađuje žig, umjesto da se unaprijed zadaje. Frekvencije na kojima se skrivaju bitovi vodenog žiga mogu se određivati na osnovu frekvencijskih komponenti signala nosioca kako bi dodati tonovi bili maskirani sadržajem tog signala. Ovaj metod ugrađivanja vodenog žiga čini sistem izuzetno otpornim na niskopropusno filtriranje. Međutim, nedostatak je u tome što bi napadač mogao frekvencijskom analizom otkriti dodate tonove i ekstrahovati vodeni žig. Takođe, kapacitet ovakvog sistema je veoma nizak.

Značajan broj tehnika u vremenskom domenu ne vrši nezavisne izmjene pojedinačnih odbiraka signala u cilju ugrađivanja bitova vodenih žigova. Pojedinačni odbirci se lako mogu poremetiti jednostavnim operacijama obrade signala. Takođe, ukoliko svaki odbirak signala treba da skriva jedan bit vodenog žiga, očuvanje kvaliteta je značajno ugroženo. Zbog toga su se kao mnogo uspješnije pokazale tehnike koje skrivaju jedan bit vodenog žiga u čitavom intervalu odbiraka. Ove tehnike su mnogo robustnije, a i vodeni žig je manje primjetan nakon ugrađivanja.

U radu [70] umetanje se vrši poređenjem energija za tri uzastopna intervala odbiraka i njihovim skaliranjem kako bi se zadovoljio predefinisani uslov. Uslov koji je potrebno zadovoljiti zavisi od bita koji se dodaje. Ova procedura sastoji se iz nekoliko koraka i ne može biti izražena u jednoj jednačini. Ekstrakcija bitova vrši se ispitivanjem uslova za odnos energija u uzastopnim intervalima odbiraka i određivanjem ekstrahovanog bita na osnovu uslova koji se ispostavi kao tačan.

Tehnika iz [71] za umetanje i detekciju vodenog žiga koriste histogram audio signala. Bit vodenog žiga se dodaje preraspodjelom broja odbiraka u tri uzastopna bina u histogramu kako bi ispunio jedan od dva uslova za ugrađivanje odabranog bita. Detekcija je neinformisana i ugrađeni bit se identifikuje tako što se za svaka tri uzastopna bina u histogramu votermarkovanog signala provjeri koji od dva uslova je ispunjen.

Skoriji rad [72] proširuje ovu ideju i za umetanje koristi specifično dizajniranu vrstu histograma, koja predstavlja robustnu karakteristiku signala. Signal se najprije podijeli na intervale, a zatim se umetanje jednog bita vodenog žiga vrši pomjeranjem vrijednosti u histogramu koja odgovara intervalu u koji se sakriva bit. Uvedeni histogram se izračunava tako što se interval odbiraka signala podijeli na grupe od po L_g odbiraka, a zatim se za svaku od grupa izračunava sljedeća vrijednost:

$$e_g = \sum_{i=0}^{L_g} (-1)^i \binom{L_g - 1}{i} x_g(i), \tag{26}$$

gdje su $x_g(i)$ odbirci iz iste grupe. Vrijednost E, za čitav interval, se izračunava sumiranjem vrijednosti za sve grupe: $E = \sum e_g$. Umetanje bita vodenog žiga u

posmatrani interval vrši se promjenom vrijednosti E. U slučaju ugrađivanja bita 1, mijenjaju se vrijednosti odbircima signala tako da se vrijednost E uveća za predefinisanu vrijednost α, kojom se kontroliše intenzitet vodenog žiga. Prilikom ugrađivanja bita 0, vrijednost E se ne mijenja. Na ovaj način detekcija je olakšana, jer se ponovnim izračunavanjem ovog histograma može zaključiti da li je vrijednost za posmatrani interval pomjerena ili ne i time odrediti koji bit vodenog žiga odgovara tom intervalu. U pomenutim radovima se ne razmatra diferencijacija votermarkovanih i nevotermarkovanih signala.

Jedna grupa votermarking tehnika vrši umetanje vodenog žiga u vještački dodatom ehu audio signala. Eho je efekat kojim se pojedini djelovi audio signala ponavljaju sa kašnjenjem. Ljudski sluh evoluirao je tako da filtrira kratke odjeke zvuka, što ih čini pogodnim za dodavanje vodenih žigova. Ova tehnika je prvi put primijenjena u [73]. Sastoji se u kreiranju vještačkog kratkog eha audio signala, slabog intenziteta, u kojem su skriveni bitovi vodenog žiga. Dodavanjem većeg broja eha signalu, povećava se kapacitet votermarking sistema. Fundamentalni nedostatak ovih tehnika je što ne mogu ugrađivati vodene žigove u intervalima tišine.

Ugrađivanje vodenog žiga u ehu audio signala može se predstaviti pomoću konvolucije, sljedećom generalizovanom jednakošću:

$$y = x * h, \tag{27}$$

gdje je h tzv. eho kernel, a „$*$" predstavlja operator konvolucije. Objašnjenje koncepta konvolucije dato je u Prilogu C.

Glavni detalj po kojem se tehnike zasnovane na ehu razlikuju je u dizajnu kernela. U osnovnom obliku je $h(n) = \delta(n) + \alpha\delta(n - d)$, δ je jedinična delta funkcija. Parametar α kontroliše jačinu eha, a parametar $d \in \{d_0, d_1\}$ predstavlja kašnjenje eha, čija dužina je određena vrijednošću bita koji se ugrađuje. Ovo je najčešći način za skrivanje informacija u ehu signala. Alternativa je dodavanje bitova u amplitudi eha. Svakako, podaci se mogu uspješno sakriti dok god se amplituda i kašnjenje eha drže pod kontrolom i ne izazivaju smetnje pri slušanju. Smatra se da kašnjenje do 1 ms ne može biti detektovano ljudskim slušnim sistemom.

U kasnijim radovima uvedeni su pozitivni i negativni eho kerneli [74], kao i koncept kernela koji djeluju unaprijed ili unazad [75], čime je uopšteni dizajn eho kernela proširen na sljedeći oblik:

$$h(n) = \delta(n) + \sum_{i=0}^{\mathcal{S}_{eho}} \alpha_{1,i}\delta(n - d_{1,i}) - \alpha_{2,i}\delta(n - d_{2,i}) + \alpha_{3,i}\delta(n + d_{3,i}) - \alpha_{4,i}\delta(n + d_{4,i}), \tag{28}$$

gdje je $\mathcal{S}_{eho}$ broj skupova eho kernela. Svaki skup sadrži po četiri eho kernela. Prema

jedakosti (28) prvi i treći kernel su pozitivni, dok su drugi i četvrti negativni. Takođe, prva dva kernela djeluju unaprijed, dok treći i četvrti djeluju unazad.

Detekcija vodenog žiga kod eho kernel tehnika vrši se kepstralnom analizom. Mana prvobitnih tehnika je što i napadač može otkriti eho, odnosno vodeni žig, kepstralnom analizom, bez ikakvog dodatnog znanja. Istraživači ulažu napore u osmišljavanje kvalitetnijih eho kernela kako bi se unaprijedili votermarking sistemi koji su bazirani na njima. U radu [76] definisan je sljedeći kernel:

$$h(n) = \delta(n) + \alpha \mathrm{PN}(n - d), \tag{29}$$

koji je uveo korišćenje pseudo-slučajne sekvence PN u ovu grupu votermarking tehnika. Vodeni žig se ekstrahuje korelacijom kepstralnog sadržaja votermarkovanog signala sa pseudo-slučajnom sekvencom. Stoga, korišćenje pseudo-slučajne sekvence pospješuje sigurnost ove metode, dok god je ta sekvenca tajna.

Tehnika iz [77] takođe vrši umetanje vodenog žiga u ehu audio signala. Međutim, u cilju poboljšanja sigurnosti prije samog dodavanja vodenog žiga, signal se razbija na skup podsignala $\{x_1, x_2, ..., x_{sub}\}$, čijim sabiranjem se može rekonstruisati originalni signal. Procedura umetanja se sprovodi u domenu Furijeove transformacije kako bi se podsignali sortirali u opadajući poredak po energiji $\{\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_{sub}\}$. Zatim se eho kernel primjenjuje nad svaka dva uzastopna podsignala u vremenskom domenu:

$$y(n) = x(n) + \sum_{s=1}^{\frac{sub}{2}} (\tilde{x}_{2s-1} * h)(n) - (\tilde{x}_{2s} * h)(n). \tag{30}$$

Eho kernel $h(n) = \alpha(\delta(n - d) - \delta(n - d - d') + \delta(n + d) - \delta(n + d - d'))$ ima dva parametra d i d'. U proceduri ekstrakcije se ovi podsignali rekonstruišu istim postupkom kao i prilikom umetanja, pa se kepstralnom analizom identifikuju bitovi vodenog žiga. Skrivanjem vodenog žiga u energetski izbalansiranim podsignalima napadaču je otežano da konvencionalnom analizom pronađe kepstralne pikove koji bi ga uputili ka razotkrivanju vodenog žiga. Sama operacija umetanja izvršava se u vremenskom domenu, što ovu tehniku svrstava u istu grupu kao i ostale tehnike iz ove sekcije. Ipak, dio procedura umetanja i detekcije koji je najznačajniji za pospješivanje sigurnosti vodenog žiga obavlja se u transformacionom domenu, što upućuje na pretpostavku da su votermarking tehnike u ovom tipu domena superiornije.

4.2 Tehnike za ugrađivanje vodenog žiga u transformacionim domenima

Tehnike koje sprovode ugrađivanje vodenog žiga u transformacionom domenu preferirane su u istraživačkoj zajednici. Smatra se da postižu veću robustnost od tehnika u vremenskom domenu, ali još uvijek ne postoje studije koje bi podržale ovu tvrdnju. Ukoliko je signal reprezentovan u vremenskom domenu, većina efekata izmijeniće većinu ako ne i sve odbirke signala. Koeficijenti u transformacionim domenima nisu toliko osjetljivi, pa se zbog toga smatra da su pogodniji za skrivanje bitova vodenog žiga. Još jedan razlog upotrebe transformacionih domena je i iskorišćenje nedostataka ljudskog slušnog sistema u pogledu osjetljivosti na određene frekvencije. Primjena transformacija razbija signal na više različitih vremensko-frekvencijskih komponenti. Ljudsko uho ima tendenciju da maskira slabije frekvencije u prisustvu jačih frekvencijskih komponenti. Izdvajanjem ovih komponenti u frekvencijskom domenu mogu se lakše sakriti podaci u zvuku.

Sistemi koji implementiraju ove tehnike iziskuju dodatni korak izračunavanja koeficijenata transformacije na početku procedura umetanja i detekcije. Takođe je potrebno, računanjem inverzne transformacije, iz modifikovanih koeficijenata rekonstruisati signal u vremenskom domenu na kraju procedure umetanja. Uglavnom se sve operacije vrše u domenima poznatih unitarnih ortogonalnih transformacija poput DFT, DCT i DWT. Kepstralni domen takođe predstavlja prostor za vršenje votermarking operacija [78]. Ovaj domen karakterističan je po svojoj otpornosti na različite audio efekte. Tehnike su često dizajnirane tako da se mogu primijeniti na koeficijente bilo koje transformacije, bez značajnog pada u performansama po bilo kojem od kriterijuma.

Umetanje vodenog žiga u najmanje značajnim bitovima moguće je i u ovim domenima [79, 80]. Za umetanje se koriste posljednji biti koeficijenata odabrane transformacije. Međutim ni ove tehnike, kao i one u vremenskom domenu, nisu naišle na širok prijem u praksi, zbog svoje inferiorne robustnosti.

Primjenom transformacije poput DFT, audio signal se može modelovati kao niz kompleksnih brojeva $X(k) = a_k + jb_k$. Svakoj od frekvencijskih komponenti, predstavljenoj koeficijentom $X(k)$ mogu se izračunati amplituda i faza. Ovo su dvije veoma važne karakteristike koje se često koriste u analizi signala, pa se mogu koristiti i u šemama za ugrađivanje i detekciju vodenih žigova.

$$A(k) = \sqrt{a_k^2 + b_k^2}, \tag{31}$$

$$\phi(k) = \arctan \frac{b_k}{a_k}. \tag{32}$$

Amplituda $A(k)$ određuje koliko je intenzivno prisustvo posmatrane sinusoidne komponente u signalu u odnosu na druge. Faza $\phi(k)$ određuje kako je ta sinusoidna komponenta usklađena u vremenu sa ostalim sinusoidama u signalu. Koeficijente DFT moguće je na jednostavan način rekonstruisati iz amplitude i faze:

$$X(k) = A(k)e^{j\phi(k)}. \tag{33}$$

Još jedno od ograničenja ljudskog sluha je neosjetljivost na fazu zvučnog signala u većini situacija. Za razliku od slika, faza kod audio signala ne nosi značajnu količinu informacija. Ovo fazu naizgled čini pogodnom za umetanje vodenih žigova, jer se ona može intenzivnije modifikovati bez osjetnog ugrožavanja kvaliteta zvuka. Na ovaj način može se postići veća otpornost na šum, filtriranje, kompresiju i slične efekte. U radu [2] predložena je sljedeća šema za umetanje vodenog žiga modulacijom faze signala:

$$Y(k) = X(k)e^{\alpha(k)v(k)}. \tag{34}$$

Međutim, maliciozni korisnik takođe može iskoristiti svojstvo nečujnosti faze i u potpunosti je izmijeniti prilikom napada. Time će se izgubiti informacija o vodenom žigu, dok će informacija koju nosi signal ostati u velikoj mjeri očuvana. Ovo je glavni razlog iz kojeg skrivanje vodenih žigova u fazi audio signala nije preporučljivo.

Skrivanje vodenog žiga u amplitudi signala je komplikovanije, ali poželjnije. Amplituda nakon umetanja mora biti gotovo savršeno rekonstruisana jer će slušaoci greške u rekonstrukciji amplitude mnogo lakše detekovati. Međutim, uspješnim skrivanjem vodenog žiga u amplitudi obezbjeđuje se visoka sigurnost vodenog žiga. Ova procedura se sprovodi tako što se u spektru audio signala najprije pronađu pogodna mjesta za skrivanje podataka, a zatim se vrši frekvencijsko maskiranje. Frekvencijsko maskiranje je efekat koji se javlja kada se dva slična zvuka pojavljuju istovremeno. Jači zvuk tada maskira slabiji onemogućavajući mu time da bude percipiran.

Širenje spektra je jedna od najčešće korišćenih tehnika u digitalnom votermarkingu. Uvedena je s ciljem izbjegavanja skrivanja bitova vodenog žiga isključivo u visokofrekventnim komponentama signala, koje je lako eliminisati iz signala a da to bude neprimjetno. Bitovi vodenog žiga se ovom metodom ugrađuju blagim modifikacijama koeficijenata u svim djelovima frekvencijskog opsega. Primjena ove tehnike donosi nekoliko benefita. Širenjem vodenog žiga u obliku šuma male amplitude po čitavom frekvencijskom opsegu smanjuje se šansa za kreiranjem čujnih artefakata. Takođe, skrivanjem vodenog žiga na različitim frekvencijama osigurava se veći stepen pravilno detektovanih bitova, jer je vodeni žig isprepletan sa signalom, pa ga je teže eliminisati dodavanjem šuma, niskopropusnim filtriranjem ili drugim audio

efektima. Gotovo svi moderni votermarking sistemi koje svoje operacije vrše u frekvencijskom domenu inkorporiraju ovu tehniku u nekom obliku. Ona ne isključuje upotrebu drugih votermarking tehnika koje će biti pomenute u nastavku ove sekcije.

Korišćenje tehnike širenja spektra predloženo je u [43] za votermarking digitalnih slika. Ova tehnika može biti predstavljena aditivnim modelom (2) ili multiplikativnim modelom:

$$Y(k) = X(k) + \alpha X(k)v(k). \tag{35}$$

Tehnika širenja spektra je dalje unapređivana i prilagođena za primjenu nad audio signalima. Prva publikacija u kojoj je širenje spektra iskorišteno za umetanje vodenih žigova u audio signale je [81]. Model koji je korišten za umetanje identičan je jednakosti (2). Modifikacija koja je napravljena je izbjegavanje umetanja bitova vodenog žiga u tihim segmentima signala, kao i onim segmentima izuzetno bogatim zvukom, kako bi se što bolje očuvao kvalitet signala. Ovi segmenti su određivani empirijski. U detekciji su korelacioni testovi zamijenjeni kepstralnim filtriranjem, što je poboljšalo performanse, a vršeno je i redudantno umetanje.

Autori u [82] predlažu sljedeći model za umetanje vodenih žigova:

$$Y(k) = X(k) + (\alpha - \alpha'\Phi(X, v))v(k), \tag{36}$$

gdje je $\Phi(X, v) = \frac{\sum_k X(k)v(k)}{\sum_k v(k)^2}$. Parametrom α' se kontroliše uticaj signala nosioca na proces detekcije. Rad [83] dodatno je unaprijedio tehniku širenja spektra proširivanjem sheme za umetanje (35) funkcijom Φ koja je osmišljena tako da maksimizuje detekcionu statistiku:

$$Y(k) = X(k) + \alpha\Phi(X, v)X(k)v(k). \tag{37}$$

U radu [3] se vodeni žigovi kodiraju pseudo-slučajnim sekvencama brojeva 1 i -1 i koriste se varijabilne vrijednosti za parametar kojim se kontroliše snaga vodenog žiga α:

$$Y(k) = X(k) + \alpha(k)\frac{X(k)\mathrm{PN}(k)}{|X(k)\mathrm{PN}(k)|}\mathrm{PN}(k). \tag{38}$$

Sve operacije obavljaju se u DCT domenu. Detektor je korelacioni. Izračunava se linearna korelacija (5) segmenta ulaznog signala i svih pseudo-slučajnih sekvenci i kao rezultat uzima se ona sekvenca koja rezultuje maksimalnom vrijednošću korelacije sa ulaznim signalom.

Tehnika širenja spektra, pored svojih prednosti, ima i određene nedostatke. Ispostavlja se da može izazvati smetnje u signalu, koje pojedini slušaoci percipiraju kao „zujanje". Osim toga, ukoliko dođe do promjena u visinama tonova, detekcija

vodenog žiga može postati nemoguća, budući da ovaj efekat izaziva pomjeranje frekvencijskih komponenti u signalu, a time i bitova vodenog žiga. Stoga su potrebne naprednije tehnike kako bi se prevazišli ovi nedostaci.

Modulacija kvantizovanog indeksa (*engl. quantization index modulation* - QIM) je veoma rasprostranjena tehnika za umetanje vodenih žigova. Sastoji se iz dva koraka, modulacije i kvantizacije. U koraku modulacije se odabrani koeficijent ili grupa koeficijenata pomjera (modulira), najčešće za predefinisanu vrijednost. U drugom koraku primjenjuje se kvantizacija, odnosno zaokruživanje vrijednosti koeficijenata na cjelobrojni umnožak odabranog koraka kvantizacije. Ove operacije treba da budu dovoljno blage da ne ugroze kvalitet signala, ali i dovoljno jake da bi se u procesu detekcije moglo raspoznati kakva je promjena izvršena kako bi se bitovi vodenog žiga ekstrahovali. Ova tehnika je originalno predložena u [84] za sve vrste multimedija. Motiv za njeno uvođenje bila je slaba robustnost tadašnjih tehnika zasnovanih na širenju spektra. QIM šema za umetanje vodenih žigova može se generički predstaviti sljedećim izrazom:

$$Y(k) = \begin{cases} \left\lfloor \left\lfloor \frac{X(k)}{q} \right\rfloor + \alpha \right\rfloor q, & v(k) = 1 \\ \left\lfloor \left\lfloor \frac{X(k)}{q} \right\rfloor - \alpha \right\rfloor q, & v(k) = -1. \end{cases} \tag{39}$$

Parametar q predstavlja odabrani kvantizacioni faktor. Vodeni žig kodiran je vrijednostima 1 i -1. Povećavanje vrijednosti parametra Q čini sistem robustnijim, ali isto tako postoji veća vjerovatnoća za pojavom artefakata koji degradiraju zvuk. Parametrom α se kontroliše modulacija. Moguće je odabrati različite vrijednosti za modulaciju prilikom umetanja različitih bitova vodenog žiga. Šeme za detekciju kod QIM votermarkinga su neinformisane i obično ne iziskuju iscrpnu pretragu prostora vodenih žigova. Usko su povezane sa šemom umetanja. Potrebno je za dobijeni koeficijent $Z(k)$ zaključiti kojim pravilom je kvantizovan. QIM metode postižu izuzetnu robustnost na aditivni šum, ali su očigledno osjetljive na efekat skaliranja amplitude.

QIM tehnika je od strane svojih autora veoma dobro utemeljena u radu [84]. Shodno tome, budući radovi nisu bili fokusirani na modifikacije i poboljšanja ove tehnike, već pokušavaju QIM uklopiti u svoja rješenja. Autori u [5] QIM tehniku primjenjuju nad srednjom vrijednošću DWT koeficijenata trećeg nivoa. Tehnikom iz [16] se umetanje vodenog žiga vrši u DCT domenu. DCT koeficijenti se podijele na dvije grupe $\{X_0, X_1\}$, slučajnim izborom. QIM se primjenjuje nad normom prve grupe koeficijenata $\|X_0\|$:

$$\|Y_0\| = \begin{cases} \left\lfloor \frac{\|X_0\|}{q} \right\rfloor q + \frac{q}{2}, & v(k) = 1 \\ \left\lfloor \frac{\|X_0\|}{q} + 0.5 \right\rfloor q, & v(k) = -1. \end{cases} \tag{40}$$

Uticaj ovih operacija na koeficijente votermarkovanog signala može se opisati sljedećom jednakošću:

$$Y_0(k) = \frac{\|Y_0\|}{\|Y_0 + \varepsilon\|} \left(X_0(k) + \varepsilon(k)\right), \tag{41}$$

gdje je ε vektor koji utiče na intenzitet kojim se bitovi vodenog žiga ugrađuje u signal. Autori ovog rada tvrde da se kvantizacijom u prostoru norme, umjesto u prostoru koeficijenata, poboljšava robustnost vodenog žiga. Druga grupa DCT koeficijenata se koristi da kompenzuje varijaciju u energiji koja se unosi kvantizacijom koeficijenata iz prve grupe. Na taj način se unapređuje kvalitet votermarkovanog signala. Detekcija je jednostavna. Za svaki od koeficijenata iz prve grupe se nezavisno određuje skriveni bit vodenog žiga:

$$\hat{v}(k) = \begin{cases} 1, & \left| \frac{\|Z_0\|}{q} - \left\lfloor \frac{Z_0(k)}{q} \right\rfloor - 0.5 \right| \leq 0.25 \\ -1, & \text{inače.} \end{cases} \tag{42}$$

Nakon toga se većinskim glasanjem može doći do konačne odluke. Shema za podjelu koeficijenata i parametar q moraju biti poznati detektoru, što ih čini slabim tačkama ove tehnike.

Autori u [6] koriste QIM i SVD za umetanje vodenih žigova u stereo audio signale. SVD je tehnika za dekompoziciju matrica koja se često koristi u votermarkingu, zbog robustnosti singularnih vrijednosti. Proizvoljna realna ili kompleksna matrica X ranga r_X se može ovom tehnikom predstaviti na sljedeći način:

$$X = U \Lambda V^T, \tag{43}$$

gdje su U i V unitarne matrice, a Λ je Žordanova matrica sa r_X nenultih singularnih vrijednosti. Kako je stereo signal x podijeljen na dva kanala: lijevi x_l i desni x_d, to se on može predstaviti kao matrica sa dva reda $X = [x_l, x_d]$. Dekompozicija ove matrice na singularne vrijednosti rezultuje sa dva nenulta člana matrice Λ, označene sa λ_0 i λ_1. Kvantizacija se vrši nad ovim vrijednostima, a zatim se votermarkovani signal dobija njihovim uvrštavanjem nazad u izraz (43). Ovi koraci iz procedure umetanja, sadržani su i u detekciji. Nad dobijenim signalom z vrši se SVD, pa se u skupu referentnih vodenih žigova pokušava pronaći onaj čijim umetanjem se dobija signal na najmanjem rastojanju od z. Ovakav pristup detekciji je veoma nepraktičan, jer iako je provjera za jedan vodeni žig efikasna, čitav skup vodenih žigova u nekim aplikacijama može biti izuzetno veliki.

SVD tehnika koristi se i u [7], da bi se iz matrice DCT koeficijenata, za dva uzastopna segmenta signala, izračunala karakteristična vrijednost nazvana koeficijent frekvencijske singularne vrijednosti (*engl. frequency singular value coefficient* - FSVC). Ova tehnika modifikuje FSVC vrijednost i na taj način se umeće jedan

bit vodenog žiga u signal. Korak kvantizacije se ne sprovodi, pa se ova tehnika ne može svrstati u QIM grupu. Ono što ovu tehniku čini kvalitetnijom od [6] je što se ne mora vršiti iscrpno pretraživanje skupa vodenih žigova prilikom detekcije, kao i određeni nivo otpornosti FSVC karakteristike na efekte desinhronizacije.

Pečvork algoritam (*engl. patchwork algorithm*) iz [85] primijenjen je za votermarking audio signala u [86]. Pečvork algoritam se primjenjuje tako što se koeficijenti transformacije X podijele u dva jednaka, disjunktna skupa X_0 i X_1. Elementi skupova nisu determinisani, već se slučajno uzorkuju. U elemente tih skupova se ugrađuje bit vodenog žiga na sljedeći način:

$$Y_i(k) = X_i(k) + (-1)^i \alpha v(k), \tag{44}$$

gdje $i \in \{0, 1\}$ Parametrom α se kontroliše intenzitet modifikacije nad signalom. Bitovi vodenog žiga se, u ovom slučaju, kodiraju kao pozitivne vrijednosti ($v(k) > 0$ $\forall k$). Nakon primjene (44), koeficijenti Y_0 i Y_1 se spajaju u jedan skup kako bi se signal mogao emitovati.

Šema za detekciju kod pečvork algoritma je informisana. Koeficijenti se dijele u dva skupa na isti način kao i kod umetanja. Ugrađeni bit vodenog žiga se ekstrahuje iz dobijenog signala z na osnovu sljedeće statistike:

$$\hat{v}(k) = \frac{\sum_k (Z_0(k) - Z_1(k)) - \sum_k (X_0(k) - X_1(k))}{2K}. \tag{45}$$

U radu [87] predložena je modifikacija ovog algoritma kojom se unapređuju robustnost i nečujnost. Šema umetanja proširuje se dodatnim članovima koji će testnu statistiku prilikom ekstrakcije bitova vodenog žiga učiniti prepoznatljivijom. U ovoj šemi skupovi koeficijenata X_0 i X_1 se dodatno dijele u dvije grupe, prema bitu koji se ugrađuje. Na ovaj način formiraju se četiri skupa koeficijenata $X_{i,j}$, $i \in \{0, 1\}$, $j \in \{0, 1\}$, a jedan bit vodenog žiga ugrađuje se po sljedećem pravilu:

$$Y_{i,j}(k) = X_i(k) + (-1)^i \alpha \Phi_j(X) \Phi'_j(X), \tag{46}$$

gdje je:

$$\Phi_j(X) = \operatorname{sgn}\left\{ \sum_k (X_{0,j}(k) - X_{1,j}(k)) \right\} \tag{47}$$

i

$$\Phi'_j(X) = \sqrt{\frac{\left(\sum_k \left(X_{0,j}(k) - \overline{X}_{0,j} \right)^2 - \sum_k \left(X_{1,j}(k) - \overline{X}_{1,j} \right)^2 \right)}{K(K-1)}}. \tag{48}$$

$\overline{X}_{0,j}$ i $\overline{X}_{1,j}$ predstavljaju srednju vrijednost odgovarajuće grupe koeficijenata. U toku detekcije se izračunavaju dvije testne statistike: $T_0 = \frac{\left(\overline{Z}_{0,0} - \overline{Z}_{1,0}\right)^2}{\Phi'_0(Z)^2}$ i $T_1 = \frac{\left(\overline{Z}_{0,1} - \overline{Z}_{1,1}\right)^2}{\Phi'_1(Z)^2}$

na skupu dobijenih koeficijenata Z. Njihovim poređenjem sa definisanim pragom se određuje da li je dati skup koeficijenata označen vodenim žigom, i ako jeste, koji bit je ugrađen u njega.

Slaba tačka osnovnog pečvork algoritma je što se šema za podjelu koeficijenata u dva skupa mora proslijediti detektoru kao tajni ključ. Ovo ostavlja prostor napadaču da presretanjem ovog ključa ovlada tehnikom detekcije i time ugrozi sigurnost sistema.

Većina energije audio signala koncentrisana je na niskim frekvencijama. Iz ovog razloga se vrijednosti koeficijenata $X(k)$ uglavnom smanjuju kako se k povećava. Autori [15] vođeni ovim zapažanjem vrše preuređivanje DCT koeficijenta tako da srednja vrijednost energije u uzastopnim grupama koeficijenata, nakon preuređivanja, bude izbalansirana. Jedan bit vodenog žiga se zatim umeće u dvije uzastopne grupe koeficijenata modifikacijom njihovih srednjih vrijednosti. Preuređivanje koeficijenata umanjuje degradaciju kvaliteta signala, jer će ove dvije grupe biti dobro energetski izbalansirane. Ovim se ujedno i prevazilazi pomenuti nedostatak osnovnog pečvork algoritma jer se otklanja potreba za proslјeđivanjem ključa za raspodjelu koeficijenata koja je ovdje određena pravilom za preuređivanje.

Sve pomenute tehnike u transformacionom domenu mogu se uporedo koristiti u cilju dizajniranja što kvalitetnijeg votermarking sistema. Zapravo, tehnika širenja spektra prisutna je u svim ostalim metodama. Vodeni žig se dodaje, odnosno širi, na sve koeficijente odabrane transformacije. Jednakost (2) sadržana je, u nekom obliku, u svim ostalim pravilima koja su pomenuta u ovoj sekciji. Ni ostale tehnike nisu međusobno isključive. Svaka od njih donosi nova sredstva koja se mogu koristiti u izgradnji votermarking sistema. QIM tehnike uvele su korak kontrolisane kvantizacije, a pečvork algoritmom uvedena je dvokanalna shema za umetanje i detekciju vodenih žigova.

Na primjer, rad [4] objedinjuje pečvork algoritam i QIM tehniku. Nakon podjele DCT koeficijenata audio segmenta na dva skupa, za svaki skup se zasebno računa karakteristična vrijednost koju autori nazivaju logaritamska sredina frekvencijskog domena (*engl. frequency domain logarithmic mean* - FDLM). Kvantizacijom FDLM vrijednosti ova dva skupa koeficijenata u audio signal se ugrađuje jedan bit vodenog žiga.

Tehnika iz [88] kombinuje dvokanalnu šemu za umetanje iz pečvork algoritma i umetanje bitova vodenog žiga korekcijom normi vektora koeficijenata. Najprije se pomoću DWT izdvajaju niskofrekventne komponente signala, nad kojima se zatim primjenjuje DCT, od čijih koeficijenata se formiraju dva vektora X_0 i X_1. U narednom koraku vrši se izjednačavanje normi ovih vektora na njihovu prosječnu

vrijednost:

$$\overline{\|X\|} = \frac{\|X_0\| + \|X_1\|}{2}. \tag{49}$$

Zatim se u zavisnosti od toga koji bit se ugrađuje jedna od normi povećava, a druga smanjuje za fiksnu vrijednost.

$$\|Y_i\| = \begin{cases} \overline{\|X\|} + (-1)^i\alpha, & v(k) = 1 \\ \overline{\|X\|} - (-1)^i\alpha, & v(k) = -1, \end{cases} \tag{50}$$

gdje je $i \in \{0, 1\}$. Parametrom α se kontroliše intenzitet umetanja jednog bita vodenog žiga, odnosno uticaj na očuvanje kvaliteta signala.

Šema za detekciju je jednostavna i neinformisana. Postupak umetanja se ponavlja do izračunavanja normi vektora Z_0 i Z_1 za dobijeni signal z. Zatim se vrši ekstrakcija bita vodenog žiga po sljedećem pravilu:

$$\hat{v}(k) = \begin{cases} 1, & \|Z_0\| > \|Z_1\| \\ -1, & \|Z_0\| \leq \|Z_1\|. \end{cases} \tag{51}$$

5 Neuronske mreže

Za sistem se može reći da uči ukoliko poboljšava svoje performanse s prikupljenim iskustvom, odnosno nakon što se snabdije novim zapažanjima o postavljenom problemu. Ovaj proces se naziva mašinsko učenje. Cilj mašinskog učenja je stvaranje sistema, odnosno modela, koji mogu rješavati postavljene zadatke, bez potrebe da budu eksplicitno programirani. Modeli mašinskog učenja pogodni su za korišćenje u situacijama kada je teško ili nemoguće unaprijed predvidjeti sve moguće scenarije ili formulisati precizna pravila za rešavanje problema.

Neuronske mreže su velika familija modela mašinskog učenja. Jedna neuronska mreža je skup međusobno povezanih računskih jedinica, odnosno čvorova. Čvorovi su organizovani u slojeve, od kojih svaki vrši različite transformacije svojih ulaza. Na taj način, neuronskim mrežama mogu se modelovati veoma složene funkcije i rješavati veoma kompleksni problemi. Mreže sa velikim brojem slojeva se nazivaju „dubokim". Zbog specifičnih problema koji se javljaju pri njihovom obučavanju i tehnika koje su nastale u cilju njihovog rješavanja, izdvojila se posebna grana mašinskog učenja nazvana „duboko" učenje (*engl. deep learning*).

U zajednici je vođena višegodišnja diskusija oko upotrebe termina „neuralna" umjesto "neuronska", argumentovana potrebom za razlikovanjem između bioloških i vještačkih mreža neurona. Međutim, u savremenoj literaturi oba termina se i dalje ravnopravno koriste, pa smo se u ovoj disertaciji opredijelili za termin „neuronska mreža", kako bismo opisali vještačke modele inspirisane strukturom i funkcionisanjem bioloških neurona.

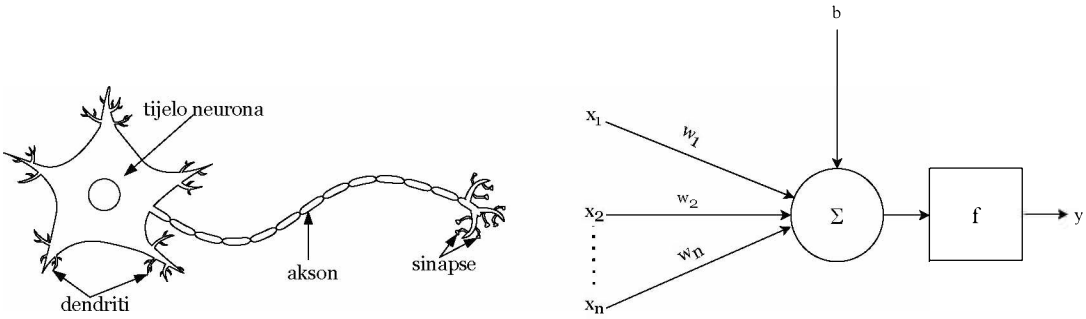
Razvoj neuronskih mreža počeo je još sredinom XX vijeka radom Mekaloha i Pitsa [89], gdje je izložena ideja da se kreira računski model oponašanjem neurona u biološkom mozgu. Prvi model neurona koji je mogao biti obučavan, nazvan perceptron, razvio je Rozenblat [90]. Slaganjem više perceptrona u slojeve, dobijene su prve neuronske mreže. Međutim, u daljem razvoju su neuronske mreže sve više gubile sličnost sa biološkim mozgom. Primarni cilj je bilo poboljšavanje empirijskih rezultata, što je obično u neuronske mreže uvodilo koncepte koji nisu utemeljeni u neurologiji ili biologiji. Iako dalji razvoj nije u potpunosti slijedio oponašanje biološkog mozga, od te inicijalne ideje neuronskim mrežama ostao je naziv.

Do sada su se u razvoju vještačke inteligencije desila dva perioda stagnacije, odnosno takozvane zime. Ovo su bili periodi izuzetno umanjenog interesovanja i finansiranja istraživanja na polju vještačke inteligencije. Prva zima trajala je od 1974. do 1980. godine, a druga od 1987. do 1993. godine. Pored teorijskih nedoumica i nesuglasica, jedan od glavnih uzroka ovih zastoja u razvoju vjestačke inteligencije je

što tadašnji računari nisu imali dovoljno procesorske moći za obučavanje praktičnih neuronskih mreža. Ipak, i u ovim periodima pojavilo se nekoliko značajnih otkrića među kojima je i algoritam propagacije unazad (*engl. backpropagation*) [91] koji se i danas dominantno koristi za obučavanje neuronskih mreža. Razvoj moćnih grafičkih kartica i distribuiranih sistema, kao i dostupnost velikih korpusa podataka prouzrokovana ekspanzijom interneta početkom ovog vijeka omogućile su obučavanje znatno većih neuronskih mreža nego ranije, čime je otpočela era dubokog učenja. Danas se u istraživanja u oblasti vještačke inteligencije ulažu značajna sredstva. Veliki entuzijazam vlada i u istraživanju i u primjeni vještačke inteligencije. U svemu tome prednjače neuronske mreže. Međutim, preveliki publicitet, neprirodno velika obećanja inženjera i očekivanja korisnika mogu da se otrgnu kontroli i da nas dovedu u još jednu zimu vještačke inteligencije.

Duboke mreže pokazale su se naročito efikasnim u pogledu visokodimenzionalnih podataka kakve su slike, zvučni signali i tekst. Neuronske mreže su trenutno najbolji računski modeli za prepoznavanje objekata na slikama, sintetisanje teksta i slika, prepoznavanje i sintetisanje govora, mašinsko prevođenje, itd. Koriste se i u drugim oblastima, gdje u kombinaciji sa drugim tehnikama takođe nadmašuju konkurenciju. Razlozi njihove uspješnosti u različitim zadacima još uvijek nisu u potpunosti razjašnjeni, ali neke prednosti su očigledne. Drugi modeli mašinskog učenja obično primjenjuju mali broj operacija nad ulazom, odnosno imaju kratke računske putanje. Takođe, ulazne promjenljive često na nezavisan način utiču na izlazne vrijednosti, bez međusobne interakcije. Dodatno, modeli dubokog učenja pokazuju sposobnost učenja relevantnih apstraktnih karakteristika ulaza, dok se u drugim modelima mašinskog učenja mora izvršiti manuelna definicija i selekcija atributa. U mnogim realnim problemima je izuzetno komplikovano dizajnirati odgovarajući skup karakteristika ulaza koje se prosljeđuju modelu. Ovim nedostacima se ograničavaju mogućnosti ovih modela u modelovanju koncepata iz realnog svijeta koji mogu biti izuzetno kompleksni. Osnovna ideja dubokog učenja je upravo u kreiranju modela sa dugim računskim putanjama i složenim interakcijama među promjenljivima koji su u stanju da na adekvatan način predstave procese iz realnog svijeta.

Mašinsko učenje može se, kada je to potrebno, posmatrati i kao problem aproksimacije funkcija. Svi prethodno navedeni primjeri mogu se tako predstaviti. U svakom od njih je, za date vrijednosti ulaza, potrebno nizom matematičkih operacija proizvesti odgovarajući izlaz, odnosno modelovati odgovarajuću funkciju. Prema nekoliko univerzalnih teorema o aproksimaciji, dovoljno složene neuronske mreže mogu aproksimirati bilo koju neprekidnu funkciju sa proizvoljnom preciznošću. Neuronske mreže sa jednim [92, 93] ili dva [94] skrivena sloja sa odgovarajućim aktivacionim funkcijama i obilnim ali konačnim brojem neurona imaju mogućnost univerzalne



(a) Biološki neuron **(b)** Vještački neuron

Slika 7: Skice biološkog i vještačkog neurona.

aproksimacije. U jednom novijem radu [95] dokazana je teorema univerzalne aproksimacije i za duboke konvolucione mreže sa dovoljnim brojem slojeva. Ovo je naznaka da se one mogu koristiti i za rješavanje svih problema digitalnog votermarkinga.

U nastavku ovog poglavlja detaljno su objašnjeni svi koncepti iz oblasti dubokog učenja korišteni u ovom radu za kreiranje sistema vodenog žiga. Opisani su gradivni elementi konvolucionih neuronskih mreža i operacije koje se u okviru njih vrše. Takođe, predstavljene su različite procedure i tehnike koje se primjenjuju kako bi se stvorili što bolji modeli. Svi uvedeni koncepti predočeni su sa visokim nivoom detaljnosti i obuhvataju značajan spektar pojmova u oblasti dubokog učenja.

5.1 Osnove neuronskih mreža

Biološki neuron ima mnoštvo dendrita preko kojih prima ulazne signale, koji se zatim obrađuju u tijelu ćelije i prosljeđuje rezultujući signal preko aksona i sinapsi sljedećim neuronima. Na taj način izlaz jednog neurona postaje ulaz drugog. Skica biološkog neurona data je na Slici 7a. Vještački neuron takođe može imati više ulaza, koje reprezentujemo vektorom $x = (x_1, x_2, ..., x_n)$ i jedan izlaz y. Svakom ulazu pridružen je težinski koeficijent. Slika 7b prikazuje blok šemu vještačkog neurona. Neuron računa ponderisanu sumu svojih ulaza, a zatim na dobijenom rezultatu primjenjuje nelinearnu funkciju f kako bi dobio izlaz y.

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right), \tag{52}$$

gdje je $w = (w_1, w_2, ..., w_n)$ vektor težinskih koeficijenata iz skupa realnih brojeva. Ponderisanoj sumi ulaza dodaje se i slobodni član b (*engl. bias*) kako bi vještački neuron mogao modelovati i funkcije koje ne prolaze kroz koordinatni početak. Funkcija f se naziva aktivacionom funkcijom jer se njome kontroliše da li neuron treba

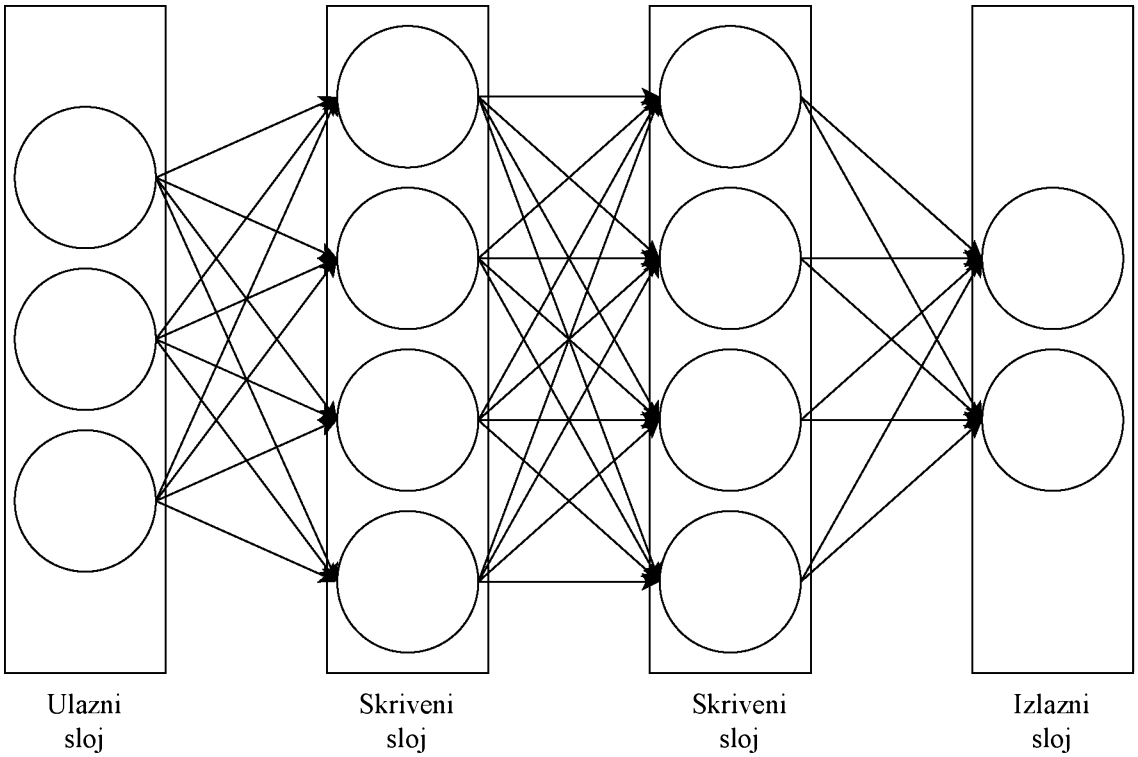
da se aktivira i u kojoj mjeri za date ulaze. Više detalja o aktivacionim funkcijama i načinu njihovog izbora biće dato u Sekciji 5.3.

Neuronske mreže formiraju se grupisanjem neurona u slojeve i njihovim povezivanjem u proizvoljnim obrascima. Na taj način mogu se modelovati veoma složene funkcije koje su kompozicija funkcija koje računaju pojedinačni neuroni. Svakom neuronu pridruženi su težinski koeficijenti i slobodni član. Ove vrijednosti jednim imenom nazivaju se parametri neuronske mreže. Obučavanje neuronske mreže je zapravo procedura prilagođavanja ovih parametara kako bi se riješio postavljeni zadatak, odnosno što uspješnije izmodelovala ciljna funkcija. Pored parametara, u procesu obučavanja neuronskih mreža koriste se i druge vrijednosti koje se unaprijed zadaju i kojima se kontroliše tok ovog procesa. U cilju njihovog razlikovanja od parametara mreže ove vrijednosti se nazivaju hiperparametri.

Slojevi neuronske mreže dijele se na nekoliko tipova. Sloj koji prima ulazne podatke naziva se ulazni sloj. Ovaj sloj ne vrši nikakvu transformaciju ulaza, već samo učitava podatke i prosljeđuje ih dalje. Sloj koji se nalazi na kraju mreže i koji emituje finalni izlaz naziva se izlazni sloj. Moguće je da neuronska mreža ima više ulaznih i izlaznih slojeva. Između ulaznog i izlaznog sloja nalazi se proizvoljan broj tzv. skrivenih slojeva. U porijeklu imena ovih slojeva nema dubljeg značenja, osim potrebe da se razlikuju od ulaznih i izlaznih slojeva. Skriveni slojevi su srž neuronskih mreža. U njima se vrše sve operacije u cilju transformisanja dobijenog ulaza u željeni izlaz.

Šema po kojoj se kreiraju slojevi neurona i veze između njih naziva se arhitekturom neuronske mreže. Iako arhitektura neuronske mreže može biti proizvoljna, u literaturi se ustalila kategorizacija mreža u odnosu na njihovu arhitekturu. U zavisnosti od toga da li se povezivanjem neurona formira acikličan ili graf sa ciklusima razlikujemo nepovratne (*engl. feedforward*) i povratne (rekurentne) neuronske mreže. Nepovratne su one mreže u kojima neuroni formiraju usmjeren acikličan graf. Svaki neuron u takvoj mreži izračunava vrijednost svog izlaza i prosljeđuje ga na ulaz svojim sljedbenicima. Informacije se prostiru od ulaza mreže prema izlazu i nema povratne sprege. Sa druge strane, rekurentne mreže pohranjuju međurezultate ili izlaze svojih slojeva nazad na njihov ulaz.

Osnovni i najjednostavniji tip nepovratnih neuronskih mreža su tzv. potpuno povezane neuronske mreže. Kod ovog tipa mreža neuron iz jednog sloja povezan je sa svim neuronima u sljedećem sloju. Neuroni u istom sloju nisu međusobno povezani. Primjer potpuno povezane neuronske mreže sa jednim ulaznim, dva skrivena i jednim izlaznim slojem dat je na Slici 8. Neuroni su predstavljeni kružnicama, a veze su predstavljene strelicama. Svakoj vezi dodijeljen je jedan parametar. Glavni nedostatak potpuno povezanih mreža je izuzetno veliki broj parametara što otežava njihovo



Slika 8: Skica neuronske mreže sa dva potpuno povezana skrivena sloja.

obučavanje. Potrebno je mnogo više memorije, iteracija algoritma za obučavanje, ali i podataka za obučavanje. Takođe se veoma često ispostavlja da je veliki dio ovih veza suvišan, tj. da ne doprinose u donošenju odluka, odnosno ne utiču na vrijednost izlaza. Iz tog razloga, uvođene su nove vrste neuronskih mreža kako bi se prevazišli ovi problemi. Potpuno povezani slojevi nisu izbačeni iz upotrebe, oni se i dalje koriste, obično na samom kraju neuronske mreže, kada se umanje dimenzije ulaznih podataka.

5.2 Konvolucione neuronske mreže

Kod potpuno povezanih slojeva neuronskih mreža ulaz se posmatra kao vektor međusobno nezavisnih vrijednosti. Ovo je veoma jaka pretpostavka i često nije tačna. Na primjer, kod slika su susjedni pikseli međusobno povezani jer pripadaju istom objektu. Kod zvuka postoji zavisnost između susjednih vremenskih i frekvencijskih prozora, jer pripadaju istom tonu ili glasu. Ni tekst nije skup nasumično nabacanih slova ili riječi, i tu itekako postoje međusobne zavisnosti.

Obučavanje potpuno povezanih slojeva je potpuno identično bilo da se radi o realnim signalima, bilo o slučajnim permutacijama njihovih odbiraka, jer one ne uzimaju u obzir relacije koje mogu postojati među odbircima ulaznih signala. Osim

toga, multimedijalni podaci su izrazito visokodimenzionalni. Čak i najmanje slike imaju na hiljade piksela. Korišćenje neuronskih mreža sa isključivo potpuno povezanim slojevima bi rezultovalo milijardama parametara. U tako velikom prostoru gotovo je nemoguće, u realnom vremenu sa postojećim računarskim resursima, doći do dobrih vrijednosti parametara. Stoga je potrebno kreirati slojeve neuronske mreže koji prihvataju na ulaz samo mali region ulaznog signala. Ovim načinom kreiranja slojeva međusobna zavisnost ne bi bila u potpunosti zanemarivana, barem ne u tim regionima, a ujedno bi se smanjio broj parametara mreže. Ono što je još neophodno ovim slojevima je prostorna, odnosno vremenska, invarijantnost. Odnosno da pomjeranja u ulaznom signalu, bilo u prostoru ili u vremenu, izazivaju istovjetne promjene i u izlazu tih slojeva. Ovdje se zapravo govori o mogućnosti neurona da se aktiviraju na stimuluse u bilo kom dijelu ulaznog signala. Na primjer, jedan objekat može se nalaziti na različitim mjestima na slici. Jedan ton se može pojaviti na više mjesta u muzičkom snimku. Takođe, riječ može biti izgovorena u bilo kom dijelu audio snimka ili napisana u bilo kom dijelu rečenice. Slojevi neuronske mreže treba da odreaguju na isti način nezavisno od toga u kom dijelu ulaznog signala se javlja stimulus.

Konvolucioni slojevi kreirani su upravo prema ovim zahtjevima, a imajući prvenstveno u vidu sliku kao primjer ulaznog signala i njene odlike [96]. Kao i potpuno povezane neuronske mreže i konvolucione se sastoje od neurona koji računaju ponderisanu sumu svojih ulaza nad kojom se primjenjuje funkcija aktivacije. Razlika je u načinu povezivanja neurona. Kod konvolucionih slojeva, težinski koeficijenti organizuju se u male grupe koje se nazivaju filtri (kerneli). Operacija koja se vrši nad filtrom i ulazom sloja je konvolucija. Odatle je ova vrsta slojeva i dobila svoje ime. Obučavanje konvolucionih slojeva svodi se na traženje optimalnih vrijednosti za koeficijente unutar filtara. Njihovim uvrštavanjem u arhitekturu neuronske mreže dobija se konvoluciona neuronska mreža. Ideja filtriranja slike pomoću kernela primjenom konvolucije utemeljena je i razvijena u oblasti obrade signala, odakle je prenijeta u oblast dubokog učenja.

Konvolucioni filtar se ponaša kao detektor neke karakteristike ulaznog signala. Njime se prepoznaje ista karakteristika na različitim pozicijama u ulazu. Jedan konvolucioni sloj obično se sastoji od više ovakvih filtara kako bi se u istom sloju mogle detektovati različite karakteristike ulaznog signala. Praktikuje se da svi filtri u jednom sloju budu istih dimenzija, iako to nije nužno. Čak i sa povećanim brojem filtara u konvolucionom sloju, broj parametara u ostaje znatno manji u odnosu na potpuno povezane slojeve i ne zavisi od dimenzija ulaza.

Nastankom konvolucionih slojeva napravljen je naizgled otklon od modelovanja neurona u ljudskom mozgu. Ipak, mogu se primijetiti sličnosti sa načinom na koji funkcioniše ljudski vizuelni sistem [97], a koje su razmatrane u toku osmišljavanja

konvolucionih slojeva. Konvoluciona neuronska mreža sastoji se od niza slojeva koji prepoznaju karakteristike ulaznog signala i ta saznanja prosljeđuju narednim slojevima u mreži. U svakom sloju prepoznaju se informacije na različitom nivou apstrakcije. Raniji slojevi prepoznaju jednostavnije koncepte, dok kasniji slojevi koriste akumulirane informacije iz prethodnih slojeva kako bi došli do apstraktnijih karakteristika ulaza. Analizom aktivacija neurona konvolucionih mreža na slikama može se zaključiti da prvi slojevi prepoznaju ivice, sljedeći oblike, pa zatim jednostavne objekte, da bi posljednji slojevi bili u stanju da razaznaju veoma kompleksne objekte. Na sličan način funkcioniše i ljudski vizuelni sistem. Istraživači su otkrili nekoliko regiona u ljudskom vizuelnom korteksu koji na različite načine doprinose vizuelnoj percepciji.

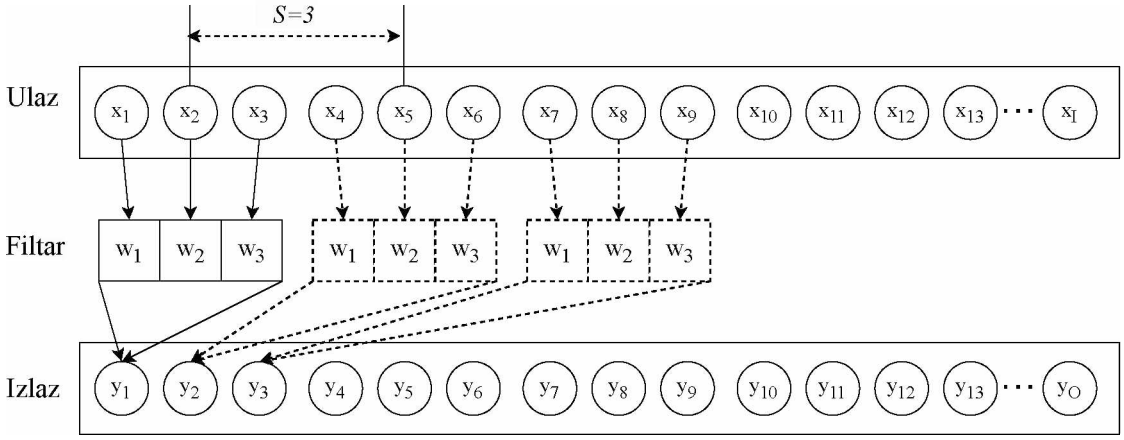
U savremenim neuronskim mrežama konvolucioni slojevi se veoma često koriste u kombinaciji sa potpuno povezanim slojevima. Konvolucioni slojevi nalaze se na početku mreže i koriste se za učenje apstraktnih svojstava ulaznih signala (*engl. feature extraction*) i smanjivanje dimenzija (*engl. dimensionality reduction*), da bi se u posljednjim slojevima mreže nalazili potpuno povezani slojevi za donošenje odgovarajuće odluke.

Konvolucione neuronske mreže primjenjuju se u različitim oblastima sa mnogo uspjeha. Mogu se koristiti sa svim vrstama multimedijalnih signala. Najuspješnija je njihova primjena u domenu kompjuterske vizije gdje daleko nadmašuju konkurenciju u zadacima kao što su klasifikacija slika [98, 99] i prepoznavanje objekata [100, 101]. U sferi obrade zvuka, konvolucioni slojevi korišteni su za pretvaranje govora u tekst [102, 103], ali i neke druge zadatke. Konvolucioni slojevi dominantno su korišteni i u ovom radu za ekstrakciju apstraktnih svojstava audio signala.

5.2.1 Konvolucioni slojevi

Konvolucija je prostorno i vremenski invarijantna operacija, pa stoga ispunjava traženo svojstvo neuronskih mreža. Pravila za izračunavanje konvolucije u slojevima sa jednodimenzionim i dvodimenzionim ulazom objašnjena su u Prilogu C. Isti postupak može se objasniti na drugačiji način. Jedan odbirak konvolucije dobija se računanjem skalarnog proizvoda filtra i jednog parčeta ulaznog signala koje je istih dimenzija kao i filtar. Konačan rezultat se dobija se pomjeranjem filtra dimenzije K signalom, sa korakom S (*engl. stride*), i računanjem skalarnog proizvoda filtra i parčeta signala koje u tom koraku filtar pokriva. Opisani postupak ilustrovan je na Slici 9, na primjeru jednodimenzionog signala i jednog filtra konvolucionog sloja.

Za razliku od potpuno povezanih slojeva, gdje se svaka izlazna vrijednost računa na osnovu zasebnog skupa težinskih koeficijenata kojih ima isto koliko i vrijednosti



Slika 9: Ilustracija izračunavanja konvolucije s filtrom dimenzije $K = 3$ i korakom veličine $S = 3$.

u ulazu sloja, u konvolucionim slojevima isti filtar (skup težinskih koeficijenata) primjenjuje se nad čitavim ulazom. Na ovaj način neophodni broj parametara potreban za računanje izlaza se drastično smanjuje, jer se isti parametri koriste za računanje različitih vrijednosti izlaza. Jedan konvolucioni sloj sadrži više ovakvih filtara čija primjena će proizvesti različite izlaze, koji se zatim prosljeđuju narednim slojevima mreže.

Dimenzije izlaza potpuno povezanog sloja su fiksne i jednake broju neurona u tom sloju. Veličina rezultata konvolucije jednaka je broju različitih pozicija na kojima se može postaviti filtar prilikom pomjeranja signalom. Na ovu vrijednost utiče nekoliko parametara. Oblik rezultata konvolucije zavisi od dimenzija ulaza, dimenzija filtra i veličine koraka pomjeranja S. Sa Slike 9 je očigledno da će primjenom filtra dimenzije $K > 1$ ili koraka $S > 1$ dimenzije rezultata konvolucije biti manje od dimenzija ulaznog signala. U nekim situacijama je ovo poželjno, dok u nekima nije, pa se tada može vršiti dopunjavanje ulaza nulama (*engl. zero padding*) kako bi se očuvale dimenzije signala i nakon primjene konvolucije. Ukoliko je jednodimenzioni signal sa N odbiraka dopunjen sa po D nula na početku i kraju, broj odbiraka njegove konvolucije sa filtrom dimenzije K, sa korakom S izračunava se na sljedeći način:

$$O = \left\lfloor \frac{N + 2D - K}{S} \right\rfloor + 1 \tag{53}$$

Nakon konvolucije filtra sa ulaznim signalom rezultatu se dodaje i vrijednost slobodnog (*bias*) parametra. Svaki filtar u konvolucionom sloju ima poseban *bias* parametar. Što je veća vrijednost ovog parametra to neuronska mreža daje više značaja rezultatu tog filtra. Posljednja operacija koja se vrši u konvolucionom sloju je primjena odabrane aktivacione funkcija nad izlazima svih filtara. Kao rezultat ovih

operacija dobija se tzv. mapa karakteristika (*engl. feature map*). Jedna vrijednost u ovoj mapi predstavlja nivo aktivacije filtra u dijelu ulaznog signala koji je posmatran tom prilikom. Na ovaj način, svaki filtar prepoznaje stimuluse u malim regionima ulaznog signala, poželjno različite od ostalih filtara.

Ulaz i izlaz konvolucionog sloja mogu sadržati više kanala. Kanal je konvencionalni termin koji predstavlja jednu komponentu signala. Jedan kanal se može posmatrati kao signal sam za sebe. Na primjer, za reprezentaciju slika u boji koriste se tri ili četiri kanala, zavisno od modela boja. Stereo zvuk se predstavlja sa dva kanala, po jedan za lijevi i desni zvučnik. Takođe, mape karakteristika koje se dobijaju kao izlaz konvolucionih slojeva mogu se posmatrati kao posebni kanali u ulazima dubljih slojeva mreže.

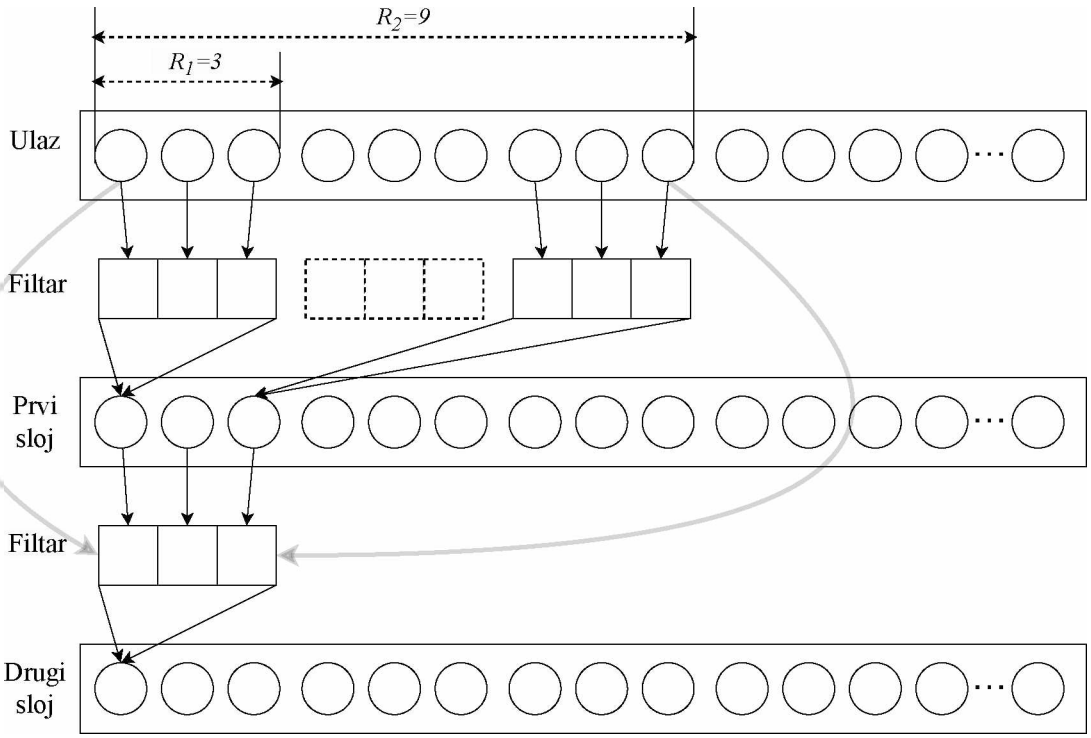
Postojanje više kanala uvodi dodatnu dimenziju za ulaze konvolucionih slojeva u mreži. Međutim, po ovoj dimenziji se ne vrši dijeljenje parametara (filtara), već se svakom filtru konvolucionog sloja dodaje nova dimenzija čija je veličina jednaka broju kanala u ulazu. Ovako prošireni filtri se zatim konvoluiraju sa višekanalnim ulazom. Dodavanje nove dimenzije filtrima za posljedicu ima povećanje broja parametara u sloju C puta, gdje je C broj kanala u reprezentaciji ulaznog signala.

5.2.2 Transponovana konvolucija

Transponovana konvolucija je operacija kojom se rekonstruišu dimenzije ulaza konvolucije. Ulaz ove operacije se posmatra kao rezultat konvolucije. Transponovanom konvolucijom se uz dati filtar dobija izlaz koji ima isti oblik kao ulaz te inicijalne konvolucije. Transponovanu konvoluciju ne treba poistovjećivati sa inverznom konvolucijom, tzv. dekonvolucijom. Za razliku od dekonvolucije, transponovana konvolucija ne rekonstruiše vrijednosti ulaza, već samo njegov oblik. U praksi se sve češće koristi za projekciju mapa karakteristika u višedimenzione prostore.

Konvolucija i transponovana konvolucija su kompatibilne operacije. Transponovana konvolucija se može posmatrati kao konvolucija sa korakom pomjeranja $S_T < 1$ (*engl. fractionally strided convolution*), te se ona može izvesti konvolucijom datog filtra i ulaza, transformisanog na odgovarajući način kako bi se simulirao korak pomjeranja $S_T < 1$. Da bi se ovo izvelo, potrebno je između svih odbiraka u ulazu dodati po $S - 1$ nula, gdje je S korak s kojim je vršena inicijalna konvolucija. Na ovaj način će se filtar „sporije" pomjerati pri računanju transponovane konvolucije. U praksi se koriste efikasnije matrične operacije za izračunavanje transponovane konvolucije.

Dimenzije izlaza transponovane konvolucije su nešto komplikovanije za procije-



Slika 10: Receptivna polja filtara konvolucione neuronske mreže.

niti. Za konvoluciju izvršenu sa filtrom dimenzije K, korakom S i ulazom od N odbiraka sa dodatih D nula na početku i kraju, odgovarajuća transponovana konvolucija vrši se sa filtrom dimenzije $K_T = K$, korakom $S_T = 1$ i dodavanjem po $D_T = K - P - 1$ nula na početku i kraju ulaza. Ukoliko se ona vrši nad ulazom od N_T odbiraka, dobija se rezultat dimenzije:

$$O_T = S((N_T - 1)S - 1) + D'_T + K - 2D, \tag{54}$$

gdje je $D'_T = (N + 2D - K) \bmod S$, broj nula koji se, pored D_T nula, dodaje na kraj ulaza transponovane konvolucije kako bi se u potpunosti rekonstruisale dimenzije ulaza inicijalne konvolucije.

5.2.3 Receptivno polje konvolucionog filtra

Receptivno polje filtra čine sve vrijednosti u inicijalnom ulazu mreže koje utiču na izlaz tog filtra. Ovo je jedan od najznačajnijih koncepata u konvolucionim neuronskim mrežama i posvećuje mu se posebna pažnja prilikom dizajniranja arhitekture modela. Skice receptivnih polja filtara za prva dva sloja konvolucione neuronske mreže date su na Slici 10. Što se filtar nalazi dublje u konvolucionoj mreži, veće mu je receptivno polje, što se vidi na primjeru 1D konvolucije sa slike. Ovo je veoma

važno, jer na taj način filtri u dubljim slojevima na osnovu šireg konteksta mogu prepoznavati kompleksnija svojstva ulaznih signala.

Veličina receptivnog polja filtra zavisi od njegove veličine, od veličine receptivnog polja filtra iz prethodnog sloja i od koraka pomjeranja. Ako je receptivno polje filtra u i-tom sloju mreže R_i, receptivno polja filtra u $i+1$-om sloju mreže obuhvata:

$$R_{i+1} = R_i + (K_{i+1} - 1) \prod_{j=1}^{i-1} S_j \tag{55}$$

odbiraka inicijalnog ulaza, s izuzetkom filtara iz prvog konvolucionog sloja, čije je receptivno polje jednako veličini filtra. Vrijednost S_j predstavlja veličinu koraka pomjeranja u j-tom konvolucionom sloju mreže, a K_{i+1} je veličina filtra iz $i+1$-og sloja. Dakle, receptivno polje se eksponencijalno povećava i u svakom sloju mreže ono treba da bude pažljivo podešeno kako bi se mogla procijeniti uloga filtara tog sloja.

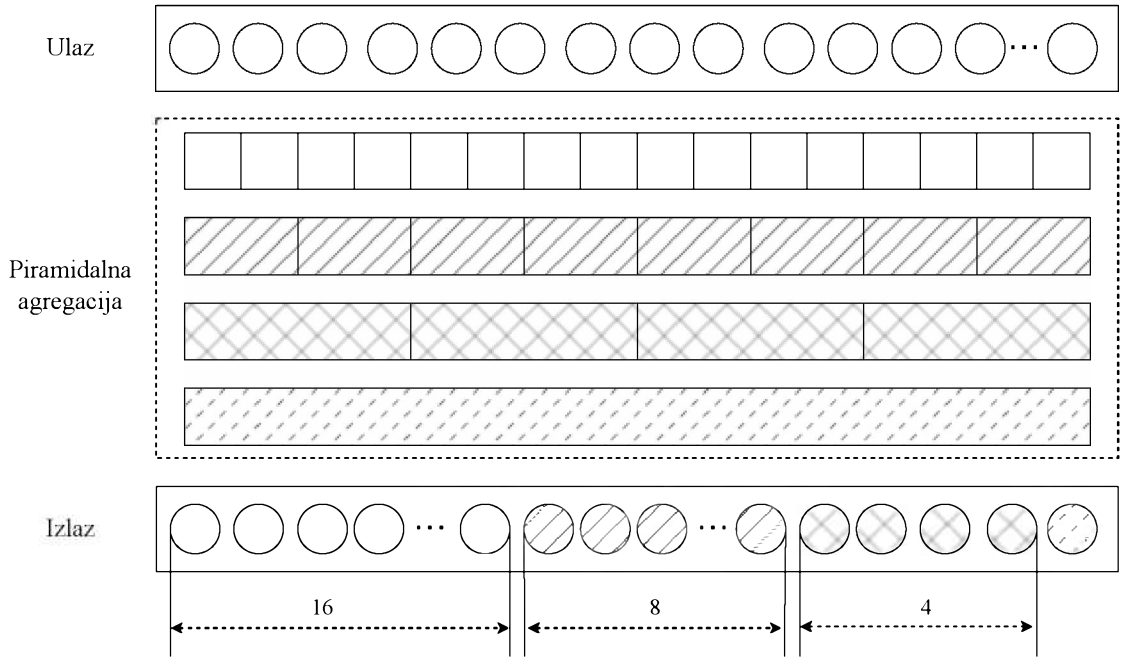
5.2.4 Smanjivanje dimenzija

U neuronskim mrežama često je potrebno da dimenzije izlaznog sloja mreže budu manje od dimenzija ulaza. Na primjer, kada se vrši klasifikacija ulaznih signala u jednu od kategorija čiji je broj unaprijed poznat. Smanjenje dimenzionalnosti je potrebno i kako bi se broj parametara mreže držao ograničenim. Apstraktne karakteristike ulaznog signala je teže prepoznati, pa se u dubljim slojevima neuronskih mreža često koristiti veći broj konvolucionih filtara kako bi se ove karakteristike detektovale. Primjenjivanje filtara nad ulazom originalnih dimenzija dovelo bi do značajnog povećanja broja parametara u kasnijim slojevima. Smanjenjem dimenzija ulaza se ovaj efekat suzbija.

Iz jednakosti (53) može se zaključiti da se smanjenje dimenzija može postići povećavanjem koraka S. Izražen efekat smanjenja dimenzija postiže se i korišćenjem agregacionih slojeva (*engl. pooling*). Ovi slojevi primjenom agregatne funkcije (na primjer, prosjeka ili maksimuma) smanjuju dimenzije ulaza na način što više vrijednosti iz mape karakteristika agregiraju u jednu. Agregacioni slojevi se mogu posmatrati i kao konvolucioni slojevi sa jednim filtrom, gdje su vrijednosti u filtru fiksirane i ne primjenjuje se aktivaciona funkcija na izlazu filtra.

5.2.5 Piramidalna agregacija

Operacija konvolucije, opisana u Prilogu C, može se primijeniti na ulazima proizvoljnih dimenzija. Dimenzije izlaza mogu se izračunati pomoću (53). Međutim,



Slika 11: Skica strukture sloja za piramidalnu agregaciju sa 4 nivoa rezolucije.

neuronske mreže obično imaju izlaze fiksne veličine. Takođe, veliki broj arhitektura neuronskih mreža nakon konvolucionih slojeva, koji služe za izvlačenje karakteristika ulaznog signala, sadrže nekoliko potpuno povezanih slojeva koji daju konačan izlaz. Za razliku od konvolucionih slojeva, dimenzije ulaza potpuno povezanih slojeva ne mogu varirati.

Piramidalna agregacija [104] je operacija koja iz ulaza proizvoljnih dimenzija izvlači (agregira) karakteristike fiksnih dimenzija. Ovaj postupak se sprovodi tako što se ulaz dijeli na fiksni broj segmenata nad kojima se zatim primjenjuje neka od tradicionalnih agregatnih funkcija, poput maksimuma ili prosjeka. Agregiranje se obično vrši na nekoliko različitih nivoa rezolucije, pa je otuda ova agregacija nazvana piramidalnom. Nerijetko se na jednom nivou vrši i globalna agregacija, odnosno primjena odabrane agregatne funkcije nad čitavim ulazom, koja kao izlaz daje samo jednu vrijednost. Na kraju se izlazi sa svih nivoa rezolucije nadovezuju i prosljeđuju narednim slojevima u mreži, koji su obično potpuno povezani. Upotrebom piramidalne agregacije se otklanja potreba da ulazi neuronske mreže budu fiksnih dimenzija.

Primjer sloja piramidalne agregacije prikazan je na Slici 11. Agregiranje se vrši na 4 nivoa rezolucije. Na svakom nivou se veličina regiona za agregiranje prilagođava tako da se signal podijeli na predefinisani broj segmenata. Prvi nivo agregacije dijeli signal na 16 segmenata, što rezultuje sa 16 vrijednosti u izlazu. Naredni nivoi su

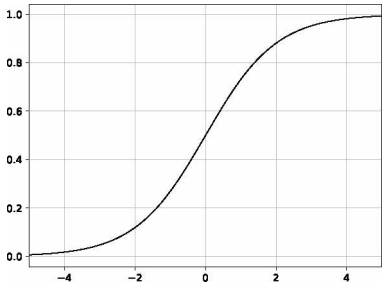
manje rezolucije, jer se dijeljenje vrši na 8 i 4 segmenta, respektivno. Posljednja je globalna agregacija koja daje jednu vrijednost na izlazu. Na slici su nivo piramidalne agregacije i njemu odgovarajuće vrijednosti u izlaznom vektoru označeni istim obrascem. Izlazi sa sva četiri nivoa se nadovezuju i time se dobija vektor sa $16 + 8 + 4 + 1 = 29$ elemenata. Ovaj vektor imao bi isti broj elemenata, bez obzira na to koje su dimenzije ulaza.

5.3 Aktivacione funkcije

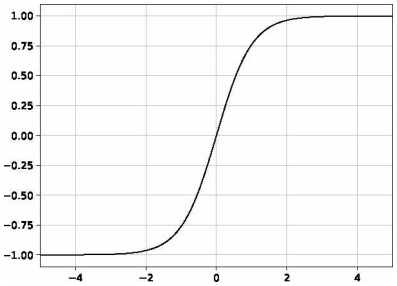
Primjena aktivacione funkcije je posljednja operacija koja se vrši u jednom sloju neuronske mreže. Ona određuje u kojoj mjeri će se neuron pobuditi za date ulaze i po pravilu je ista za sve neurone u sloju. Veća vrijednost sume ulaza upućuje da se dobijeni ulaz dobro poklapa sa težinskim koeficijentima, pa neuron treba snažnije da se aktivira. Dakle, aktivaciona funkcija treba da bude rastuća. Pored toga, za pozitivne ulaze treba da bude pozitivna, a za negativne manja ili jednaka nuli. Aktivacione funkcije su obično i diferencijabilne jer se neuronske mreže najčešće obučavaju metodama zasnovanim na gradijentnom spustu, pa je stoga potrebno izračunavati prvi, a nekada i drugi izvod aktivacione funkcije.

Veoma je važno i da većina aktivacionih funkcija u neuronskoj mreži bude nelinearna. Ukoliko ne bi koristili nelinearne aktivacione funkcije, neuronska mreža predstavljala bi kompoziciju linearnih funkcija, što je takođe linearna funkcija. Ovime bi se izrazito ugrozila sposobnost neuronskih mreža da rješavaju različite zadatke, jer bi bile ograničene na modelovanje samo jedne vrste funkcija. Uvođenjem nelinearnosti, neuronske mreže mogu modelovati, odnosno aproksimirati veoma složene, zapravo proizvoljne funkcije.

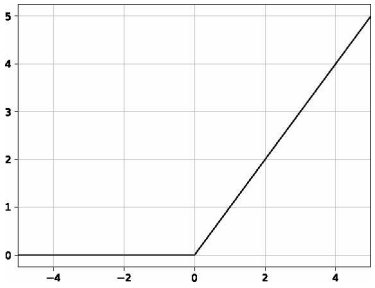
Postoji veliki broj funkcija koje ispunjavaju gore pomenute uslove i koje bi se mogle koristiti kao aktivacione. Istraživači su, tokom godina, predložili veliki broj funkcija koje bi se mogle koristiti u ovu svrhu. Neke su se pokazale uspješnijim od ostalih i njima će u ovom poglavlju biti posvećena pažnja. Odzivi svih razmatranih aktivacionih funkcija dati su na Slici 12. Međutim, ni izbor među tim aktivacionim funkcijama nije jednostavan i predstavlja jedan veoma važan korak prilikom dizajniranja neuronske mreže. Aktivaciona funkcija za izlazni sloj mreže mora biti odabrana u skladu sa vrijednostima koje mreža treba da produkuje. U skrivenim slojevima ovaj zadatak je još komplikovaniji jer izbor aktivacione funkcije može značajno da utiče na sposobnost tih slojeva da uče.



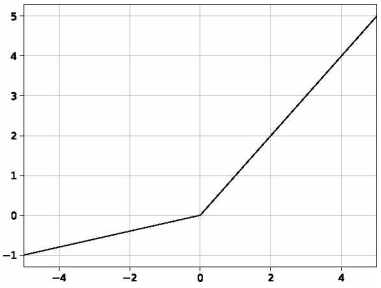
(a) Sigmoid



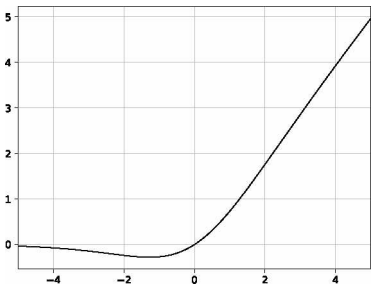
(b) Hiperbolički tangens



(c) ReLU



(d) Propustljiva ReLU ($\upsilon = 2$)



(e) *Swish* ($\upsilon' = 1$)

Slika 12: Odzivi aktivacionih funkcija.

5.3.1 Funkcija identiteta

Funkcija identiteta $f(x) = x$ je najjednostavniji oblik aktivacione funkcije. Ona zadovoljava sva svojstva koja se očekuje da aktivacione funkcije ispune, osim svojstva nelinearnosti. Iz tog razloga se uglavnom ne koristi u skrivenim slojevima neuronskih mreža, već samo u izlaznom sloju kada mreža na izlazu treba da produkuje vrijednosti koje su u neograničenom intervalu $(-\infty, +\infty)$.

5.3.2 Sigmoid

Sigmoid je aktivaciona funkcija koja je karakteristična po svom obliku koji liči na latinično slovo „S". Definiše se sljedećom jednakošću:

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{56}$$

Očigledno se radi o nelinearnoj funkciji. Diferencijabilna je i izvod se računa na jednostavan način $\frac{df}{dx}(x) = f(x)(1 - f(x))$.

Sigmoid funkcija ograničava izlazne vrijednosti neurona na interval $(0, 1)$. Ona izuzetno dobro interpretira aktivaciju bioloških neurona koji se mogu nalaziti u dva stanja, neaktivnom, tj. prigušenom i aktivnom, odnosno pobuđenom. Vrijednost sigmoid funkcije može se tumačiti kao vjerovatnoća da je odgovarajući neuron pobuđen. Ovo je bio motiv za uvođenje sigmoid funkcije i razlog zbog kojeg je ona dugo bila najzastupljenija aktivaciona funkcija u neuronskim mrežama. Međutim, vremenom su prepoznata praktična ograničenja ove funkcije. Kod sigmoid funkcije veoma je izražen problem isčezavanja gradijenata (*engl. vanishing gradients*), o kojem će biti više riječi u Sekciji 5.5.3. Ova funkcija se brzo približava nuli na jednom, a jedinici na drugom kraju, da gotovo postane konstantna. Zbog toga su izvodi te funkcije u većini tačaka mali, skoro jednaki nuli što značajno usporava konvergenciju mreže jer se vrijednosti parametara u jednom koraku algoritma obučavanja neznatno mijenjaju. Otežavajuća okolnost za optimizaciju je i što vrijednosti funkcije nisu centrirane u nuli, već u 0.5. Takođe, sigmoid funkcija je računski skupa jer uključuje računanje vrijednosti e^{-x}. Zbog navedenih nedostataka sigmoid funkcija se rijetko koristi u skrivenim slojevima mreže. Još uvijek se koristi u LSTM ćelijama rekurentnih neuronskih mreža. Upotrebljava se i u izlaznim slojevima kada je izlaz iz mreže vjerovatnoća, što je slučaj kod zadataka klasifikacije. *Softmax* funkcija je uopštenje sigmoid funkcije i koristi se u izlaznom sloju mreže za višeklasnu klasifikaciju. U ovom slučaju $f : \mathbb{R}^{N_c} \to (0, 1)^{N_c}$, gdje je N_c ukupni broj klasa. Posljednji sloj mreže dizajnira se tako da proizvede po jedan realan broj za svaku od klasa. Nad tim

vektorom brojeva se zatim primjenjuje sljedeća funkcija:

$$f_i(x) = \frac{e^{x_i}}{\sum_{j=1}^{N_c} e^{x_j}}, \tag{57}$$

gdje je $x = (x_1, x_2, ..., x_{N_c})$ ulazni vektor, a $f_i(x)$ vjerovatnoća da ulazni signal pripada i-toj klasi, $i = 1, 2, ..., N_c$.

5.3.3 Hiperbolički tangens

Hiperbolički tangens (tanh) je funkcija slična sigmoidu. Ona je skalirana i pomjerena verzija sigmoid funkcije i ima sličan „S" oblik. Definisana je sa:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{58}$$

Ova aktivaciona funkcija uvedena je kako bi se prevazišli nedostaci sigmoida. U početku je njena upotreba bila intenzivna, ali su je u međuvremenu neke druge funkcije prevazišle. Za razliku od sigmoid funkcije, hiperbolički tangens je ograničen na interval $(-1, 1)$ i centriran je u nuli, pa je manja šansa da se mreža „zaglavi" u toku treninga. Ova funkcija ima nešto strmije gradijente, ali i dalje postoji problem nestajućih gradijenata. Takođe je računski zahtjevna zbog članova e^x i e^{-x}. Iz ovih razloga, isto kao i sigmoid, ne koristi se u skrivenim slojevima mreže, osim u rekurentnim neuronskim mrežama i u izlaznim slojevima.

5.3.4 ReLU

ReLU (*engl. Rectified Linear Unit*) je dobra alternativa sigmoidu i hiperboličkom tangensu. Uvođenje ove funkcije bilo je jedno od najvažnijih otkrića u dubokom učenju [105]. ReLU funkcija je veoma jednostavna:

$$f(x) = \max(0, x) \tag{59}$$

Sastoji se iz dvije linearne komponente. Za negativne ulaze vrijednost funkcije je jednaka nuli, a za pozitivne ulaze ReLU je funkcija identiteta. Iako je čine dvije linearne funkcije, ReLU funkcija je nelinearna. Vrijednosti su joj u intervalu $(0, +\infty)$. Ova funkcija nema toliko izražen problem sa isčezavanjem gradijenata i nije računski zahtjevna jer nema eksponencijalnih članova, pa je konvergencija mreža sa ovom aktivacionom funkcijom znatno efikasnija. Takođe, ReLU ima konstantan izvod u oba dijela x ose. Za $x \leq 0$ izvod je jednak 0, a za $x > 0$ izvod je 1, što takođe ubrzava izračunavanja. ReLU je, zbog svih ovih prednosti, zamijenila funkcije sigmoid i

hiperbolički tangens kao podrazumijevana aktivaciona funkcija u potpuno povezanim i konvolucionim slojevima neuronskih mreža. Većina najznačajnijih dostignuća u dubokom učenju postignuta je uz korišćenje upravo ove aktivacione funkcije. Ipak, uprkos izuzetno širokoj i uspješnoj primjeni i ReLU funkciji uočene su mane. Ova funkcija nije diferencijabilna u nuli i nije centrirana u nuli. Takođe nije ograničena odozgo, pa njene vrijednosti mogu biti izuzetno velike i izazvati numeričke nestabilnosti prilikom izračunavanja i reprezentacije u računaru. Najveći problem ReLu funkcije je tzv. problem umiranja neurona (*engl. dying ReLU problem*). U zavisnosti od inicijalnih vrijednosti parametara, neuroni mogu doći u stanje u kojem postanu neaktivni za sve moguće ulaze. Ovo se može desiti zbog toga što je za negativne vrijednosti ulaza ReLU funkcija jednaka nuli. Kako je i izvod u tim tačkama jednak nuli, vrijednosti parametara ne mogu se promijeniti i neuron će ostati „zaglavljen" u ovom stanju. Neurone koji se nađu u ovoj situaciji, smatramo umrlim. Ovaj efekat može se tolerisati ako je procenat neaktivnih neurona mali, jer moderne neuronske mreže sadrže obilje neurona. Međutim, ukoliko se ovo desi velikom broju neurona, značajno se smanjuje kapacitet modela i to postaje ozbiljan problem. Zbog svega navedenog, u poslednje vrijeme počele su se pojavljivati funkcije koje prijete da zamijene ReLU na mjestu najzastupljenije aktivacione funkcije u skrivenim slojevima neuronskih mreža.

5.3.5 Propustljiva i parametrizovana ReLU

Problem umiranja neurona koji postoji kod ReLU aktivacione funkcije može se ublažiti uvođenjem blagog nagiba u negativnom dijelu x ose. Propustljiva ReLU funkcija (*engl. leaky ReLU*)[106] je upravo takva modifikacija ReLU-a i definisana je sljedećom jednakošću:

$$f(x) = \max(vx, x). \tag{60}$$

Parametar v se zadaje manuelno i to je mala konstanta vrijednost (npr. 0.01).

Parametrizovana ReLU funkcija (PReLU)[107] ovu ideju odvodi još jedan korak dalje. Za razliku od propustljive ReLU parametar v nije unaprijed zadat, već je to još jedan od parametara čija se vrijednost traži učenjem.

Ove dvije funkcije mogu se upotrebljavati na svim mjestima umjesto ReLU funkcije. Iako naizgled superiornije od ReLU funkcije jer ne izazivaju problem umiranja neurona, još uvijek nisu u potpunosti izbacile ReLU iz upotrebe. Korišćenjem ovih funkcija uvodi se dodatno otežanje u vidu traženja optimalne vrijednosti za parametar v što usporava proces obučavanja mreže. Ukoliko se ne izabere dobra vrijednost za v mogu se značajno ugroziti performanse modela.

5.3.6 *Swish* funkcija

Swish funkcija dobija se množenjem sigmoid funkcije sa ulazom x:

$$f(x) = \frac{x}{1 + e^{-v'x}}. \tag{61}$$

Ovo je nelinarna funkcija čiji je oblik sličan ReLU funkciji, pa nema problema isčezavanja gradijenata. Prema definiciji iz rada [108], funkcija *swish* sadrži parametar v', čija se vrijednost uči, poput parametra v PReLU funkcije. Međutim, u gotovo svim praktičnim realizacijama uzima se $v' = 1$ i većina eksperimenata sprovodi se s tom verzijom funkcije [109]. U radu [109] se pokazuje da se korišćenjem *swish* funkcije umjesto ReLU-a mogu popraviti performanse neuronskih mreža na nizu najvažnijih zadataka. Njena jednostavnost i sličnost ReLU funkciji čine ovu zamjenu lakom, jer nema potrebe za uvođenjem dodatnih parametara i podešavanjem njihovih vrijednosti. Jedina mana *swish* funkcije je računska složenost zbog člana e^{-x} koji se ovdje opet pojavljuje.

5.4 Obučavanje neuronskih mreža

Svrha algoritama mašinskog učenja, a time i dubokih neuronskih mreža je da na optimalan način modeluju zavisnosti u određenom skupu podataka ili raspodjelu iz koje su primjerci toga skupa uzorkovani. Na osnovu dostupnih primjeraka, vrši se prilagođavanje parametara modela. Ovaj postupak naziva se obučavanje, a skup koji se pritom koristi je analogno nazvan skup za obučavanje. Iako se parametri modela prilagođavaju skupu za obučavanje, cilj ovog procesa je generalizacija, odnosno postizanje dobrih performansi modela na novim, prethodno neviđenim, primjercima, uzorkovanim iz iste raspodjele kao i primjerci za obučavanje.

5.4.1 Funkcija gubitka

Da bi se procedura obučavanja mogla uspješno sprovoditi, odnosno da bi model vremenom poboljšavao performanse, mora postojati način da se izmjeri greška koju on u određenom trenutku obučavanja pravi, tj. koliko se razlikuju trenutni izlazi modela od željenih vrijednosti. Ovu grešku je zatim potrebno kao povratnu informaciju proslijediti modelu i na odgovarajući način prilagoditi svaki njegov parametar kako bi se greška smanjila. U ovu svrhu u dubokom učenju koriste se funkcije gubitka (*engl. loss functions*).

Jedna postavka vrijednosti svih parametara u neuronskoj mreži predstavlja jednu tačku u tzv. parametarskom prostoru modela. Funkcija gubitka preslikava tačku

iz parametarskog prostora u skalarnu vrijednost kojom se kvantifikuje uspješnost u modelovanju primjeraka iz skupa za obučavanje. Poboljšanja ove vrijednosti vode proceduru obučavanja ka dostizanju boljeg modela. Vrijednost funkcije gubitka u osnovi predstavlja prosječnu mjeru razlike između očekivanih i stvarnih vrijednosti koje model sa datom postavkom parametara daje na izlazu, odnosno prosječnu grešku modela. Ipak, funkcije gubitka treba razlikovati od metrika, jer njihove vrijednosti ne moraju biti jednostavne za interpretirati, već služe za usmjeravanje procedure obučavanja. Funkcije gubitka ne moraju isključivo sadržati ocjenu greške modela na skupu za obučavanje, već i druge članove koji će poslužiti u poboljšavanju performansi modela. Na primjer, korišćenje funkcije gubitka kojom se isključivo ocjenjuje greška modela na skupu za obučavanje može izazvati efekat preprilagođavanja (*engl. overfitting*). Model u tom slučaju postaje previše specijalizovan za podatke korišćene u obučavanju i apsorbuje šumove i druge nepoželjne karakteristike podataka, što rezultuje lošim performansama na novim podacima, odnosno inferiornom sposobnošću generalizacije. Stoga, funkcije gubitka nerijetko proširuju članovima koji služe sprečavanju ove pojave.

Način definisanja funkcije gubitka u najvećoj mjeri zavisi od vrste zadatka za koji se pokušava napraviti model. Uopšteno gledano, modelom mašinskog učenja se, za date vrijednosti ulaznih promjenljivih, pokušavaju proizvesti izlazi koji odgovaraju raspodjeli ciljnih promjenljivih. Stoga se funkcija gubitka u osnovi može posmatrati kao mjera sličnosti raspodjele izlaza modela i raspodjele iz koje su uzorkovane ciljne vrijednosti iz skupa za obučavanje. Razlika, a samim tim i sličnost, dvije raspodjele mjeri se unakrsnom entropijom (*engl. cross-entropy* - CE). Opšta definicija unakrsne entropije za raspodjele $\mathcal{Q}$ i $\mathcal{Q}'$ je:

$$\mathrm{CE}(\mathcal{Q}, \mathcal{Q}') = \mathbb{E}_{\zeta \sim \mathcal{Q}(\zeta)} \left[\log \left(\mathcal{Q}'(\zeta) \right) \right] = \int \mathcal{Q}(\zeta) \log \left(\mathcal{Q}'(\zeta) \right) d\zeta. \tag{62}$$

U mašinskom učenju $\mathcal{Q}$ je stvarna (empirijska) raspodjela, a $\mathcal{Q}'$ je raspodjela koju daje model. Model mašinskog učenja gotovo nikada ne može u potpunosti reprodukovati raspodjelu $\mathcal{Q}$ jer ona uglavnom nije poznata, već je model aproksimira na osnovu primjeraka iz skupa za obučavanje.

Ako je zadatak modela klasifikacija, odnosno dodjeljivanje primjeraka iz skupa odgovarajućoj klasi, on se modeluje tako što se na izlazu za svaku od klasa procjenjuje vjerovatnoća da joj neki primjerak iz skupa pripada. Unakrsna entropija se tada računa kao:

$$J_{CE}(\Theta) = \frac{1}{M} \sum_{m=1}^{M} \sum_{c=1}^{N_c} y_c^{(m)} \log \left(\hat{y}_c^{(m)}(\Theta) \right). \tag{63}$$

M je broj primjeraka u uzorku za obučavanje, a N_c je ukupan broj klasa. Vrijednost $y_c^{(m)}$ označava ispravnu vjerovatnoću. Ova vrijednost je jednaka 0, ako primjerak m

ne pripada klasi c, odnosno 1 ako pripada. Vrijednost $\hat{y}_c^{(m)}(\Theta)$ predstavlja vjerovatnoću da primjerak m pripada klasi c, generisanu modelom sa vektorom parametara Θ. Ova vrijednost je takođe iz segmenta $[0, 1]$.

Najčešće korištena funkcija gubitka u regresionim zadacima je srednja kvadratna greška (*engl. mean squared error* - MSE). Računa se na jednostavan način:

$$J_{\text{MSE}}(\Theta) = \frac{1}{M} \sum_{m=1}^{M} \left(y^{(m)} - \hat{y}^{(m)}(\Theta) \right)^2 . \tag{64}$$

Vrijednosti $y^{(m)}$ i $\hat{y}^{(m)}(\Theta)$ su očekivani izlaz i izlaz modela za m-ti primjerak iz skupa, respektivno. Pod pretpostavkom da ciljna promjenljiva ima Gausovu raspodjelu, srednja kvadratna greška može se posmatrati kao unakrsna entropija te raspodjele i raspodjele izlaza modela.

Srednja apsolutna greška (*engl. mean absolute error* - MAE) je takođe veoma zastupljena u regresiji:

$$J_{\text{MAE}} = \frac{1}{M} \sum_{m=1}^{M} |y^{(m)} - \hat{y}^{(m)}(\Theta)|. \tag{65}$$

Ova funkcija se, kao i srednja kvadratna greška, može protumačiti kao unakrsna entropija dvije raspodjele.

Dakle, iako nominalno postoje različite funkcije gubitka za različite zadatke modela mašinskog učenja, može se smatrati da se u obučavanju modela kao osnova za funkciju gubitka univerzalno koristi unakrsna entropija.

5.4.2 Gradijentni spust

Obučavanje modela mašinskog učenja, odnosno neuronskih mreža, sastoji se u traženju vrijednosti parametara modela kojima bi se minimizovala vrijednost funkcije gubitka na skupu za obučavanje. Dakle, obučavanje neuronskih mreža je optimizacioni zadatak. Kompleksnost neuronskih mreža prouzrokuje da njihove funkcije gubitka budu veoma složene i nekonveksne. Trenutno ne postoji algoritam koji može garantovati pronalazak globalnog minimuma takvih funkcija. Za obučavanje se koriste iterativni algoritmi koji postepeno smanjuju vrijednost funkcije gubitka na skupu za obučavanje do veoma malih, ali ne garantovano i najmanjih mogućih vrijednosti. Parametri neuronske mreže su na početku slučajno zadati, pa se u uzastopnim iteracijama algoritma mijenjaju u smjeru smanjivanja greške koju neuronska mreža pravi.

Optimalne vrijednosti parametara neuronske mreže traže se u neprekidnom prostoru. Jedan način pretraživanja takvog prostora je njegova diskretizacija, a potom

primjena neke od tehnika lokalnog traženja na diskretnom prostoru. Kontinualni prostor preslikava se u koordinatnu mrežu određene rezolucije. Ova mreža formira se tako što se po svakoj od dimenzija (za svaki od parametara) odabere vektor pomjeranja Δ i iz kontunualnog prostora se odabiraju tačke sa tim fiksnim korakom. Drugi postupak kojim bi se moglo kretati po prostoru parametara je uzorkovanje skupa tačaka (vrijednosti parametara) iz Δ-okoline trenutne tačke i pomjeranje ka onoj tački koja najviše umanjuje vrijednost funkcije gubitka. Ovi pristupi imaju nekoliko nedostataka. Skup tačaka koje se na ovaj način mogu pretražiti je ograničen. Ishod optimizacionog algoritma umnogome zavisi od izbora vrijednosti za Δ. Mana ovog pristupa je i što su svi djelovi prostora reprezentovani na isti način, sa istom rezolucijom. Vrijednosti funkcije gubitka nisu ravnomjerne u svim djelovima prostora za pretragu. Djelove prostora sa manjim vrijednostima funkcije gubitka treba pretraživati sa većom rezolucijom, dok se u lošijim djelovima mogu praviti veći koraci bez pokrivanja velikog broja tačaka.

Praćenje promjena vrijednosti funkcije u obližnjim tačkama i kretanje u smjeru najboljeg progresa, na čemu su zasnovane pomenute metode, jednim imenom se naziva empirijski gradijent. Gradijent funkcije $f : \mathbb{R}^n \to \mathbb{R}$ inače predstavlja vektor parcijalnih izvoda te funkcije po svakom od parametara:

$$\nabla_x f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), ..., \frac{\partial f}{\partial x_n}(x) \right), \tag{66}$$

gdje je $x = (x_1, x_2, ..., x_n)$. Vrijednost gradijenta daje veličinu i pravac najstrmije padine po svakoj od dimenzija. Na osnovu tih vrijednosti može se usmjeravati algoritam za traženje ekstremuma funkcije. Kako je funkcija gubitka za neuronsku mrežu obično dostupna u analitičkom obliku, mogu se računati njeni parcijalni izvodi, odnosno gradijent. Obučavanje mreže se na ovaj način može sprovoditi analitičkim, a ne empirijskim postupkom. Iz ovog razloga se za obučavanje neuronskih mreža u praksi ustalila grupa algoritama zasnovanih na gradijentnom spustu. U osnovi tih algoritama je sljedeće pravilo:

$$\Theta^t \leftarrow \Theta^{t-1} - \nu \nabla_\Theta J(\Theta^{t-1}), \tag{67}$$

gdje je Θ^t vektor parametara neuronske mreže u iteraciji t, J funkcija gubitka, a ν hiperparametar koji se naziva stopa učenja (*engl. learning rate*). Pravilo (67) govori da se korak u cilju minimizacije funkcije gubitka treba napraviti u smjeru suprotnom od gradijenta, jer je to smjer u kojem će vrijednost funkcije najviše opasti. Veličina tog koraka kontroliše se vrijednošću hiperparametra ν, koja se ručno inicijalizuje, ali se može mijenjati u toku treninga.

Funkcija gubitka, poput i same neuronske mreže, je zapravo kompozicija velikog broja funkcija. Prilikom računanja gradijenta te funkcije, odnosno parcijalnih izvoda,

mora se primjenjivati pravilo za računanje izvoda kompozicije funkcija (*engl. chain rule*):

$$\frac{\partial(f \circ g)}{\partial x} = \frac{\partial f}{\partial g}\frac{\partial g}{\partial x}. \tag{68}$$

Vrijednost funkcije gubitka, odnosno greške koju pravi neuronska mreža, se izračunava na osnovu razlike između očekivanih izlaza i izlaza neuronske mreže. U tom procesu podaci se prosljeđuju od ranijih slojeva mreže ka kasnijim. Nasuprot tome, u procesu računanja parcijalnih izvoda po parametrima mreže, na osnovu pravila (68), podaci o grešci koju model pravi se od čvorova izlaznog sloja prosljeđuje čvorovima prethodnih slojeva. Iz tog razloga se ovaj algoritam za računanje gradijenta naziva algoritam propagacije unazad. Korišćenje ovog pristupa za obučavanje neuronskih mreža prvi put je predloženo u radu [91].

Da bi se izračunali parcijalni izvodi po parametrima proizvoljnog čvora a u mreži, najprije je potrebno izmjeriti koliki je doprinos tog čvora u grešci koji pravi model. Taj doprinos je parcijalni izvod funkcije gubitka po izlazu čvora a, tj. $\partial J/\partial y_a$, gdje je y_a izlaz čvora a.

Greška koju pravi čvor a utiče na greške čvorova u dubljim slojevima mreže. Stoga se doprinos grešci za čvor a izračunava tako što se čvoru a „proslijedi" dio greške iz svih čvorova koji na ulaz dobijaju vrijednost izračunatu u čvoru a. Na primjer, ukoliko se izlaz čvora a prosljeđuje na ulaz čvorovima $a'_1, a'_2, ..., a'_l$ sa aktivacijama $y_{a'_1}, y_{a'_2}, ..., y_{a'_l}$, doprinos grešci za čvor a dobija se sumiranjem djelova greške proslijeđenih čvoru a preko svakog od čvorova $a'_1, a'_2, ..., a'_l$. Prilikom računanja dijela greške koji će biti proslijeđen određenom konekcijom koristi se pravilo (68). Izraz za određivanje $\partial J/\partial y_a$ izgleda ovako:

$$\frac{\partial J}{\partial y_a} = \sum_{i=1}^{l} \frac{\partial J}{\partial y_{a'_i}}\frac{\partial y_{a'_i}}{\partial y_a}. \tag{69}$$

Ovaj postupak počinje određivanjem parcijalnih izvoda funkcije J po izlaznim čvorovima mreže $\hat{y}$, tj. $\partial J/\partial \hat{y}$. Ovi izvodi određuju se na jednostavan način, diferenciranjem izraza za funkciju gubitka. U narednim koracima se izračunati izvodi koriste u izračunavanju izvoda za čvorove koji im prethode, u skladu sa pravilom (69) i na taj način se informacije o grešci propagiraju sve do čvorova ulaznog sloja.

Na osnovu vrijednosti $\partial J/\partial y_a$ određuju se parcijalni izvodi po parametrima čvora a i vrši se njihovo ažuriranje pravilom (67). Ako je θ_a jedan od parametara čvora a, parcijalni izvod funkcije gubitka po tom parametru računa se primjenom pravila (68):

$$\frac{\partial J}{\partial \theta_a} = \frac{\partial J}{\partial y_a}\frac{\partial y_a}{\partial \theta_a}. \tag{70}$$

Uzastopnom primjenom pravila (67) smanjuje se vrijednost funkcije gubitka sve dok se ne ispuni neki od uslova zaustavljanja. Postoje različiti kriterijumi kada se gradijentni spust može zaustaviti. Osnovni je kada gradijent bude jednak nuli, odnosno kada se dođe u tačku lokalnog ekstremuma. Međutim, to je rijetko kada slučaj. Najčešće se unaprijed definiše broj iteracija gradijentnog spusta. Pored toga, gradijentni spust se može zaustaviti kada nekoliko uzastopnih iteracija ne donese značajna poboljšanja u vrijednosti funkcije gubitka ili kada gradijenti ili vrijednost funkcije gubitka padne ispod unaprijed zadate granice. U praksi se koristi i tehnika ranog zaustavljanja (*engl. early stopping*) kojom se uporedo prate performanse sistema na skupu za validaciju, a obučavanje se zaustavlja kada neuronska mreža nekoliko iteracija ne poboljšava rezultate tom skupu.

Gradijentni spust je ključni algoritam za obučavanje dubokih neuronskih mreža. Međutim, u osnovnom obliku rijetko se koristi u praksi zbog niza nedostataka koji su vremenom uočeni. U nastavku sekcije biće izloženi glavni nedostaci gradijentnog spusta, zajedno sa modifikacijama koje su predložene kako bi se oni prevazišli.

U osnovnoj verziji gradijentnog spusta, vrijednost funkcije gubitka se izračunava na osnovu čitavog skupa za obučavanje. Skupovi za obučavanje neuronskih mreža su izuzetno obimni, pa bi prolazak čitavim skupom zarad samo jedne promjene vrijednosti parametara mreže oduzimao mnogo vremena i usporio konvergenciju. Kao alternativa je predložen stohastički gradijentni spust u kojem se gradijent izračunava na osnovu uzorka primjeraka iz skupa za obučavanje fiksne veličine (*engl. minibatch*). Upotrebom stohastičkog gradijentnog spusta se osigurava da je vremenska složenost računanja gradijenta nezavisna od veličine skupa za obučavanje. U tom pogledu, ona se može smatrati i konstantnom. Veličina uzorka koji se koristi je veoma važan hiperparametar ovog algoritma [110]. Veći uzorak daje bolju procjenu gradijenta. Međutim, preveliki uzorci mogu prouzrokovati preveliko prilagođavanje modela podacima za obučavanje ili dovesti proceduru optimizacije do „lošeg" lokalnog minimuma. Manji uzorci unose šum u proces obučavanja i čine neuronsku mrežu manje podložnom preprilagođavanju.

Nerijetko funkcija gubitka neuronskih mreža u blizini lokalnog ima oblik jaruge (*engl. ravine*). Korišćenjem fiksne vrijednosti stope učenja može se desiti da korak gradijentnog spusta bude preveliki i da procedura optimizacije prelazi sa jednog na drugi kraj jaruge, ne spuštajući se ka lokalnom minimumu. Takođe, fiksna vrijednost stope učenja u drugim djelovima prostora parametara uzrokuje premale korake, naročito u početnoj fazi obučavanja modela i time bespotrebno usporava konvergenciju. Nekada gradijenti izračunati na osnovu malih uzoraka mogu imati veliku varijansu u blizini lokalnih minimuma i odvesti obučavanje u pogrešnom smjeru. Ovo se može prevazići povećanjem uzorka pri kraju obučavanja. Međutim, to je veoma neprak-

tično. Da bi se poboljšala konvergencija, često je djelotvorno koristiti stopu učenja koja se smanjuje tokom vremena. Nekim unapređenjima gradijentnog spusta pokušava se doći do najboljeg rasporeda za vrijednosti stope učenja, odnosno najbolje vrijednosti za veličinu koraka pomjeranja u parametarskom prostoru.

Kao efikasne su se pokazale metode koje ovaj problem prevazilaze uvođenjem momentuma u proceduru optimizacije [111]. Uvedena je promjenljiva η koja ujedno predstavlja smjer i momentum, odnosno brzinu kretanja u prostoru parametara. Ova ideja ima analogiju u fizici, odnosno kretanju čestica. Negativni gradijent predstavlja silu koja pomjera česticu u prostoru po Njutnovim zakonima kretanja. U fizici je momentum jednak proizvodu mase i brzine. U ovom slučaju se uzima jedinična masa, pa se momentum i brzina mogu posmatrati ekvivalentno. U promjenljivoj η^t akumuliraju se gradijenti iz prethodnih iteracija algoritma po sljedećem pravilu:

$$\eta^t \leftarrow \mu\eta^{t-1} - \nabla_\Theta J(\Theta^{t-1}). \tag{71}$$

Zatim se vektor parametara mreže ažurira sa:

$$\Theta^t \leftarrow \Theta^{t-1} + \nu\eta^t. \tag{72}$$

Kod gradijentnog spusta veličina koraka zavisila je samo od vrijednosti stope učenja ν i norme gradijenta. Kod momentum tehnika zavisi dodatno i od poklapanja gradijenata. Veličina koraka se povećava ako je nekoliko uzastopnih gradijenata istog smjera, dok ukoliko imamo promjenu znaka u gradijentu, momentum se smanjuje. Parametar μ uveden je kao koeficijent trenja, kako bi se brzina smanjivala tokom vremena i da bi algoritam konvergirao lokalnom minimumu, što se inače možda ne bi desilo.

Nesterovljev momentum, zasnovan na [112], je varijanta standardnog momentum algoritma koja je u radu [113] predložena za obučavanje modela dubokog učenja. Osnovna razlika između Nesterovljevog momentuma i standardne momentum metode je u tački u kojoj se izračunava gradijent funkcije gubitka. Nesterovljevim momentumom se gradijent funkcije računa nakon što se na postojeću vrijednost parametara primijeni trenutna brzina (momentum) η^t. Time se pravilo (71) transformiše u:

$$\eta^t \leftarrow \mu\eta^{t-1} - \nabla_\Theta J(\Theta^{t-1} + \nu\mu\eta^{t-1}). \tag{73}$$

Momentum tehnike prevazilaze neke od problema gradijentnog spusta, ali čine to uvođenjem dodatnog hiperparametra μ. Postoji čitava grupa tehnika koja se zasniva na prilagođavanju stope učenja ν za svaki od parametara neuronske mreže zasebno.

AdaGrad tehnika [114] vrši prilagođavanje stope učenja na sljedeći način:

$$\begin{aligned} \varphi^t &\leftarrow \varphi^{t-1} + \nabla_\Theta J(\Theta^{t-1}) \odot \nabla_\Theta J(\Theta^{t-1}) \\ \Theta^t &\leftarrow \Theta^{t-1} - \frac{\nu}{\epsilon + \sqrt{\varphi^t}} \nabla_\Theta J(\Theta^{t-1}), \end{aligned} \tag{74}$$

Operacija $\odot$ predstavlja Hadamardov proizvod vektora, a φ^t je vektor čiji su elementi sume kvadrata svih parcijalnih izvoda po parametrima modela od početka obučavanja do iteracije t. Parametar ε je mala konstanta koja se dodaje zbog numeričke stabilnosti. Ovim načinom ažuriranja stope učenja se postiže efekat da se parametrima koji su kroz istoriju obučavanja imali velike gradijente smanjuje stopa učenja, dok se veća stopa učenja, odnosno veći koraci, primjenjuje u dijelu parametarskog prostora sa blažim nagibom. AdaGrad tehnika ima neka pogodna svojstva, ali čuvanje cjelokupne istorije gradijenata može prouzrokovati preuranjeno smanjivanje stope učenja za neke parametre, dok se nekim parametrima zbog prvobitno malih gradijenata može neopravdano povećavati stopa učenja. AgaGrad ostvaruje brzu konvergenciju u minimizaciji konveksnih funkcija. Međutim, oblici funkcija gubitka neuronskih mreža su veoma nepravilni. Nagibi po različitim smjerovima se intenzivno mijenjaju, pa se stoga odlučivanje o stopi učenja na osnovu nagiba koji su računati znatno ranije u procesu obučavanja nije pokazalo pogodnim za ovu vrstu modela dubokog učenja.

RMSProp tehnika [115] pokušava prevazići navedeni nedostatak AdaGrad tehnike eksponencijalnim smanjivanjem uticaja ranijih gradijenata na obučavanje. Na ovaj način se gradijenti iz daleke istorije suštinski odbacuju i veći uticaj na korak gradijentnog spusta se daje lokalnoj strukturi funkcije gubitka. RMSProp tehnika ažurira parametre neuronske mreže sljedećim pravilima:

$$\begin{aligned} \varphi^t &\leftarrow \rho\varphi^{t-1} + (1-\rho)\nabla_\Theta J(\Theta^{t-1}) \odot \nabla_\Theta J(\Theta^{t-1}) \\ \Theta^t &\leftarrow \Theta^{t-1} - \frac{\nu}{\sqrt{\epsilon + \varphi^t}} \nabla_\Theta J(\Theta^{t-1}). \end{aligned} \tag{75}$$

Uveden je hiperparametar ρ, koji obično uzima vrijednosti iz skupa $\{0.9, 0.99, 0.999\}$. Njegovom vrijednošću se kontroliše uticaj ranijih vrijednosti gradijenata na veličinu koraka u parametarskom prostoru.

RMSProp se može uvezati i sa idejom Nesterovljeve momentum tehnike, čime se parametri ažuriraju sa:

$$\begin{aligned} \varphi^t &\leftarrow \rho\varphi^{t-1} + (1-\rho)\nabla_\Theta J(\Theta^{t-1} + \nu\mu\eta^{t-1}) \odot \nabla_\Theta J(\Theta^{t-1} + \nu\mu\eta^{t-1}) \\ \eta^t &\leftarrow \mu\eta^{t-1} - \frac{1}{\sqrt{\varphi^t}} \nabla_\Theta J(\Theta + \nu\mu\eta^{t-1}) \\ \Theta^t &\leftarrow \Theta^{t-1} + \nu\eta^t. \end{aligned} \tag{76}$$

Adam [116] je optimizacioni algoritam koji se može posmatrati kao RMSProp sa momentom, uz nekoliko propratnih detalja. Naziv ovog algoritma izveden je iz fraze „adaptivni momenti” (*engl. adaptive moments*). Ovim algoritmom ustanovljen

je sljedeći niz pravila za ažuriranje parametara modela dubokog učenja:

$$\begin{aligned}
\eta^t &\leftarrow \mu\eta^{t-1} + (1-\mu)\nabla_\Theta J(\Theta^{t-1}) \\
\hat{\eta}^t &\leftarrow \frac{\eta^t}{1-\mu^t} \\
\varphi^t &\leftarrow \rho\varphi^{t-1} + (1-\rho)\nabla_\Theta J(\Theta^{t-1}) \odot \nabla_\Theta J(\Theta^{t-1}) \\
\hat{\varphi}^t &\leftarrow \frac{\varphi^t}{1-\rho^t} \\
\Theta^t &\leftarrow \Theta^{t-1} - \nu\frac{\hat{\eta}^t}{\epsilon + \sqrt{\hat{\varphi}^t}}.
\end{aligned} \tag{77}$$

Momentum je u okviru Adam algoritma uključen unutar vektora η^t. Ovo je najjednostavniji način za dodavanje momentuma RMSProp tehnici i nema jaku teorijsku potporu. Hiperparametri μ i ρ obično uzimaju vrijednosti 0.9 i 0.999, respektivno, iako se ovaj algoritam smatra prilično robustnim na izbor ovih vrijednosti. RMSProp tehnika manifestuje nestabilnosti na početku obučavanja, zbog inicijalizacije vektora η_0 i φ_0 na nulu. Adam algoritam izbjegava ovaj problem uvođenjem korekcionog faktora, odnosno vektora $\hat{\eta}^t$ i $\hat{\varphi}^t$ koji zavise od trenutnog broja iteracije t.

Konačno, algoritam Nadam [117] uključuje Nesterovljev momentum u Adam algoritam. Efikasnost momentum tehnike se dodatno poboljšava uvođenjem rasporeda vrijednosti za parametar μ u zavisnosti od iteracije t. To rezultuje sljedećim skupom pravila:

$$\begin{aligned}
\eta^t &\leftarrow \mu^t\eta^{t-1} + (1-\mu^t)\nabla_\Theta J(\Theta^{t-1}) \\
\hat{\eta}^t &\leftarrow \frac{\eta^t}{1-\prod_{i=1}^{t+1}\mu^t} \\
\overline{\eta}^t &\leftarrow (1-\mu^t)\frac{\nabla_\Theta J(\Theta^{t-1})}{1-\prod_{i=1}^{t}\mu_i} + \mu^{t+1}\hat{\eta}^t \\
\varphi^t &\leftarrow \rho\varphi^{t-1} + (1-\rho)\nabla_\Theta J(\Theta^{t-1}) \odot \nabla_\Theta J(\Theta^{t-1}) \\
\hat{\varphi}^t &\leftarrow \frac{\varphi^t}{1-\prod_{i=1}^{t+1}\rho^t} \\
\Theta^t &\leftarrow \Theta^{t-1} - \nu\frac{\overline{\eta}^t}{\epsilon + \sqrt{\hat{\varphi}^t}}.
\end{aligned} \tag{78}$$

Način izbora tehnike za optimizaciju nije jasno određen. Tehnike sa adaptivnom stopom učenja smatraju se superiornijima i lakšim za korišćenje zbog manje osjetljivosti na vrijednosti hiperparametara. Međutim, nijedna od tehnika se nije izdvojila kao najbolja i sve se gotovo ekvivalentno koriste.

5.4.3 Transfer učenja

Transfer učenja je koncept u psihologiji koji predstavlja prenošenje znanja sa jedne vještine na novu, do tada nesavladanu, vještinu. Postojanje ovog efekta kod ljudi je očigledno. Na primjer, osoba koja zna da svira neki muzički instrument obično će lakše naučiti da svira novi, nego osoba koja nema iskustva u muzici. Znanja i vještine se mogu prenositi i sa jedne naučne oblasti na drugu, sa jednog sporta na drugi, itd.

Koncept transfera učenja postoji i intenzivno se koristi u dubokom učenju. Ova tehnika se sprovodi na veoma jednostavan način, pošto mehanizmi transfera učenja u ljudskom mozgu još uvijek nisu dovoljno razjašnjeni da bi mogli biti oponašani vještačkim neuronskim mrežama. Transfer učenja između dvije vještačke neuronske mreže sastoji se u postavljanju vrijednosti parametara modela obučenog na jednom zadatku kao inicijalnih vrijednosti za model koji će dalje biti obučavan na drugom zadatku. Očekuje se da će ovako postavljen model biti bolje ili barem brže obučen za postavljeni zadatak od modela koji bi startovao obučavanje bez transfera učenja. Odabir zadatka sa kojeg se vrši transfer učenja nije trivijalan. Dostupnost velikog broja modela, obučenih na različitim skupovima, predstavlja značajan resurs u ovom pogledu. Međutim, potrebna je odgovarajuća ekspertiza kako bi se zaključilo da li se znanja stečena na tom zadatku mogu iskoristiti u stvaranju boljeg modela za postojeći zadatak.

5.4.4 Inicijalizacija parametara

Kada ne postoji odgovarajući model s kojeg bi se prenijelo znanje, algoritmima kojima se obučavaju neuronske mreže, zasnovanim na metodi gradijentnog spusta, potrebno je zadati tačku iz koje počinje optimizacija funkcije gubitka. Od izbora ove polazne tačke zavisi i konačan ishod obučavanja mreže. Inicijalizacija parametara neuronske mreže je procedura kojom se težinskim koeficijentima i slobodnim članovima svih neurona u mreži dodjeljuju početne vrijednosti, prije nego otpočne proces obučavanja.

Međutim, loše odabrane inicijalne vrijednosti parametara mogu prouzrokovati nestajanje ili eksploziju gradijenata, ili značajno otežati, nekada čak i onemogućiti, pronalazak optimalnih vrijednosti za te parametre. U situacijama kada procedura obučavanja pronalazi tačku lokalnog minimuma, odnosno konvergira, od izbora polazne tačke zavisi da li će ta tačka lokalnog minimuma imati dovoljno malu vrijednost. Inicijalizacija takođe može uticati i na grešku u generalizaciji, odnosno grešku koju model pravi na skupu za testiranje. Nekada dvije postavke parametara neuronske

mreže sa istom vrijednošću funkcije greške, mogu proizvesti značajno različite greške na skupu za testiranje. Nakon što je uočen značajan uticaj inicijalizacije na ishod i efikasnost treninga, i njoj se počela posvećivati sve veća pažnja.

Inicijalizovanje svih parametara mreže konstantnom vrijednošću nije prikladno, jer je gradijentni spust deterministički algoritam. Svi neuroni jednog sloja neuronske mreže obično na ulazu dobijaju identične vrijednosti i nad njima primjenjuju identičnu aktivacionu funkciju. Kada bi i parametri tih neurona bili jednaki, izvodi funkcije gubitka po tim parametrima bi, u tom slučaju, takođe bili identični. Neuroni tog sloja bi se kroz iteracije algoritma obučavanja mijenjali na isti način, odnosno učili bi iste karakteristike ulaznog signala, što nije poželjno. Ovo nema efekta na *bias* parametre, pa se oni mogu inicijalizovati konstantom, najčešće nulom.

Težinskim koeficijentima je najprikladnije na početku dodijeliti slučajno izgenerisane vrijednosti. Ovim načinom zadavanja inicijalnih vrijednosti parametara smanjuje se šansa da dva neurona u mreži računaju istu vrijednost. Prve ideje, koje su se dugo zadržale, bile su da se inicijalne vrijednosti za parametre uzorkuju iz uniformne raspodjele $\mathcal{U}(-\varepsilon, \varepsilon)$ ili normalne raspodjele $\mathcal{N}(0, \sigma^2)$. Ove raspodjele podešene su tako da im je matematičko očekivanje 0, a disperzija dovoljno mala kako bi se izbjegla eksplozija gradijenata i divergencija prilikom obučavanja mreže. Uzorkovanje vrijednosti parametara iz raspodjele sa velikom disperzijom pomaže u smanjenju broja suvišnih neurona, ali može dovesti do akumuliranja velikih vrijednosti tokom propagacije vrijednosti unaprijed ili unazad kroz mrežu. Velike vrijednosti na ulazu mogu izazvati eksploziju vrijednosti aktivacionih funkcija poput ReLU, hiperboličkog tangensa, *Swish*, i njihovih gradijenata, ili saturaciju funkcija kao što je sigmoid, gdje dolazi do gubitka gradijenata za te neurone. Dodjeljivanjem premalih vrijednosti parametrima može se prouzrokovati suprotan efekat, odnosno nestajanje gradijenata i usporavanje treninga. S povećanjem broja slojeva, ovaj problem se pogoršava. Do izbora optimalne vrijednosti za disperziju raspodjele koja se koristi za inicijalizaciju parametara mreže dolazi se uravnotežavanjem ovih efekata, što je veoma izazovan zadatak.

Vremenom je uočeno da se specifičnijim heurističkim tehnikama inicijalizacije mogu značajno umanjiti šanse da dođe do nestajanja ili eksplozije gradijenata, ali i ubrzati obučavanje neuronskih mreža. Ove tehnike, koje su postale de fakto standard u dubokom učenju, služe se informacijama kao što su vrsta aktivacione funkcije i ulazni i izlazni stepen neurona prilikom dodjeljivanja inicijalnih vrijednosti težinskim koeficijentima. Prema ovim tehnikama, u cilju izbjegavanja efekata nestajanja i eksplozije gradijenata treba se pridržavati sljedećih pravila. Srednja vrijednost aktivacija svih slojeva treba da bude jednaka nuli. Takođe, disperzija aktivacija treba da bude ujednačena u svim slojevima. Ukoliko su ova dva uslova zadovoljena, vrijednost

gradijenta koja se propagira unazad kroz slojeve neće se množiti sa prevelikim ili premalim vrijednostima i stići će do ulaznog sloja bez eksplodiranja ili nestajanja.

Glorotova inicijalizacija je jedna od modernijih šema za inicijalizaciju. Nazvana je po svom autoru, koji je ovaj način inicijalizacije parametara neuronske mreže predstavio u radu [118]. Prema ovoj šemi inicijalizacije, težinskim koeficijentima se dodjeljuju vrijednosti uzorkovanjem iz uniformne raspodjele:

$$\mathcal{U}\left(-\sqrt{\frac{6}{N_{in}+N_{out}}}, \sqrt{\frac{6}{N_{in}+N_{out}}}\right), \tag{79}$$

gdje je N_{in} ulazni stepen neurona u sloju, a N_{out} je ukupan broj izlaza iz sloja. Umjesto uniformne raspodjele, gotovo ekvivalentno se može uzeti i normalna (Gausova) raspodjela $\mathcal{N}(0, \frac{6}{N_{in}+N_{out}})$. Posljedice izbora jedne ili druge raspodjele nisu detaljnije proučavane. Upotreba ulaznih i izlaznih stepena neurona u izvođenju raspodjele za uzorkovanje inicijalnih vrijednosti parametara pomaže u izjednačavanju disperzija aktivacija i disperzija gradijenata svih slojeva mreže. Raspodjela koja se koristi u Glorotovoj inicijalizaciji izvedena je s pretpostavkom da se neuronska mreža sastoji isključivo od niza matričnih množenja, bez nelinearnosti. Ova pretpostavka očigledno ne važi u gotovo svim neuronskim mrežama. Međutim, empirijski je pokazala dobre osobine kada se koristi sa odgovarajućom vrstom slojeva, što je razlog iz kojeg se i dalje intenzivno primjenjuje.

Glorotova inicijalizacija se trenutno koristi za inicijalizaciju parametara slojeva koji kao aktivacionu funkciju koriste sigmoid ili hiperbolički tangens, koje su simetrične u odnosu na koordinatni početak. Nije se pokazala dobro u praksi kada je korišćena u slojevima sa ReLU aktivacionom funkcijom i drugim funkcijama koje ne ispunjavaju ovaj uslov simetričnosti. U istom radu u kojem je uvedena aktivaciona funkcija PReLU, predložena je i nova strategija za inicijalizaciju parametara [107]. Takođe je po svom autoru nazvana Hiova inicijalizacija. Ovaj način inicijalizacije trenutno se najčešće koristi u slojevima sa ReLU aktivacionom funkcijom i njenim varijantama (propustljivi ReLU i PReLU). Ukoliko se u slojevima neuronske mreže koristi ReLU aktivaciona funkcija, izvodi se sljedeća raspodjela za uzorkovanje incijalnih vrijednosti parametara: $\mathcal{N}(0, \frac{2}{N_{in}})$. Kada se u računanje vrijednosti ReLU aktivacione funkcije uključi i parametar v (čime se dobijaju propustljiva ReLU i PReLU), raspodjela za inicijalizaciju dobija sljedeći oblik: $\mathcal{N}(0, \frac{2}{N_{in}(1+v^2)})$.

Obje ove strategije kreirane su sa namjerom da se neuronska mreža dovede u stanje koje je prikladno za obučavanje. Međutim, malo je poznato da li se, nakon što obučavanje počne, ovo stanje održava. Jedino što se može tvrditi je da će inicijalne vrijednosti parametara biti raznolike. Drugi uticaji ovih strategija na proces obučavanja moraju biti podrobnije analizirani.

5.5 Metode unapređenja obuke neuronskih mreža

Obučavanje neuronskih mreža je složen proces praćen brojnim izazovima, poput prevelikog prilagođavanja podacima za obučavanje, eksplozije ili nestajanja aktivacija i/ili gradijenata. Zbog toga su razvijane metode usmjerene na suzbijanje ovih problema. Metode koje su korištene u ovom radu izložene su u nastavku ove sekcije. Primjena ovih metoda često dovodi do poboljšanja performansi i stabilnosti tokom procedure obučavanja, ali nije garancija njene uspješnosti. Uspješnost obučavanja neuronskih mreža zavisi od mnogih drugih faktora: kvaliteta podataka, odabira odgovarajuće arhitekture modela, pravilnog podešavanja vrijednosti hiperparametara, kao i pravovremenog i adekvatnog prilagođavanja procedure obučavanja.

5.5.1 Skaliranje ulaznih atributa

Skaliranje ulaznih atributa neuronskih mreža je veoma važan korak koji prethodi njihovom obučavanju. Sve ulazne vrijednosti poželjno je svesti na sličan opseg, sa relativno malim vrijednostima, kako bi se podaci učinili pogodnijim za modelovanje.

Ukoliko ulazni atributi nisu u istoj skali, gradijenti po svakoj od ulaznih promjenljivih će se drastično razlikovati, što može dovesti do toga da su izlazi modela dominanto bazirani na vrijednosti atributa sa većim opsegom. Skaliranje je poželjno i ukoliko su sve ulazne vrijednosti neuronske mreže u istom opsegu. Prevelike vrijednosti ulaza mogu dovesti do eksplozije neurona sa aktivacionom funkcijom ReLU, kao i odgovarajućih gradijenata. Ovo za posljedicu ima divergenciju ili oscilovanje optimizacione procedure zbog prevelikih koraka gradijentnog spusta. Nasuprot tome, u neuronima sa sigmoid aktivacijom doći će do saturacije gradijenata, što će izazvati male korake u gradijentnom spustu i usporavanje obučavanja modela.

Skaliranjem ulaznih podataka se numerički stabilizuje, a samim tim i ubrzava, obučavanje neuronske mreže. Pomaže se da ova procedura postane manje osetljiva na inicijalizaciju parametara mreže. Takođe se sprečava da atributi sa većim vrijednostima dominiraju prilikom obučavanja modela.

Postoji više metoda skaliranja podataka. Najčešće korišćene su min-max skaliranje i standardizacija. Metodom min-max skaliranja se sve ulazne vrijednosti svode na opseg $[0, 1]$, sljedećom transformacijom:

$$\hat{x} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}, \tag{80}$$

gdje su x_{min} i x_{max} najmanja i najveća vrijednost atributa x u skupu za obučavanje.

Standardizacija je drugi metod skaliranja ulaznih atributa neuronske mreže kojim se sve vrijednosti svode na standardnu normalnu raspodjelu $\mathcal{N}(0,1)$. Standardizacijom se ulazni atributi transformišu na sljedeći način:

$$\hat{x} = \frac{x - \mu_x}{\sqrt{\sigma_x^2}}. \tag{81}$$

Vrijednost μ_x predstavlja srednju vrijednost atributa x, a vrijednost σ_x je njegova standardna devijacija. Ove vrijednosti izračunavaju se na jednostavan način, na osnovu svih primjeraka iz skupa za obučavanje.

Izbor metode za skaliranje zavisi od prirode podataka. Postojanje izuzetaka u skupu (*engl. outliers*), sa izuzetno velikim ili izuzetno malim vrijednostima, može poremetiti min-max skaliranje. Standardizacijom se mijenja originalna raspodjela podataka. Stoga je min-max skaliranje bolje koristiti kada je poznato da ulazni podaci ne prate normalnu raspodjelu. Standardizacija je znatno manje osjetljiva na izuzetke i preferira se kada ulazni atributi imaju normalnu raspodjelu. Međutim, u praksi se uglavnom prednosti jedne ili druge metode obično utvrđuju eksperimentalnim putem na konkretnom problemu.

Deskaliranje, odnosno destandardizacija podataka su operacije suprotna skaliranju i standardizaciji. Ovim operacijama se podaci vraćaju u prvobitni opseg, prije skaliranja i standardizacije, što je nekada potrebno kako bi se interpretirale izlazne vrijednosti modela.

5.5.2 Normalizacija po seriji

U Sekciji 5.4 je kao jedan od glavnih problema modela mašinskog učenja istaknuto preveliko prilagođavanje podacima za obučavanje, odnosno neadekvatna sposobnost generalizovanja. Ovaj problem se prepoznaje po prevelikoj razlici između grešaka modela na skupu za obučavanje i grešaka na testnom skupu. U cilju njegovog sprečavanja, osmišljene su različite metode regularizacije modela. Tehnika ranog zaustavljanja spomenuta je u Sekciji 5.4. Model mašinskog učenja može se regularizovati i dodavanjem odgovarajuće norme vektora parametara u funkciju gubitka, čime se suzbijaju vrijednosti koje parametri mogu uzeti i sprečava pretjerano prilagođavanje. Međutim, ova metoda nije široko zastupljena u neuronskim mrežama. Nasumično izostavljanje dijela neurona u slojevima neuronske mreže (*engl. dropout*) [119], odnosno izjednačavanje njihovih aktivacija sa nulom u toku obučavanja, sprečava koadaptaciju neurona i međusobno ispravljanje grešaka, odnosno podstiče neuronsku mrežu na generalizaciju. Čak se i tehnike inicijalizacije parametara, opisane u Sekciji 5.4.4, mogu svrstati u grupu tehnika za regularizaciju. Heurističkim tehnikama inicijalizacije parametara neuronskih mreža, opisanim u prethodnoj sekciji, pokušavaju

se ograničiti aktivacije svih neurona u mreži kako bi se izbjegli efekti eksplozije ili nestajanja gradijenata, ali i zadala dobra polazna tačke za proceduru obučavanja. Međutim, kada obučavanje počne, uticaj ovih tehnika na izlaze neurona gotovo da u potpunosti isčezava. Time se i povećava šansa da procedura obučavanja završi u lokalnom ekstremumu funkcije gubitka koji neće rezultovati dobrim performansama na testnom skupu.

U nastavku sekcije će više pažnje biti posvećeno metodi normalizacije po seriji (*engl. batch normalization*). Ova metoda predložena je u radu [120] i zbog svojih praktičnih doprinosa postala je nezaobilazna gradivna cjelina gotovo svih savremenih neuronskih mreža, pa tako i onih kreiranih tokom ovog istraživanja.

Normalizacija po seriji, vrši skaliranje ulaza slojeva neuronske mreže, slično standardizaciji ulaznih atributa iz sekcije 5.5.1. Međutim, ona se može primijeniti na ulazu bilo kojeg sloja mreže. Normalizacijom po seriji se takođe uvode dodatni parametri, koji se podešavaju tokom obučavanja, kako bi se normalizacija prilagodila specifičnostima svakog sloja. Ako posmatramo jednu seriju (*engl. batch*) ulaza u sloj neuronske mreže $x^{(1)}, x^{(2)}, ..., x^{(m)}$, normalizacijom po seriji se oni trasformišu sljedećim pravilom:

$$\hat{x}^{(i)} = \gamma \frac{x^{(i)} - \mu_x}{\sqrt{\varepsilon + \sigma_x^2}} + \beta, \tag{82}$$

gdje je μ_x srednja vrijednost, a σ_x estimacija standardne devijacije na seriji ulaza $\left\{x^{(i)}, i \in \{1, 2, ..., m\}\right\}$. Hiperparametar ε je uveden kako bi se izbjeglo dijeljenje sa nulom, a γ i β su parametri čija se vrijednost podešava tokom obučavanja. Normalizacija po seriji se uglavnom primjenjuje neposredno prije aktivacione funkcije. Međutim, ova praksa nije dovoljno argumentovana, pa se u nekim mrežnim arhitekturama može naći i nakon aktivacije.

Normalizacijom se aktivacije slojeva neuronske mreže svode na standardnu normalnu raspodjelu, sa srednjom vrijednošću 0 i disperzijom 1. Uvođenjem parametara γ i β ostavlja se prostor da aktivacije slojeva imaju u izvjesnoj mjeri različite raspodjele, ali se i dalje drže pod kontrolom. Parametrom γ se kontroliše disperzija aktivacija sloja, a parametrom β pomjeraj srednje vrijednosti.

Kada se obučavanje modela okonča, slojevi normalizacije po seriji koriste se tako što se parametri γ i β fiksiraju na vrijednosti koje su uzeli na kraju obučavanja, a za μ_x i σ_x se koriste globalna srednja vrijednost, odnosno standardna devijacija, izračunata na osnovu svih serija podataka u toku obučavanja.

Normalizacija po seriji ima niz pozitivnih efekata na proceduru obučavanja neuronske mreže. Algoritmom propagacije unazad vrše se izračunavanja parcijalnih

izvoda po parametrima neuronske mreže od poslednjeg ka prvom sloju. Parcijalni izvod daje smjer i intenzitet promjene svakog od parametara, ali pod pretpostavkom da ostali parametri u prethodnim slojevima ostanu nepromijenjeni. Međutim, algoritam obučavanja istovremeno ažurira sve parametre, pa ova pretpostavka ne stoji. Nakon što se izvrši promjena parametara u ranijim slojevima, ažuriranje kasnijih slojeva može se ispostaviti kao suboptimalno. Ovaj problem za sada nije u potpunosti riješen, ali kako bi se umanjio neophodno je da procedure obučavanja koriste male vrijednosti stope učenja, kao i da parametri mreže budu dobro inicijalizovani. Normalizacijom po seriji se aktivacije slojeva u dobroj mjeri ograničavaju. Tada se prilikom ažuriranja parametara u jednom sloju može pretpostaviti raspodjela aktivacija prethodnih slojeva, pa se promjene vrijednosti mogu ležernije izvršavati. Eksperimenti sprovedeni u radu [120], ali i mnogi drugi praktični primjeri pokazuju da primjena normalizacije po seriji vodi ka bržoj i stabilnijoj konvergenciji procedure obučavanja neuronske mreže. Uslijed stabilnije konvergencije, smanjuje se potreba za pažljivom inicijalizacijom parametara, normalizacijom ulaznih podataka i podešavanjem hiperparametara mreže, poput stope učenja. Normalizacija po seriji sprečava i da parametri i izlazne vrijednosti slojeva postanu premale i time dovedu do gubitka informacija prilikom prenosa kroz mrežu. Regularizacija modela je takođe jedna od posljedica primjene ove metode. Razlike u distribuciji podataka od uzorka do uzorka uvode šum koji model odvraća od pretjeranog prilagođavanja. Povećavanjem veličine uzorka za obučavanje umanjuje se efekat ove regularizacije.

Razlozi uspješnosti normalizacije po seriji su još uvijek predmet rasprave u istraživačkoj zajednici. Postoji nekoliko hipoteza kojima se pokušava obrazložiti uticaj ove metode na obučavanje neuronskih mreža. U izvornom radu [120], autori pretpostavljaju da je uzrok efikasnosti njihove metode to što se njome suzbija unutrašnje kovarijatno pomjeranje (*engl. internal covariate shift*). Unutrašnje kovarijantno pomjeranje je efekat koji se javlja kada se mijenjaju raspodjele ulaza skrivenih (unutrašnjih) slojeva duboke neuronske mreže. Kako se koraci obučavanja neuronske mreže sprovode na relativno malim uzorcima, ova pomjeranja distribucija mogu biti izuzetno dramatična, jer se ti uzorci mogu međusobno drastično razlikovati. Normalizacijom po seriji se umanjuje ovaj efekat. Međutim, u radu [121] je ova hipoteza opovrgnuta nizom eksperimenata u kojima je pokazano da ne postoji veza između unutrašnjeg kovarijantnog pomjeranja i ishoda procedure obučavanja. Autori [121] su dali novu pretpostavku da se normalizacijom po seriji vrši reparametrizacija modela koja optimizacioni problem preslikava u prostor sa mnogo pogodnijim osobinama.

5.5.3 Preskačuće veze

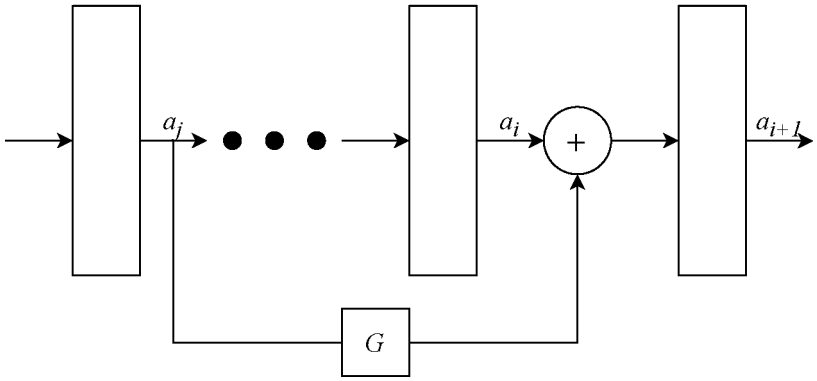
Naizgled, dodavanje većeg broja slojeva u neuronsku mrežu vodi boljim rezultatima, jer dublji slojevi mogu postepeno učiti sve kompleksnije karakteristike ulaza i na taj način poboljšavati performanse. Međutim, duboke neuronske mreže susreću se s problemom isčezavanja gradijenata. Ovo je problem koji se javlja kada parcijalni izvodi po parametrima nekih slojeva u mreži postanu toliko mali da u potpunosti zaustave dalje obučavanje neuronske mreže. Kako su parametri mreže i stopa obučavanja po apsolutnoj vrijednosti mali realni brojevi iz segmenta $[0, 1]$, slijedi da će se algoritmom propagacije unazad parcijalni izvodi postepeno smanjivati uzastopnom primjenom pravila (68). Ukoliko mreža ima preveliki broj slojeva dolazi do prevelikog smanjivanja gradijenata za parametre u početnim slojevima koji se ne mogu adekvatno ažurirati kako bi modelovali potrebne karakteristike ulaznih podataka.

Jedan sloj tipične duboke neuronske mreže kao ulaz dobija isključivo izlaz njemu prethodnog sloja. Funkcija koju neuroni tog sloja računaju može se predstaviti sljedećom jednakošću:

$$y_i = F\left(y_{i-1}\right), \tag{83}$$

gdje je y_i vektor izlaznih vrijednosti i-tog, a y_{i-1} vektor izlaznih vrijednosti $i-1$-og sloja neuronske mreže. Isčezavanje gradijenata za neurone i-tog sloja opstruiralo bi obučavanje kompletne neuronske mreže. Parametri sloja $i-1$ i njemu prethodnih slojeva se ne bi ažurirali zbog premalih gradijenata, čime bi taj dio mreže bio odsječen od ostatka i ne bi mogao doprinijeti u modelovanju tražene funkcije. Ovaj fenomen bi se mogao odraziti i na obučavanje neurona u i-tom i dubljim slojevima mreže. Parametri tih slojeva bi se nakon određenog vremena mogli zamrznuti jer ne bi bilo varijacija u njihovom ulazu. Stoga, svi slojevi u neuronskoj mreži moraju biti u stanju da barem proslijede znanje akumulirano u prethodnim slojevima, ako mu već ne mogu doprinijeti. Nikako ne smiju postati prepreka protoku informacija kroz mrežu.

Jedna od najuspješnijih tehnika za ublaživanje problema nestajućih gradijenata je dodavanje preskačućih veza (*engl. skip connections*) u mrežnu arhitekturu. Ovaj koncept uveden je u okviru ResNet modela u radu [122]. Kao što im i samo ime govori, ovim vezama se izlazi nekih slojeva sprovode na ulaz kasnijih slojeva u mreži, preskačući pritom jedan ili više slojeva, kao što je prikazano na Slici 13. Na ovaj način obezbjeđuju se alternativne putanje kojima gradijent može teći unazad od kasnijih slojeva ka početnim. Ove putanje sadrže manji broj grana, pa je samim tim i nestajanje gradijenata manje izraženo.



Slika 13: Prikaz preskačuće veze u arhitekturi neuronske mreže.

Ukoliko se u arhitekturu neuronske mreže doda preskačuća veza od j-tog do i-tog sloja ($j < i - 1$), tada neuroni i-tog sloja izračunavaju sljedeću funkciju:

$$y_i = F\left(y_{i-1} + G\left(y_j\right)\right), \tag{84}$$

gdje je G funkcija koja transformiše izlaz j-tog sloja y_j, prije sabiranja sa izlazom sloja $i - 1$. Najčešće je to funkcija identiteta ($G(x) = x$). Međutim, ukoliko se dimenzije izlaza slojeva $i - 1$ i j ne poklapaju, tada se funkcija G može iskoristiti za dopunjavanje ulaza nulama ili može predstavljati jedan konvolucioni sloj koji daje izlaz odgovarajuće dimenzije. Slika 13 i jednakost 84 predviđaju sabiranje dva ulaza i-tog sloja. Međutim moguće je korišćenje i drugih operacija, poput nadovezivanja, što će se vidjeti u nastavku disertacije.

Postojanjem preskačućih veza se osigurava da slojevi neuronske mreže u najgorem slučaju prenose informacije iz ranijih slojeva. Tipične neuronske mreže se za to moraju obučiti. Taj zadatak nije trivijalan i može se završiti neuspjehom ako procedura obučavanja nije dobro postavljena. Neuronske mreže sa preskačućim vezama su podrazumijevano u stanju da proslјeđuju informacije. Na taj način će duboke mreže imati barem jednako dobre performanse kao i plitke, a ne lošije. Uvođenje koncepta preskačućih veza omogućilo je uspješno obučavanje izuzetno kompleksnih neuronskih mreža sa više desetina slojeva.

6 Sistem vodenog žiga sa neuronskim mrežama

U ovoj disertaciji predlažemo novu paradigmu za kreiranje sistema vodenog žiga. Predloženom paradigmom predviđeno je korišćenje dubokih neuronskih mreža za obavljanje ključnih zadataka jednog sistema za umetanje vodenih žigova u digitalne audio signale. Razlozi za to su višestruki.

U prethodnim poglavljima je u više navrata isticana složenost dizajniranja procedure za umetanje vodenih žigova zbog kompleksnih zadataka koji se tom prilikom moraju riješiti. Ogromni napori su uloženi u istraživanje domena u kojima je najbolje sakriti bitove vodenog žiga i po kojim pravilima. Obično se ove operacije sprovode u domenu jedne od poznatih unitarnih ortogonalnih transformacija ili modifikacijom karakteristične vrijednosti izračunate na osnovu koeficijenata odabrane transformacije. Do metoda za umetanje se dolazi pažljivim ispitivanjem osobina audio signala i ljudskog slušnog sistema. Ipak, većina tehnika se na kraju može predstaviti kroz nekoliko pravila ili nizom od nekoliko uslova, kao što je izloženo u Sekciji 4.2. Kompleksnije zavisnosti i pravila je veoma teško izvesti tradicionalnim matematičkim aparatom. Postojeće tehnike ne ispunjavaju u potpunosti zahtjeve savremenih sistema vodenog žiga koji se suočavaju sa sve raznovrsnijim i kompleksnijim efektima i napadima, pa se čini se da je najbolji pristup prihvatiti složenost problema i riješiti ga tako što će se iskoristiti moć podataka i mogućnosti modela dubokog učenja [123].

Provlačenjem signala kroz niz slojeva neuronske mreže signal se projektuje u sintetički domen u kojem se vrši umetanje žigova. Ovaj domen može biti izrazito kompleksna transformacija izvornog domena signala koja bi se teško proizvela tradicionalnim matematičkim aparatom. Međutim, uz odgovarajuće procedure obučavanja i funkcije gubitka, može se doći do domena koji su prikladniji za skrivanje informacija od tradicionalnih. Daljom primjenom operatora u neuronskoj mreži mogu se dobiti sofisticiranija pravila za umetanje, a time i superiornije votermarking tehnike. Tradicionalni domeni i karakteristike signala izračunate u tim domenima mogu se zadržati i u ovim pristupima i koristiti kao ulaz neuronskih mreža i time eventualno dodatno poboljšati performanse.

Jedan od nedostataka tradicionalnih votermarking tehnika je što u šeme za umetanje ne uključuju informacije o raspodjeli iz koje se vrši uzorkovanje signala nosilaca. Na primjer, muzički signali generišu se iz drugačije raspodjele od govornih signala, i to treba iskoristiti u proceduri umetanja ukoliko se zna nad kojom klasom signala će sistem biti primjenjivan. Neuronske mreže implicitno modeluju distribucije svojih ulaza i na taj način integrišu znanje o svojstvima signala nosilaca. Ovo znanje se može iskoristiti za umetanje vodenih žigova, prilagođeno klasi signala, što

je izvedeno u ovom radu.

Dodatna prednost je što tehnike umetanja zasnovane na upotrebi dubokih neuronskih mreža nisu inverzibilne zbog velikog broja parametara i operacija koje se u njima vrše. Ovo ih čini otpornim na neautorizovano umetanje dok god su parametri ili arhitektura neuronskih mreža tajni.

Korišćenje neuronskih mreža ima najizraženiji uticaj na proceduru umetanja. Neuronske mreže mogu se obučiti da vrše redudantno umetanje bitova vodenog žiga, u djelovima signala koji imaju minimalan ili nikakav uticaj na ljudsku percepciju. Međutim, i drugi djelovi votermarking sistema mogu imati koristi od implementacije tehnika dubokog učenja. Ukoliko bi procedura detekcije bila realizovana kao duboka neuronska mreža, takvu proceduru bilo bi izuzetno teško rekonstuisati i vršiti neovlašćenu detekciju. Takođe, fabrikovanje lažnih primjera ovako kompleksnom detektoru bio bi teško izvodljiv poduhvat.

Postizanje otpornosti sistema na različite efekte i napade je najizazovniji zadatak za sisteme vodenog žiga. Mnoštvo i raznovrsnost efekata, odnosno napada, kojima signal može podleći prije detekcije predstavljaju odličan preduslov za primjenu neuronskih mreža. Iluzorno je očekivati da inženjeri predvide svaki scenario napada i svaki mogući efekat i preduzmu potrebne akcije za njegovu prevenciju. Neuronske mreže, sa određenim nivoom generalizacije, mogu se iskoristiti za suprotstavljanje različitim varijacijama velikog broja efekata. To se može postići otkrivanjem robustnih karakteristika signala nosioca i umetanjem bitova vodenog žiga u njih, ili inverzijom napada prilikom detekcije vodenog žiga. Inverzija napada bi se u tradicionalnim metodama morala vršiti direktnim pristupom, iscrpnim pretraživanjem svih mogućnosti i pozivanjem detekcije za svaku od njih. Ovo je izuzetno neefikasno i rezultovalo bi velikim brojem detekcija, čime se povećava šansa da dođe do greške. Neuronske mreže bi kroz brojne primjere mogle naučiti svojstva napada i efikasnije pronaći njegov inverz.

Dodatno, neuronske mreže se već koriste u napadima na votermarking sisteme prilikom generisanja dipfejkova. Stoga, logično bi bilo iskoristiti ih i na drugoj strani, za ublažavanje ili potpuno suzbijanje posljedica koje ovi falsifikovani signali mogu izazvati.

6.1 Arhitektura sistema

U osnovi naš sistem sastoji se iz dvije karike, umetača i detektora. Umetač prihvata dva ulaza, vodeni žig i signal nosilac i proizvodi signal sa umetnutim vodenim žigom. Umetač koristi svojstva ulaznog signala u obavljanju svojih zadataka što

ovu komponentu čini informisanom i, samim tim, značajno težom za invertovati. Detektoru se na ulaz dovodi signal sa umetnutim vodenim žigom, ili neki drugi signal koji nije prošao kroz umetač. Detektor u našem sistemu je neinformisan, jer, prema opisanoj postavci ulaza, ne raspolaže izvornim oblikom audio signala. Odsustvo vodenog žiga na ulazu detektora onemogućava primjenu korelacije pri detekciji, ali čini detektor znatno sigurnijim i primjenljivijim u praksi. Ustaljene i provjerene kriptografske tehnike kojima se sistemi vodenog žiga brane od neovlašćenog umetanja i neovlašćene detekcije mogu se nesmetano integrisati u ovaj sistem. Kreiranje osnovnih komponenti sistema vodenog žiga sa ovakvim odlikama je u skladu sa savremenim trendovima u digitalnom votermarkingu.

Nepovratne neuronske mreže su temeljni djelovi ovog sistema, dok rekurentne nisu razmatrane. Rekurentne mreže se koriste kada u ulaznim podacima postoji kauzalitet, odnosno kada vrijednosti odbiraka sa različitih pozicija u ulazu utiču na vrijednosti na drugim pozicijama. Iako u audio signalima uglavnom postoji uzročno-posljedična povezanost nekoliko uzastopnih frejmova, vodeni žig može se nezavisno dodavati u svakom od njih. Dovoljan je kontekst od nekoliko stotina ili hiljada susjednih odbiraka i ne moraju se uzimati u obzir dugoročnije zavisnosti. Među bitovima vodenog žiga najčešće ne postoje bilo kakvi odnosi, jer su oni nasumično izgenerisani, pa upotreba rekurentnih slojeva nad ovim podacima nije opravdana. Iz ovih razloga su, u ovom radu, za realizaciju votermarking sistema odabrane nepovratne neuronske mreže. Preciznije, ključni djelovi sistema vodenog žiga realizovani su kao konvolucione neuronske mreže, pošto se predložene arhitekture u najvećoj mjeri sastoje od konvolucionih slojeva. Konvolucija je sveprisutna operacija u obradi signala. Veliki broj efekata nad različitim vrstama signala može se realizovati konvolucijom, uz odgovarajuće filtre. Takođe, konvolucija se već uspješno koristi u oblasti votermarkinga, za umetanje vodenih žigova pomoću eho kernela, što je pomenuto u Sekciji 4.2. Ove odlike istakle su konvoluciju, odnosno konvolucione neuronske mreže, kao adekvatnu tehniku za realizaciju procedura za umetanje i detekciju vodenih žigova.

Jedna neuronska mreža u votermarking sistemu može istovremeno obavljati više uloga. Zadaci različitih djelova sistema ne moraju biti jasno razgraničeni kao što je opisano u Sekciji 2. Konvolucioni filtri u mreži umetača mogu vršiti transformaciju ulaznih reprezentacija signala nosioca i vodenog žiga u sintetičke domene, pogodne za sprovođenje procedure ugrađivanja. Na kraju, ista mreža može vršiti i rekonstrukciju originalnog signala iz sintetičkog domena, odnosno biti zadužena za očuvanje kvaliteta signala nosioca. Dakle, ova neuronska mreža može ujedno imati ulogu enkodera signala nosioca i enkodera vodenog žiga, kao i ulogu dekodera signala. Zadatke detektora takođe mogu obavljati jedna ili više mreža, zavisno od dizajna sistema. Jedan dio konvolucionih filtara detektora traži najpogodniji domen

za detekciju vodenog žiga, dok drugi djelovi služe za ekstrakciju bitova vodenog žiga, ukoliko su oni prisutni u signalu.

Od neuronske mreže za detekciju se dodatno očekuje da svoje zadatke uspješno izvrši i kada je signal nakon umetanja podlegao različitim efektima i napadima. Međutim, da bi sistem vodenog žiga bio robustan nije dovoljno dizajnirati robustnu šemu za detekciju, već je neophodna usklađenost komponete za umetanje i komponente za detekciju vodenog žiga. Umetač mora učestvovati u suzbijanju posljedica ovih efekata i napada. U suprotnom, vodeni žig bi se veoma lako mogao obrisati ili degradirati, bez obzira na složenost procedure detekcije. Na primjer, ukoliko umetač ugrađuje vodene žigove isključivo u visokim frekvencijama, niskopropusni filtar bi ga u potpunosti obrisao. Ako se prisjetimo da je pri svemu ovome potrebno i čuvati kvalitet signala, ovaj zadatak postaje utoliko teži. Dakle, procedure umetanja i detekcije moraju biti u sprezi zarad ispunjavanja zahtjeva robustnosti sistema vodenog žiga. Greška u detekciji, izazvana napadima, mora se proslijediti umetaču, kako bi se prilagodila i procedura umetanja. Stoga, naš pristup predviđa da se efekti i napadi na koje se želi postići otpornost sistema realizuju kao slojevi neuronske mreže. Ovi slojevi predstavljaju posebnu komponentu sistema koja se koristiti samo pri obučavanju i prevashodno služe za protok informacija o napadima između komponenti umetača i detektora kako bi se te dvije komponente mogle uskladiti i postići otpornost na definisane napade. Prilikom propagacije unaprijed ovi slojevi aproksimiraju realne scenarije u kojima je signal izobličen kako bi se izračunala greška detekcije u tim situacijama. U propagaciji unazad, izračunava se gradijent slojeva napada, i na taj način se greška detekcije distribuira i na komponentu umetača.

U Poglavlju 2 istaknuta je potreba za pravljenjem kompromisa prilikom razvoja sistema vodenog žiga. Kreiranje sistema koji je dominantan po svim kriterijumima smatra se gotovo nemogućim poduhvatom. Na osnovu planirane primjene sistema i okruženja u kojem će biti korišten definišu se primarni kriterijumi za ocjenu uspješnosti sistema, dok se ostali smatraju sporednima i, kada je to neophodno, nauštrb tih kriterijuma se ostvaruju zacrtani ciljevi po primarnim mjerilima performansi.

Duboke neuronske mreže pokazale su se uspješnim u radu sa sirovim (neprocesiranim) podacima jer su u stanju izvući neophodne karakteristike ulaza kako bi se riješio postavljeni zadatak. U različitim zadacima obrade audio signala su se vremenom ustalile reprezentacije poput standardnog ili Mel spektrograma [124], kao i standardnog [125] ili Mel-frekvencijskog kepstruma (MFC) [126]. Ove reprezentacije koriste se u algoritmima za automatsku transkripciju govora [127, 128], prepoznavanje govornika [129], klasifikaciju zvukova [130–132], procjenu sličnosti audio signala [133], itd. U rješavanju mnogih od ovih zadataka danas prednjače duboke neuronske

mreže. Međutim, još uvijek se nije odstupilo od korišćenja frekvencijskih reprezentacija, jer su eksperimenti pokazali da njihova primjena na ulazu neuronske mreže vodi boljim rezultatima. Na primjer, savremeni enkoder-dekoder modeli za transkripciju govora [128] obučavani su na Mel spektrogramima govornih signala. Uzrok ovakvih rezultata se može tražiti u pretpostavci da se računanjem ovih vrijednosti vrši inicijalna ekstrakcija karakteristika signala i time ubrzava proces obučavanja i pojednostavljuje arhitekture neuronskih mreža, jer nije potrebno trošiti podatke i vrijeme, niti kreirati dodatne slojeve za ekstrakciju karakteristika iz sirovih signala.

Upotreba Mel skale i kepstralnih koeficijenata u domenu votermarkinga nije podrobno ispitana. Sprovedeno je svega nekoliko studija o njihovom potencijalu u ovoj oblasti [78, 134]. Najveći izazov u korišćenju reprezentacija poput MFC u votermarkingu je njihova slaba otpornost na šum. Pojava šuma na jednoj frekvenciji mijenja više Mel-frekvencijskih kepstralnih koeficijenta, jer se ovi koeficijenti računaju primjenom banke filtara među kojima postoje preklapanja. Dodatan problem je što ove reprezentacije obično uzimaju u obzir samo spektar snage (intenzitet) signala, zanemarujući njegove ostale karakteristike. Kepstrum je prvobitno definisan kao kepstrum snage (*engl. power cepstrum*). Eliminacija dijela informacija koje signal nosi mogla bi spriječiti ekstrakciju karakteristika koje su od presudnog značaja za umetanje vodenog žiga. Informacija o rasporedu energije u audio signalu je sasvim dovoljna za prethodno pomenute sisteme, koji kao izlaz ne daju audio signal, već vrše preslikavanja u druge prostore, nevezane za audio, kao što su realni brojevi, tekst ili ograničeni skup klasa. Međutim, sistem za umetanje vodenih žigova u suštini vrši transformaciju svog ulaza, pa je prostor u koji se vrši preslikavanje takođe prostor audio signala. Stoga bi korišćenje ovih komprimovanih reprezentacija iziskivalo dodatne procedure za sintetisanje signala nakon umetanja vodenog žiga kojima bi se potencijalno ugrozio kvalitet. Nabrojani razlozi uslovili su korišćenje reverzibilnih frekvencijskih reprezentacija u našem sistemu, za pripremu audio signala, prije njihovog prosljeđivanja neuronskim mrežama na dalju obradu.

U ovoj disertaciji razmatraju se dva modela arhitekture sistema vodenog žiga. Prvi model obezbjeđuje visoku otpornost na različite efekte, uključujući i desinhronizujuće, kao i očuvanje kvaliteta signala. Međutim, superiorne performanse po tim kriterijumima rezultovale su limitiranim kapacitetom ovog modela. Iz tog razloga je kreiran i model B koji omogućava umetanje znatno većeg broja bitova u jedinici vremena. Povećanje kapaciteta se u tradicionalnim votermarking sistemima izrazito negativno odražava na očuvanje kvaliteta signala. Koncipiranje modela B izvedeno je s posebnom težnjom da se ovaj problem ne pojavi, odnosno da se kvalitet signala zadrži na visokom nivou. Takođe, zadržana je otpornost na standardne efekte, čime je zadovoljen potrebni nivo robustnosti sistema vodenog žiga, dok je otpornost

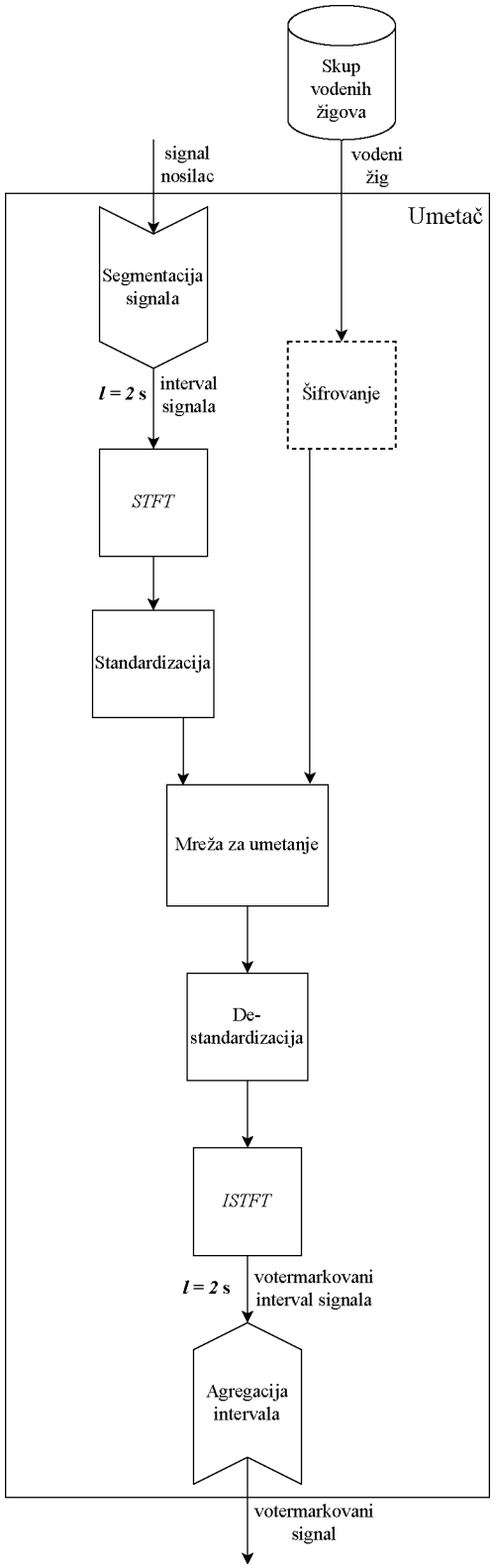
na efekte desinhronizacije marginalizovana. Ovim su ponuđena rješenja koja mogu pokriti različite realne scenarije primjene sistema vodenog žiga i dat je doprinos kompletnosti ove studije. U nastavku poglavlja detaljno su opisana pomenuta dva modela, kao i arhitekture neuronskih mreža koje ih čine.

6.1.1 Model A

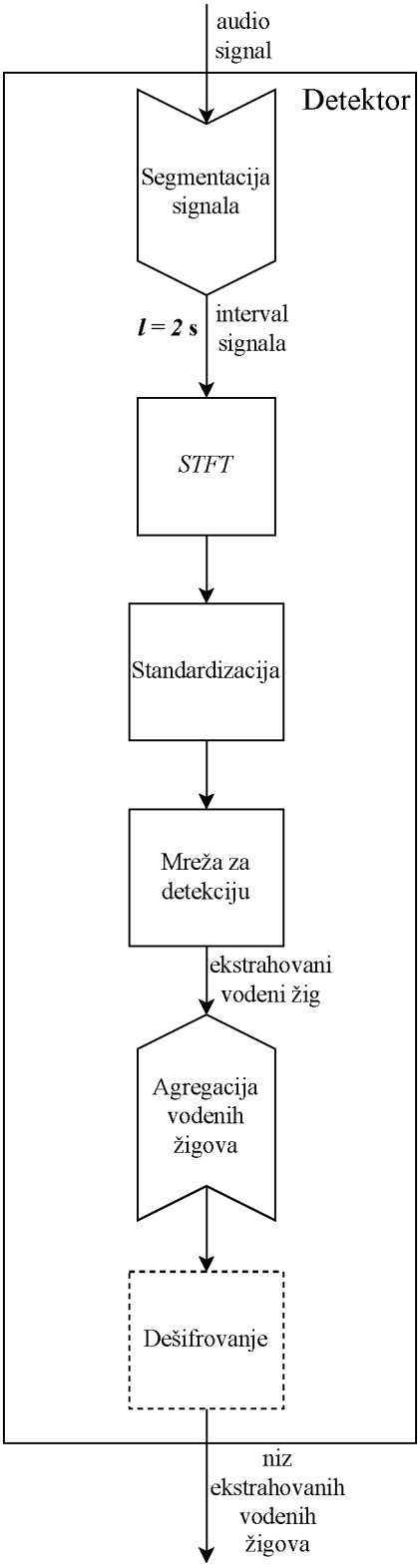
Model A prezentovan je u radovima [135–138]. Prema tom modelu, sistem vodenog žiga sastoji se od dvije komponente: umetača i detektora, u čijem središtu su dvije neuronske mreže, za umetanje i detekciju. Struktura i komponente umetača skicirane su na Slici 14, dok Slika 15 prikazuje konfiguraciju detektora. Iako se ključni koraci obrade u komponentama ovog sistema obavljaju neuronskim mrežama, umetač i detektor sadrže i prateće djelove, neophodne kako bi se obezbijedila potpuna funkcionalnost i primjenljivost sistema. Za reprezentaciju ulaza mreža za umetanje i detekciju u modelu A odabrana je kratkotrajna Furijeova transformacija (*short-time Fourier transform* - STFT). Ona je reprezentovana kao dvokanalna matrica, sa po jednim kanalom za realne i imaginarne djelove koeficijenata transformacije. Vrijednosti u oba kanala su standardizovane na nivou čitavog skupa za obučavanje, prije prosljeđivanja odgovarajućoj neuronskoj mreži. Nakon umetanja, vrijednosti STFT su vraćene u originalnu skalu kako bi se odbirci audio signala mogli pravilno rekonstruisati. Eksperimenti u radu [136] sprovođeni su i sa ulaznim signalima u vremenskom domenu. Međutim, ova reprezentacija se i u ovom slučaju pokazala inferiornom, pogotovo u pogledu robustnosti. Ovo se može opravdati činjenicom da veliki broj efekata, poput niskopropusnog filtriranja i šuma, mijenja sve odbirke signala u vremenu. U frekvencijskom domenu bi ovi efekti izmijenili samo pojedine grupe koeficijenata, dok bi se preostale nesmetano mogle koristiti za umetanje i detekciju vodenih žigova.

STFT je reverzibilna operacija, jer zadržava informaciju i o amplitudi i fazi signala, tj. audio signal se može rekonstruisati bez gubitaka. Stoga se STFT može koristiti u sistemima vodenog žiga bez negativnog uticaja na očuvanje kvaliteta signala. Dodatno, korišćenje STFT, kao vremensko-frekventne reprezentacije, omogućava umetaču da bitove vodenog žiga doda u odgovarajućim trenucima u vremenu, umjesto samo na određenim frekvencijama, kao što bi to bio slučaj sa frekvencijskim reprezentacijama poput DFT.

Zbog navedenih razloga STFT se koristi za kodiranje audio signala u modelu A. U Prilogu B izloženi su neophodni matematčki instrumenti za izračunavanje STFT. Nad vodenim žigom nisu vršene transformacije prije samog umetanja ($v = w$). On je ostavljen u originalnom obliku niza od L_w bitova i kao takav proslijeđen mreži ume-



Slika 14: Blok šema umetača za model A.



Slika 15: Blok šema detektora za model A.

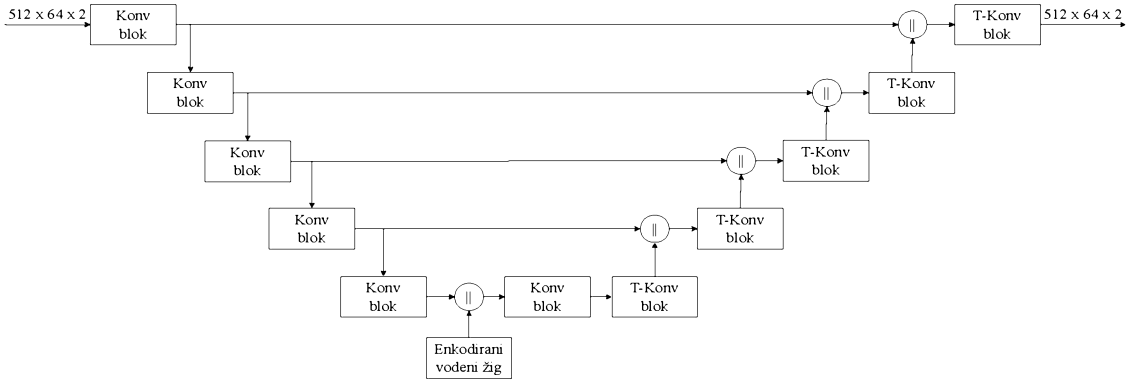
tača. Iz arhitekturalnih razloga uzeto je $L_w = 256$. Umetanje vodenih žigova se vrši u segmentima audio signala u trajanju od 2 sekunde. Signal se na početku procedure dijeli na intervale ove dužine. To iziskuje i dodavanje komponente za spajanje (agregaciju) ovih intervala unutar umetača (Slika 14), nakon što se neuronskom mrežom ugrade bitovi vodenog žiga u svaki od njih. Isti postupak segmentacije audio signala sprovodi se i na početku procedure detekcije, što je prikazano na Slici 15.

Prilikom dizajniranja modela A prioritet je dat robustnosti i očuvanju kvaliteta signala. Kompromis je pravljen u pogledu kapaciteta na način što je ograničavan broj vodenih žigova koje je sistem u stanju da ugradi. Prije obučavanja, slučajno je izgenerisan skup od $N_w + 1$ vodenih žigova koji se koriste za umetanje. Jedan vodeni žig je izdvojen iz skupa i nije korišten za umetanje, već za prepoznavanje nevotermarkovanih signala. Ukoliko detektor na izlazu da taj vodeni žig, smatra se da ulazni signal nije votermarkovan. Vrijednost parametra N_w zavisi od broja efekata na koje se želi postići otpornost. U situaciji kada je sistem testiran na otpornost od ustaljenih efekata (šum i niskopropusno filtriranje), uzeto je $N_w = 8$. Dostizanje otpornosti na sve implementirane efekte desinhronizacije u [137] prouzrokovalo je smanjenje broja vodenih žigova na $N_w = 2$. Svakako, i sa ovim brojem vodenih žigova moguće je kodirati proizvoljnu poruku i prenijeti je audio signalom adekvatne dužine.

6.1.1.1 Arhitektura neuronske mreže umetača

Kao polazište za dizajn neuronske mreže za umetanje vodenih žigova uzet je U-net model [139]. Ova arhitektura inicijalno je predložena za segmentaciju biomedicinskih slika. Međutim, kasnije je primjenjena u različitim zadacima [140–145] i postala je jedan od najznačajnih dizajn koncepata u dubokom učenju. Arhitektura mreže umetača u okviru modela A, koja je kreirana po uzoru na U-net, skicirana je na Slici 16. Ova mreža namijenjena je za obavljanje zadataka kodiranja signala nosioca i vodenog žiga, kao i za dekodiranje signala nakon umetanja.

U-net arhitektura sastoji se iz dva dijela: enkodera i dekodera. Slojevi enkodera pronalaze sintetičku reprezentaciju ulaza u prostoru niže dimenzije. Ova komponenta prati tipičnu arhitekturu konvolucione mreže. Sadrži niz blokova kojima se vrši smanjenje dimenzionalnosti. Jedan blok uključuje konvolucioni sloj, normalizaciju po seriji i aktivacionu funkciju. Pri svakom koraku decimacije, odnosno smanjenja dimenzionalnosti, broj filtara u konvolucionom sloju se duplira. U originalnom radu je predloženo da se za smanjivanje dimenzija koriste slojevi agregacije koji su sastavni dio blokova. Međutim, zbog značajnog gubitka informacija koji ovi slojevi prouzrokuju, oni ovdje nisu korišteni. Decimacija se vrši povećavanjem veličine



Slika 16: Skica neuronske mreže za umetanje u modelu A. Simbol „||" označava operaciju nadovezivanja (konkatenacije).

koraka u konvolucionim slojevima.

Dekoder je simetričan enkoderu, čime se dobija arhitektura u obliku latiničnog slova „U", odakle je ova mreža i dobila svoj naziv. Slojevi dekodera signal iz latentnog prostora vraćaju u prostor koji svojom dimenzionalnošću odgovara izvornom prostoru signala. Za povećavanje dimenzionalnosti koriste se blokovi koji se sastoje od transponovane konvolucije, normalizacije po seriji i odabrane aktivacione funkcije. Broj filtara se polovi prilikom svakog povećanja dimenzija kako bi izlaz sloja po veličini odgovarao sloju enkodera na istoj dubini. U cilju boljeg prenosa informacija vrši se premošćavanje između odgovarajućih slojeva enkodera i dekodera, kao što je prikazano na Slici 16. Na ovaj način, informacije sa nižeg nivoa apstrakcije mogu značajnije uticati na izlaz iz neuronske mreže. Na kraju dekodera opciono se dodaje još jedan decimacioni blok kako bi se dobio izlaz s odgovarajućim brojem kanala, ukoliko taj broj nije jednak broju kanala na ulazu.

U originalnom radu [139], U-net se koristi za segmentaciju, odnosno dijeljenje slike na regione od interesa. Ulaz ove procedure je slika, a izlaz je segmentaciona maska u kojoj je za svaki piksel slike označeno kojem regionu pripada. Međutim, ovo ne odgovara zadacima umetača, pa se iz tog razloga U-net mreža, u ovom radu, koristi kao autoenkoder. Autoenkoder je vrsta neuronske mreže koja se obučava da nauči identitetsku funkciju $f(x) = x$, odnosno da iskopira ulaz na izlaz. Iako naizgled djeluje kao da rješava trivijalan zadatak, vrijednost autoenkodera leži u preslikavanju originalnog prostora u prostor manje dimenzije, odnosno svojevrsnoj kompresiji podataka. U-net sadrži sve komponente koje čine jedan autoenkoder. Najprije se enkoderom ulaz predstavlja u latentnom prostoru, manje dimenzionalnosti, a zatim se dekoderom vraća u prostor originalnih dimenzija, kao što se može vidjeti na Slici 16. Taj niz transformacija kroz koje signal prolazi u U-net mreži donekle odgovara proceduri umetanja vodenih žigova. Enkoderom se može tražiti reprezentacija sig-

Tabela 1: Postavke arhitekture neuronske mreže umetača u okviru model A.

Blok	Broj filtara	Veličina filtra/koraka	Veličina izlaza
Decimacioni	16	5×5 / 2×2	$256 \times 32 \times 16$
Decimacioni	32	5×5 / 2×2	$128 \times 16 \times 32$
Decimacioni	64	5×5 / 2×2	$64 \times 8 \times 64$
Decimacioni	128	5×5 / 2×2	$32 \times 4 \times 128$
Decimacioni	256	5×5 / 2×2	$16 \times 2 \times 256$
-dodavanje vodenog žiga-			
Decimacioni	256	5×5 / 1×1	$16 \times 2 \times 256$
Interpolacioni	128	5×5 / 2×2	$32 \times 4 \times 128$
Interpolacioni	64	5×5 / 2×2	$64 \times 8 \times 64$
Interpolacioni	32	5×5 / 2×2	$128 \times 16 \times 32$
Interpolacioni	16	5×5 / 2×2	$256 \times 32 \times 16$
Interpolacioni	8	5×5 / 2×2	$512 \times 64 \times 8$
Decimacioni	2	5×5 / 1×1	$512 \times 64 \times 2$

nala nad kojom se vrši umetanje vodenog žiga, a zatim se dekoderovim slojevima može izvršiti rekonstrukcija signala. Kako je vodeni žig definisan u prostoru manje dimenzije, hipoteza je da će njegovo dodavanje biti moguće u latentnom prostoru koji kreira enkoder. Ipak, dodavanjem vodenog žiga narušava se koncept autoenkodera. Funkcija gubitka ove mreže definiše se tako da je cilj procedure obučavanja minimizovanje razlike između ulaza i izlaza modela. Ovo može predstavljati prepreku u njegovom korišćenju za umetanje vodenog žiga, jer bi autoenkoder težio da u potpunosti izbriše vodeni žig kako bi postavljeni zadatak bio što uspješnije riješen. Zbog toga se na izlaz umetača veže mreža za detekciju koja za zadatak ima da spriječi ovu pojavu. Način na koji se ovo postiže biće objašnjen u Sekciji 6.2.

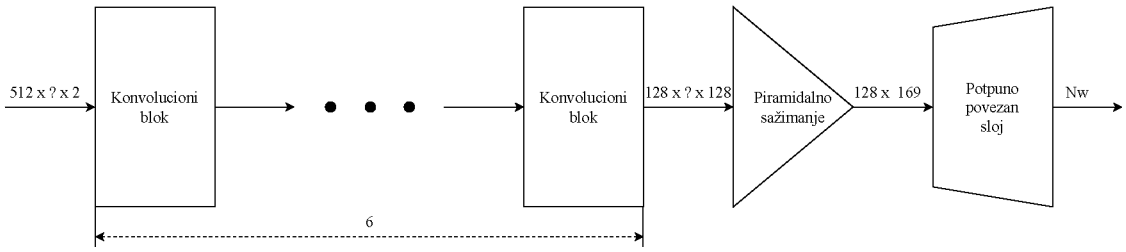
U Tabeli 1 dati su svi relevantni podaci o dizajnu arhitekture neuronske mreže za umetanje, koja je skicirana na Slici 16. U-net arhitektura iz originalnog rada [139] nije u potpunosti replicirana, već je redizajnirana kako bi eksperimentalnim postupkom dovela do najboljih performansi. U tabeli su, radi sažetijeg zapisa, isključivo dati podaci o konvolucionim slojevima. Iako mreža sadrži i druge vrste slojeva, koji će biti pomenuti u nastavku, detalji ovih slojeva nisu navedeni jer su uniformni u čitavoj mreži. U svim djelovima umetača kao aktivaciona funkcija korištena je popustljiva ReLU. Parametru v je empirijskim putem dodijeljena vrijednost 0.2. Inicijalizacija parametara vršena je Hiovom metodom, jer je ona prikladnija za korišćenje sa aktivacionim funkcijama koje nisu simetrične u odnosu na koordinatni početak.

Signal nosilac se na ulaz mreže dovodi u obliku tenzora dimenzija $512 \times 64 \times 2$. Prve dvije dimenzije su frekvencija i vrijeme. Broj 512 predstavlja broj frekvencijskih koeficijenata koji se izračunavaju u okviru STFT, a 64 je broj koraka napravljenih prilikom pomjeranja prozora signalom. S obzirom na to da je konvolucija u našem modelu definisana u polju realnih brojeva, a da su koeficijenti STFT kompleksni brojevi, ulaz je podijeljen na dva kanala koji redom predstavljaju realni i imaginarni dio Furijeove transformacije, čime je uvedena i treća dimenzija na ulazu. Ovakva reprezentacija ulaza uslovljava primjenu 2D konvolucije u svim konvolucionim slojevima modela.

Enkoder dio izgrađen je od 5 decimacionih blokova. U okviru jednog decimacionog bloka se nad ulazom primjenjuju tri operacije. Prvo se vrši konvolucija ulaza sa svim filtrima u bloku. Zatim se vrši normalizacija po seriji dobijenog izlaza konvolucije, prije nego se primjeni aktivaciona funkcija. Prvi decimacioni blok sadrži 16 filtara dimenzija 5×5, koji se primjenjuju s korakom veličine 2 po obje dimenzije. Ovo rezultuje izlazom dimenzija $256 \times 32 \times 16$, gdje je 256×32 dimenzija izlaza konvolucije, a 16 je broj filtara. U svakom narednom bloku duplira se broj filtara, a dimenzije izlaza se polove. Nakon posljednjeg, petog decimacionog bloka, ulazni signal sveden je na latentnu reprezentaciju dimenzija $16 \times 2 \times 256$.

Umetanje u latentnom domenu se vrši nadovezivanjem vodenog žiga na reprezentaciju signala nosioca koju je proizveo enkoder. Međutim, bez obzira da li se vrši nadovezivanje ili neka druga operacija, poput sabiranja ili množenja, između latentne reprezentacije signala nosioca i vodenog žiga postoje neslaganja u dimenzijama koja se moraju prevazići. Ovo se postiže tako što se vodeni žig od 256 bitova replicira do dimenzija $16 \times 2 \times 256$, kako bi veličinom odgovarao izlazu enkodera. Nadovezivanje se vrši po dimenziji kanala. Zatim se dobijeni tenzor, dimenzija $16 \times 2 \times 512$, provlači kroz još jedan decimacioni blok sa 256 filtara kako bi se dobio izlaz podesnog oblika za rekonstrukciju dekoderom i kako se ne bi narušila simetrija koju zahtijeva U-net arhitektura. Ovaj decimacioni blok razlikuje se od prethodnih po tome što se konvolucija primjenjuje s korakom veličine 1, što rezultuje nepromijenjenim dimenzijama na izlazu.

Dekoder nizom od 5 interpolacionih blokova vrši rekonstrukciju originalnog signala. Interpolacioni blok se takođe sastoji od tri sloja: transponovane konvolucije, normalizacije po seriji i aktivacione funkcije, koji se primjenjuju redom kojim su navedeni. Rekonstrukcija počinje blokom od 128 filtara, a zatim se, simetrično enkoderu, broj filtara polovi, a dimenzije izlaza dupliraju sa svakim narednim blokom. Prije svakog sljedećeg T-konv bloka vrši se nadovezivanje izlaza prethodnog i izlaza odgovarajućeg bloka enkodera, što je na Slici 16 označeno simbolom „||". Dodavanje ovih veza motivisano je ResNet arhitekturom, pomenutoj u Sekciji 5.5.3. Kako U-net



Slika 17: Skica arhitekture detektora za model A.

mreža ima veliki broj slojeva, ovo služi izbjegavanju problema nestajućih gradijenata, ubrzavanju procedure obučavanja i, u krajnjem, postizanju boljih performansi. Veličina filtara u dekoderu ista je kao i u enkoder dijelu mreže. Na kraju ovog niza blokova dobija se izlaz oblika $512 \times 64 \times 8$.

Kako bi se dobio izlaz koji po svom obliku odgovara originalnoj reprezentaciji signala nosioca primjenjuje se finalni decimacioni blok sa 2 filtra dimenzija 5×5 i korakom veličine 1. U ovom bloku se, za razliku od ostalih blokova umetača, ne može koristiti propustljiva ReLU aktivaciona funkcija. Ova funkcija nije simetrična u odnosu na koordinatni početak, pa se njome ne može na odgovarajući način modelovati izlaz koji treba da predstavlja kratkotrajnu Furijeovu transformaciju audio signala. Stoga se u posljednjem bloku umetača kao aktivaciona funkcija koristi funkcija identiteta i Glorotova tehnika inicijalizacije.

Prilagođavanje ove arhitekture za obradu sirovih audio signala zahtijeva prelazak sa 2D konvolucije na 1D konvoluciju. Takođe, nužna je i promjena dimenzija konvolucionih filtara. Po ugledu na WavNet model [146] za generisanje sirovih audio sekvenci, korišteni su filtri dužine 41 za prvi i posljednji, a dužine 21 za preostale blokove U-net arhitekture.

6.1.1.2 Arhitektura neuronske mreže detektora

Zadatak detektora je da prepoznaje da li je, i kojim vodenim žigom je označen audio signal. Ovaj zadatak se može smatrati zadatkom klasifikacije, jer je broj vodenih žigova u modelu A ograničen, pa je mreža detektora dizajnirana po ugledu na neuronske mreže za klasifikaciju. Kompletna skica arhitekture ove neuronske mreže prikazana je na Slici 17. Mreža se sastoji od nekoliko decimacionih blokova koji izvlače ključne karakteristike ulaznog signala na osnovu kojih će se izvršiti klasifikacija. Ovaj dio mreže za detekciju ima sličnu ulogu kao enkoder u umetaču, tj. traži odgovarajući domen za rješavanje postavljenog problema. Kao i kod umetača, decimacioni blok sadrži jedan konvolucioni sloj, sloj normalizacije po seriji i aktivacionu funkciju, respektivno. Na izlazu neuronske mreže za detekciju nalazi se

Tabela 2: Postavke arhitekture neuronske mreže detektora u okviru model A.

Blok	Broj filtara	Veličina filtra/koraka	Veličina izlaza
Decimacioni	32	5×5 / 2×2	$256 \times ? \times 32$
Decimacioni	32	5×5 / 2×2	$128 \times ? \times 32$
Decimacioni	64	5×5 / 1×2	$128 \times ? \times 64$
Decimacioni	64	5×5 / 1×2	$128 \times ? \times 64$
Decimacioni	128	5×5 / 1×2	$128 \times ? \times 128$
Decimacioni	128	5×5 / 1×2	$128 \times ? \times 128$
Piramidalna agregacija		$(1 \times ?), (4 \times ?)$ $(16 \times ?), (128 \times ?)$	169×128
Potpuno povezani			L_w

potpuno povezan sloj sa sigmoid aktivacijom. Svrha ovog sloja je da za svaki od bitova vodenog žiga izračuna vjerovatnoću da je njegova vrijednost 1. Broj izlaznih vrijednosti potpuno povezanog sloja jednak je dužini vodenog žiga L_w. Na osnovu vjerovatnoća $P(k)$, $k \in \{1, 2, ..., L_w\}$ koje generiše ovaj sloj vrši se ekstrakcija bitova vodenog žiga po sljedećem pravilu:

$$\hat{v}(k) = \begin{cases} 1, & P(k) > 0.5 \\ 0, & P(k) \leq 0.5. \end{cases} \tag{85}$$

Prag je očekivano postavljen na vrijednost od 0.5 s obzirom da se oba bita približno jednako često pojavljuju u svim vodenim žigovima u skupu. Sve vrijednosti na izlazu mreže detektora koje prelaze ovu granicu smatraju se bližim vrijednosti 1 i ekstrahuju se kao taj bit, a sve vrijednosti ispod granice kao bit 0.

Neposredno prije potpuno povezanog sloja dodat je sloj za piramidalnu agregaciju. Uključivanje ovog sloja u arhitekturu detektora nametnuto je postojanjem efekata desinhronizacije. Ovi efekti mogu uzrokovati promjenu dimenzije signala u vremenu, a kako izlaz detektora mora biti fiksne veličine, kao i sami vodeni žig, neophodno je, prije ekstrakcije bitova, dovesti ulaz na fiksnu veličinu. Ovo se postiže piramidalnom agregacijom koja je dodata nakon decimacionih blokova. Time je omogućeno da konvolucioni slojevi vrše ekstrakciju karakteristika direktno iz ulaznog signala i izbjegnut je nepotrebni gubitak informacija u inicijalnim slojevima mreže detektora, što dovodi do preciznije ekstrakcije bitova vodenog žiga.

Tabela 2 sadrži podatke o arhitekturi mreže detektora. Ova tabela, analogno tabeli za mrežu umetača, sadrži isključivo specifikacije konvolucionih slojeva. U skladu sa dobrom praksom da se u svim unutrašnjim slojevima mreže koriste ista aktivaciona funkcija i šema za inicijalizaciju, aktivaciona funkcija svih decimacionih blokova

detektora je propustljiva ReLU, sa parametrom $v = 0.2$ i primjenjivana je Hiova tehnika inicijalizacije.

Ulaz detektora je audio signal reprezentovan kratkotrajnom Furijeovom transformacijom, koji je potencijalno podlegao dejstvu nekih efekata. Broj procijenjenih frekvencijskih koeficijenata ostaje 512. Međutim, uslijed postojanja efekata desinhronizacije, broj vremenskih intervala je nepoznat. Početni dio detektora sačinjen je od 6 decimacionih blokova. Prvi par blokova sadrži po 32 filtra dimenzije 5×5 koji se primjenjuju s korakom veličine 2 po obje dimenzije. Naredna dva para decimacionih blokova sadrže po 64 i 128 filtara, respektivno. U ovim slojevima je veličina koraka za računanje konvolucije po dimenziji frekvencije smanjena na 1, što frekventnu dimenziju održava konstantnom, dok se veličina vremenske dimenzije polovi nakon svakog bloka.

Nakon ekstrakcije mape karakteristika ulaznog signala nizom decimacionih blokova vrši se piramidalna agregacija tih karakteristika kroz 4 nivoa rezolucije. Na svim nivoima koristi se agregatna funkcija maksimuma. Veličine regiona na kojima se primjenjuje agregatna funkcija su fiksne veličine po dimenziji frekvencije, a adaptivni po vremenskoj dimenziji, kako bi se dobio izlaz fiksne veličine. Prvim nivoom piramidalne rezolucije nastoje se očuvati vrijednosti iz svih frekvencijskih opsega. Stoga je veličina regiona za agregaciju po tom nivou jednaka 1, što će dati 128 vrijednosti na izlazu. U narednim nivoima se povećavaju dimenzije ovog regiona, odnosno smanjuje rezolucija. Na posljednjem nivou, region za agregaciju obuhvata čitav ulaz, tj. izračunava se globalni maksimum. Sve ovo u zbiru daje $128+32+16+1 = 169$ agregiranih vrijednosti za svaki kanal ulaza, odnosno izlaznu matricu dimenzija 169×128. Konačno, sve vrijednosti u izlazu dobijenom piramidalnom agregacijom se nadovezuju u jedan vektor od $169 \cdot 128 = 21632$ elemenata da bi se proslijedile potpuno povezanom sloju koji daje konačan izlaz detektora.

Slojevi detektora za signale u vremenskom domenu definisani su analogno slojevima umetača za istu vrstu signala. Korišćena je 1D konvolucija, s filtrima dimenzija 41 u prvom, 21 u drugom i trećem, 11 u četvrtom i petom i 9 u posljednjem, šestom decimacionom bloku detektora.

6.1.2 Model B

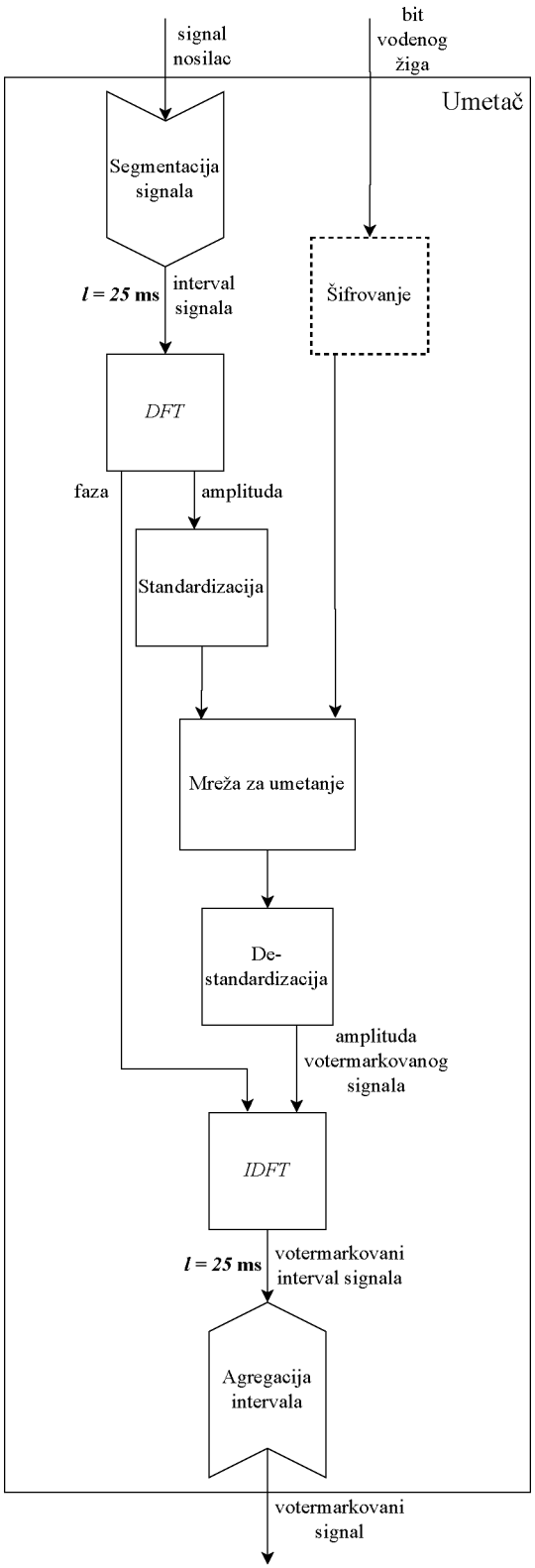
Prethodno opisani model A može se integrisati u moderne sisteme vodenog žiga i primijeniti u različitim scenarijima. Međutim, postoji osnov za njegova dalja poboljšanja, prvenstveno u pogledu kapaciteta. Povećanje kapaciteta stvorilo bi prostor za nove primjene ovog sistema i unapređenja postojećih. Votermarking sistemom većeg kapaciteta moguće je, pored autentifikacionih bitova, ugraditi i neke druge

informacije u signal. Dodatni bitovi mogu se iskoristiti za upisivanje informacija o vlasniku i drugih metapodataka u sistemima za zaštitu autorskih prava, pomenutim u Sekciji 1.3. *Second screen* aplikacije bi takođe imale očiglednih benefita od povećanja kapaciteta sistema vodenog žiga, jer bi se veća količina podataka mogla isporučiti korisnicima, čime bi se obogatio njihov sadržaj. Umetanje većeg broja bitova vodenog žiga u jedinici vremena doprinosi i sigurnosti votermarking sistema. Uzorkovanjem vodenih žigova iz većeg skupa otežava se neautorizovanim korisnicima da ga rekonstruišu, izbrišu ili promijene.

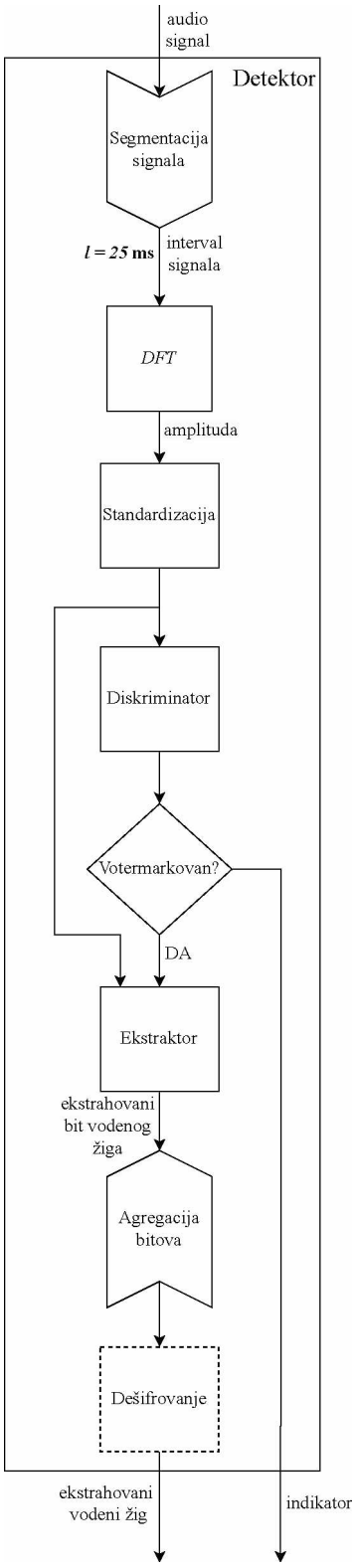
Model B osmišljen je upravo s ciljem povećanja kapaciteta modela A. U dizajnu modela A uočeno je nekoliko nedostataka koji su označeni kao uzroci manjeg kapaciteta. Ovi nedostaci su novim modelom adresirani i otklonjeni, dok su ostali djelovi sistema, koji su se pokazali efikasnim, sačuvani. Skice komponenti za umetanje i detekciju u okviru modela B date su na Slici 18 i Slici 19, respektivno.

U srži modela B, kao i modela A nalazi se operacija konvolucije, koja je dokazala svoju djelotvornost u svim aspektima jednog sistema vodenog žiga. Stoga se i u modelu B većina koraka obrade signala i vodenog žiga obavlja konvolucionim filtrima. Dakle, i model B predstavlja skup uvezanih konvolucionih neuronskih mreža. Zadržana je i ideja korišćenja frekvencijskih reprezentacija koja se pokazala uspješnom u modelu A, ali ne u cjelosti. Kao jedan od mogućih problema u modelu A identifikovano je korišćenje vremensko-frekventne reprezentacije, preciznije, uvođenje vremenske dimenzije u reprezentaciju ulaznog signala. Postojanje vremenske dimenzije navodi proceduru obučavanja na liniju manjeg otpora koja dovodi do umetanje bitova vodenog žiga samo u određenim („pogodnim") intervalima signala nosioca, ograničavajući pritom njegovu dužinu. Ovaj efekat se može djelimično spriječiti odgovarajućim napadima koji brišu pojedine intervale signala, primoravajući time sistem na ugradnju bitova vodenog žiga na različitim lokacijama. Međutim, čak i nakon najrazornijih efekata ostaju intervali u signalu koji se mogu podijeliti na „pogodne" i „nepogodne" za umetanje, što treba izbjeći i vršiti umetanje u svim djelovima signala.

Kako bi se obezbijedilo prostiranje bitova vodenog žiga duž čitavog signala, a ujedno ostvarilo i povećanje kapaciteta, signal nosilac je podijeljen na kratke vremenske intervale i u svaki interval je ugrađivan po jedan bit vodenog žiga. Za obradu su korištene samo amplitudne vrijednosti frekvencijskih komponenti (31), dok su faze komponenti signala (32) uklonjene iz procedure umetanja. Vrijednosti amplitude su standardizovane prije, a destandardizovane nakon umetanja bita vodenog žiga. Faza je dodavana tek po završetku umetanja kako bi se, prema jednakosti (33), dobili koeficijenti DFT, a zatim i rekonstruisao signal u vremenu sa inverznom DFT. Ova odluka podržana je činjenicom da je ljudsko uho gotovo neosjetljivo na promjene u



Slika 18: Blok šema umetača za model B.



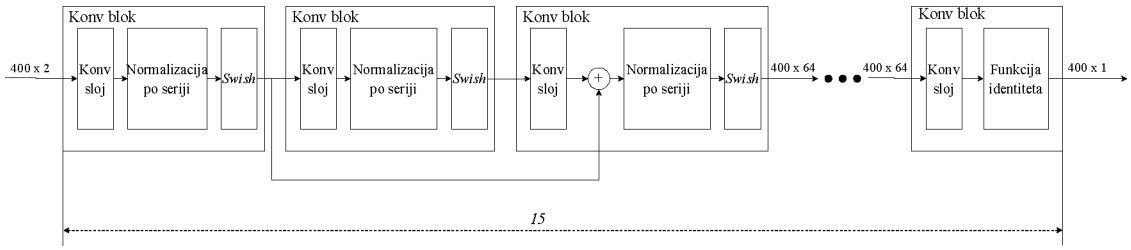
Slika 19: Blok šema detektora za model B.

fazi, pa bi korišćenje tog kanala za umetanje bilo kakvih informacija učinilo sistem veoma osjetljivim, čak i na jednostavne efekte i napade. Odabrana dužina intervala je 25 ms kako bi sistem postigao kapacitet od 40 bps čime se, u ovom pogledu, dostižu performanse konkurentne sa drugim votermarking sistemima. Opisana procedura skicirana je na Slici 18.

Umetač modela A nizom konvolucionih blokova transformiše ulaznu reprezentaciju signala, nakon čega se vrši spajanje sa vodenim žigom. Međutim, za očekivati je da se umetanje vodenog žiga može efikasnije obaviti ako bi njegova obrada započela istovremeno sa obradom signala nosioca. Iz tog razloga se u modelu B kodirani vodeni žig dovodi na ulaz prvim slojevima umetača, zajedno sa ulaznom reprezentacijom audio signala kako bi procedura umetanja žiga počela već sa prvim koracima obrade. Bitovi vodenog žiga predstavljeni su vrijednostima -1 i 1 kako bi i taj ulaz neuronske mreže bio centriran u nuli, što, kako praksa pokazuje, vodi ka stabilnijoj proceduri obučavanja i boljim rezultatima. Osim toga, ovim je osigurano da se opsezi oba ulaza mreže poklapaju, što olakšava inicijalizaciju parametara mreže i vodi stabilnijoj i efikasnijoj proceduri obučavanja. Opseg vrijednosti u latentnoj reprezentaciji signala koju proizvede enkoder modela A nije unaprijed poznat, pa samim tim nije moguće izvršiti adekvatno skaliranje vodenog žiga. Neusklađenost opsega ulaza mreže može dovesti do nesrazmjernog uticaja ulaznih atributa na izlazne vrijednosti, što može usporiti ili potpuno zaustaviti obučavanje. Dodatno, ovakva reprezentacija vodenog žiga olakšava primjenu regresionih metoda u procesu detekcije.

Detektoru modela A dodijeljena su dva zadatka, ako izuzmemo postizanje robustnosti. Prvo treba ispitati da li signal uopšte sadrži vodeni žig, nakon čega se ekstrahuju njegovi bitovi. Kreiranjem posebnih djelova sistema, predviđenih za obavljanje ovih zadataka, rasterećuje se detektor. Detektor modela B podijeljen je na dvije komponente: diskriminator i ekstraktor, koje obavljaju posebne zadatke. Diskriminator razdvaja votermarkovane i nevotermarkovane signale, a ekstraktor iz votermarkovanih signala izdvaja bitove žiga. Sumarna arhitektura detektora u okviru modela B može se vidjeti na Slici 19.

Povećanje kapaciteta, a pogotovo drastično kakvo je ostvareno modelom B, zahtijeva kompromise u drugim aspektima performansi. Olakšica je što primjene sistema vodenog žiga sa velikim kapacitetom, poput *second screen* aplikacija, pretežno ne iziskuju otpornost na veliki broj efekata. Shodno tome je odlučeno da se kompromis napravi u pogledu robustnosti. Efekti desinhronizacije, koji obično ne predstavljaju ustaljene mehanizme obrade signala, već maliciozne radnje, za kojima u ovom scenariju nema opravdanih razloga, su zanemareni. Predstavnici grupe klasičnih efekata su zadržani kako bi se obezbijedio potreban nivo robustnosti i proširile mogućnosti primjene ovog sistema.



Slika 20: Skica arhitekture neuronske mreže za umetanje u okviru modela B.

6.1.2.1 Arhitektura neuronske mreže umetača

U modelu B se za izvršavanje glavnih koraka u procesu umetanja vodenih žigova koristi jedinstvena neuronska mreža. Iako je osnovna konfiguracija umetača ostala gotovo nepromijenjena u odnosu na model A, u arhitekturu neuronske mreže za umetanje unesene su značajne promjene. Detaljna skica arhitekture ove neuronske mreže data je Slici 20.

Neuronska mreža za umetanje vodenih žigova, u okviru modela B, sastoji se od niza konvolucionih blokova. Jedan blok sadrži konvolucioni sloj, normalizaciju po seriji i aktivacionu funkciju. Blokovi nisu decimacioni, jer je vodeni žig skaliran na dimenzije signala nosioca, pa nema potrebe za smanjenjem njegovih dimenzija. U svakom sloju se vrši dopunjavanje nulama kako bi se na izlazu zadržale dimenzije dimenzije ulaza. Konvolucija u slojevima je jednodimenziona, jer je ulaz mreže dvokanalni vektor dužine 400. U prvom kanalu su 400 frekvencijskih koeficijenata za isječak signala nosioca dužine 25 ms. Drugi kanal sadrži 400 kopija bita vodenog žiga koji se ugrađuje.

Dizajn mreže umetača rezultat je usklađivanja vrijednosti nekoliko arhitekturalnih parametara. Vrijednosti ovih parametara su podešavane uzimajući u obzir vremenska i memorijska ograničenja. Arhitektura, odnosno memorijska složenost konvolucione neuronske mreže prevashodno zavisi od broja i veličine filtara u konvolucionim slojevima, kao i samog broja slojeva. Veličina uzorka za obučavanje ima uticaj na memorijske zahtjeve mreže, što se takođe odražava i na njenu arhitekturu koja mora biti adekvatne veličine, jer se procedura obučavanja mora sprovesti sa postojećim memorijskim kapacitetima.

Kao primarni parametar prilikom podešavanja arhitekture uzeta je veličina konvolucionog filtra, a vrijednosti ostalih parametara su zatim prilagođavane kako bi se postigla optimalna ravnoteža između performansi i resursa. U cilju pojednostavljenja procesa dizajna arhitekture i izbora vrijednosti parametara uzeto je da u svim konvolucionim slojevima veličina filtra bude ista. Razmatrane su veličine 21, 31, 41, koje približno odgovaraju veličini filtara u modelu A, koji su dimenzija 5×5. Broj

Tabela 3: Postavke arhitekture neuronske mreže umetača u okviru modela B.

Blok	Broj filtara	Veličina filtra/koraka	Veličina izlaza
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Konvolucioni	128	31 / 1	400×128
Izlazni	1	31 / 1	400×1

slojeva, odnosno konvolucionih blokova, biran je tako da se postepenim povećanjem receptivnog polja filtra, prema jednakosti (55), osigura da filtri u posljednjem konvolucionom sloju umetača obuhvataju čitav ulazni signal. Vodeći se ovom strategijom, s filtrima veličine 31 dobija se mreža sa 14 blokova, što je približno broju blokova u umetaču modela A. Mreža te veličine ispunjava hardverske kriterijume i dovoljna je za obavljanje zadataka umetača. Korišćenje filtara veće dimenzije, rezultuje nedovoljno dubokom arhitekturom, dok mreža s filtrima dužine 21 sadrži 20 blokova i prekoračuje memorijske okvire. Svakom konvolucionom sloju u mreži dodijeljena su po 128 filtra za ekstrakciju različitih karakteristika. Konvolucija je vršena s jediničnim korakom kako bi se očuvale ulazne dimenzije. ReLU aktivaciona funkcija iz modela A zamijenjena je modernijom varijantom, funkcijom *swish*. Za inicijalizaciju parametara odabrana je Hiova tehnika, zbog nesimetričnih aktivacionih funkcija u svim unutrašnjim blokovima. Na kraju mreže dodat je petnaesti konvolucioni sloj sa jednim filtrom i linearnom aktivacijom kako bi se izlaz umetača doveo na jednodimenzioni vektor od 400 vrijednosti kojima je reprezentovan votermarkovani isječak signala. Opisani detalji arhitekture neuronske mreže umetača predstavljeni su i Tabelom 3.

U arhitekturu mreže za umetanje uključene su preskačuće veze, koje se mogu

vidjeti na Slici 20. Ovo je urađeno kako bi se ubrzalo obučavanje, izbjegao problem nestajanja gradijenata i poboljšali rezultati mreže, prevashodno u pogledu očuvanja kvaliteta signala. Eksperimenti su pokazali da je, nakon dodavanja ovih veza u mrežu umetača, došlo do značajnog povećanja kvaliteta votermarkovanih signala, po obijema korišćenim mjerama kvaliteta od oko 20%.

6.1.2.2 Arhitektura neuronskh mreža detektora

Dizajnom modela B je predviđeno da diskriminator i ekstraktor vodenog žiga podijele dva zadatka koja je u modelu A obavljao detektor. Princip korišćenja konvolucionih neuronskih mreža za obavljanje ključnih koraka obrade je nastavljen, pa su obje komponente realizovane na taj način. S obzirom na to da su dva zadatka približno iste složenosti, ove dvije mreže imaju istu arhitekturu.

Zadatak diskriminatora je da odredi da li je u ulaznom audio signalu ugrađen bit vodenog žiga, odnosno klasifikacija signala na votermarkovane i nevotermarkovane. Ulaz ove mreže je vektor od 400 vrijednosti koji predstavlja amplitude frekvencijskih komponenti 25 ms audio signala. Mreža na izlazu daje samo jednu vrijednost P_w, koja predstavlja vjerovatnoću da je u ulaznom signalu ugrađen bit vodenog žiga. Ova vrijednost koristi se za donošenje odluke da li je signal votermarkovan, istim pravilom kojim je u modelu A vršena ekstrakcija bitova vodenog žiga:

$$v_t = \begin{cases} \top, & P_w > 0.5 \\ \bot, & P_w \leq 0.5, \end{cases} \tag{86}$$

gdje v_t predstavlja indikator prisustva bita vodenog žiga u posmatranom isječku audio signala.

Zadatak ekstraktora vodenog žiga je estimacija vrijednosti bita vodenog žiga ugrađenog u signal. Ulaz mreže ekstraktora je takođe vektor od 400 elemenata, odnosno 25 ms audio signala koji je diskriminator označio kao votermarkovan. Izlaz koji ova mreža treba da isporuči je vrijednost bita vodenog žiga v_b. U ovom modelu, bitovi vodenog žiga predstavljeni su vrijednostima -1 i 1. Ekstrahuju se na osnovu vrijednosti E_b na izlazu mreže za ekstrakciju, pomoću sljedećeg pravila:

$$v_b = \begin{cases} 1, & E_b > 0 \\ -1, & E_b \leq 0. \end{cases} \tag{87}$$

Dakle, obje mreže treba dizajnirati tako da se vektor sa 400 vrijednosti svede na jednu vrijednost. Prilikom dizajna arhitekture ovih mreža potrebno je ostvariti balans sa veličinom mreže za umetanje. Među komponentama umetača i detektora treba postići ravnotežu kako bi obje uspjele da ispune svoje zadatke. Stoga je

Tabela 4: Postavke arhitekture neuronske mreže diskriminatora i ekstraktora u okviru modela B.

Blok	Broj filtara	Veličina filtra/koraka	Veličina izlaza
Decimacioni	64	31 / 2	200×64
Decimacioni	64	31 / 2	100×64
Decimacioni	64	31 / 2	50×64
Decimacioni	64	31 / 2	25×64
Decimacioni	64	21 / 2	13×64
Decimacioni	64	11 / 2	7×64
Decimacioni	64	5 / 2	4×64
Decimacioni	64	3 / 2	2×64
Decimacioni	64	1 / 2	1×64
Izlazni	1	1 / 1	1×1

poželjno i da neuronske mreže koje im pripadaju budu približno jednake veličine. Na kraju su kreirane mreže sa po 9 konvolucionih blokova. Sastav blokova ostao je isti kao u mreži umetača. Svaki blok sadrži konvolucioni sloj, sloj normalizacije po seriji i aktivacionu funkciju. Promijenjene su postavke konvolucionih slojeva koji u ovim mrežama treba da vrše decimaciju. Decimacija je vršena sa najmanjom mogućom veličinom koraka $S = 2$ kako bi gubitak informacija između uzastopnih slojeva bio što manje izražen. U početna 4 bloka mreže uzeta je veličina filtra 31, koja je korištena i u slojevima umetača. Međutim, kako se smanjuju dimenzije ulaza u slojeve, neophodno je smanjivati i veličinu filtara. Filtri ne trebaju biti većih dimenzija od njihovog ulaza. Kada nakon primjene četvrtog bloka dimenzije ulaza postanu manje od 31, peti blok se formira sa filtrima veličine 21. U šestom bloku se koriste filtri veličine 11, u sedmom 5, u osmom 3 i veličine 1 u devetom bloku. Broj konvolucionih filtara u svim blokovima je smanjen u odnosu na mrežu za umetanje i iznosi 64. Ove mreže su s namjerom napravljene slabijima od mreže za umetanje, jer one obavljaju samo po jedan zadatak. Nasuprot tome, mreža za umetanje, pored rekonstrukcije originalnog signala, doprinosi i minimizaciji grešaka u detekciji. Zbog toga je važno da ova mreža bude moćnija, kako je mreže za detekciju ne bi nadvladale.

Na samom kraju mreža diskriminatora i ekstraktora dodat je jedan konvolucioni sloj sa jednim filtrom dimenzije 1 i jediničnim korakom, čija svrha je da generiše jednu izlaznu vrijednost. Tabela 4 sadrži podatke o arhitekturi neuronskih mreža u sklopu komponente detektora modela B. Neuronske mreže diskriminatora i ekstraktora bitova vodenog žiga razlikuju se samo u posljednjem sloju, preciznije, u

aktivacionoj funkciji izlaznog sloja. Mreža diskriminatora vrši klasifikaciju, pa je njoj na izlazu dodijeljena aktivaciona funkcija sigmoid. Mreža za ekstraciju bitova žiga se završava primjenom aktivacione funkcije hiperboličkog tangensa, čiji opseg vrijednosti odgovara načinu na koji su bitovi vodenog žiga kodirani u modelu B. Ovako dizajnirana arhitektura detektora, iako se sastoji od dvije mreže, zbirno sadrži značajno manji broj parametara od mreže za detekciju modela A, pa je efikasnost detekcije dodatna prednost modela B.

6.1.3 Slojevi za aproksimaciju napada

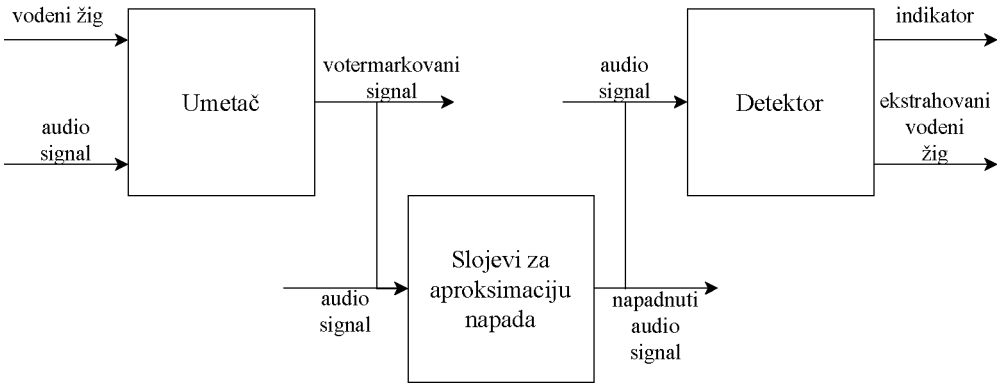
U sekcijama 2.3.1 i 2.3.2 više puta je spomenuto da postoji značajan broj efekata i napada kojima podliježu audio signali i sistemi vodenog žiga. Iz perspektive sistema vodenog žiga sve takve operacije, bile one maliciozne ili ne, smatraju se napadima jer ugrožavaju njegovo funkcionisanje. Stoga, u nastavku teksta, termin napadi obuhvata sve efekte i akcije kojima se može spriječiti izvršavanje nekog od zadataka sistema vodenog žiga.

U brojnim preglednim radovima [147–152] klasifikovani su i nabrojani neki od napada. Međutim, nijednom studijom nisu se mogle obuhvatiti sve moguće vrste napada koje su postojale u datom trenutku. Pored toga, u međuvremenu su dizajnirani i novi napadi [153, 154].

Nijedan sistem vodenog žiga ne može biti otporan na sve napade, jer je broj potencijalnih napada praktično neograničen. Stoga je u cilju postizanja što cjelovitije robustnosti potrebno napade svrstati u grupe koje čine napadi sa sličnim karakteristikama. Nakon toga, odabirom i tretiranjem predstavnika svake od grupa napada, povećavaju se izgledi da sistem bude otporan na napad koji se u toku njegovog dizajniranja, odnosno obučavanja, nije desio.

U ovom radu su analizirane su različite tehnike obrade signala koje mogu predstavljati prijetnju funkcionisanju sistema vodenog žiga. Razmatrani su efekti za čiju realizaciju je neophodan samo jedan primjerak audio signala koji treba napasti. Drugi efekti koje inteligentni napadač može iskoristiti, ako posjeduje dodatno znanje o vodenom žigu, načinu njegovog umetanja ili detekcije, kako bi ugrozio robustnost i sigurnost vodenog žiga nisu uzimani u obzir.

Iako se napadi dešavaju nakon umetanja vodenog žiga u signal, procedura detekcije ne može biti jedina odgovorna za njihovo neutralisanje. Procedura umetanja takođe mora biti na adekvatan način prilagođena napadima kako vodeni žig ne bi bio uništen ukoliko se neki od napada desi. Na primjer, ukoliko umetač bitove vodenog žiga ugrađuje isključivo u jednom opsegu frekvencija, nakon filtriranja votermarko-



Slika 21: Skica sistema vodenog žiga sa slojevima za aproksimaciju napada.

vanog signala odgovarajućim filtrom vodeni žig će biti obrisan i detektor ga ni na koji način ne bi mogao rekonstruisati. Takođe, ukoliko umetač vodeni žig smješta samo u jednom intervalu signala, brisanjem tog intervala nepovratno bi se izgubio i vodeni žig.

U tradicionalnim pristupima naučnici istražuju napade od interesa kako bi prepoznali njihove karakteristike i osmislili šeme umetanja i detekcije otporne na posmatrane napade. Upotreba neuronskih mreža olakšava i ovaj aspekt dizajniranja sistema vodenog žiga, jer se osobine napada izdvajaju procesom obučavanja i susretanjem mreža sa njihovim različitim instancama. Ovaj način dizajniranja votermarking sistema zahtijeva da napadi budu realizovani kao slojevi neuronske mreže, odnosno kao diferencijabilne funkcije. To je potrebno kako bi se obezbijedio protok gradijenata od mreže detektora ka mreži umetača, tj. kako bi se informacije o greškama u detekciji izazvane napadima mogle prenijeti komponenti za umetanje. Uslijed zahtjeva za diferencijabilnošću, neki napadi se ne mogu u potpunosti autentično simulirati, već se realizuju njihove aproksimacije. Aproksimacije treba da budu što preciznije kako bi se sistem bolje pripremio na pojavu realnih napada.

U ovoj disertaciji izdvojene su četiri kategorije napada: šumovi, skaliranje amplitude, filtriranje i efekti desinhronizacije. Ovim kategorijama obuhvaćene su ključne vrste potencijalnih napada na sisteme vodenog žiga. Izabrani su predstavnici sve četiri kategorije, koji su uključeni u modele sistema, između komponenti umetača i detektora, kao što je prikazano na Slici 21. U model A su integrisani predstavnici svih kategorija, dok su iz modela B odstranjeni efekti desinhronizacije. Mjere kojima se procjenjuje koliko su uspješno odabrani napadi suzbijeni uključene su u analizu performansi sistema. Svi napadi su parametrizovani, a vrijednosti parametara su podešene tako da napadi ne utiču previše na perceptivni kvalitet signala. Napadi gube smisao ako su toliko razorni da je očigledno da je signal izmijenjen operacijama obrade zvuka.

6.1.3.1 Ustaljeni audio efekti

Promjene u signalu nekada mogu nastati njegovim sasvim regularnim procesiranjem što za posljedicu ne bi smjelo imati greške u detekciji vodenih žigova. Stoga je otpornost na ova procesiranja neophodna gotovo svim sistemima vodenog žiga namijenjenim praktičnoj primjeni. U ovoj disertaciji su izdvojena tri najčešća tipa ovih efekata, zajedno nazvanih ustaljenim audio efektima. Slojevi koji ih aproksimiraju dodati su u arhitekturu predloženih modela.

Šumovi su jedan od najčešćih tipova efekata koji se mogu pojaviti u različitim vrstama signala, uključujući i audio signale. Predstavljaju bilo kakve neželjene i slučajne varijacije, odnosno smetnje, kojima se degradira signal. Šumovi su ujedno i najopštija vrsta napada na sisteme vodenog žiga, jer se izmjene koje u signal unosi primjena i nekih drugih operacija mogu, barem djelimično, modelovati kao šum. Stoga otpornost modela dubokog učenja na šumove u signalu doprinosi njegovoj sposobnosti da generalizuje, odnosno ostvari određeni nivo otpornosti i na nepoznate efekte. Ova karakteristika šumova pruža još jedan razlog za njihovo uključivanje u proceduru obučavanja

Skaliranje amplitude je veoma čest, ali ujedno i veoma jednostavan efekat koji ne narušava relativni odnos odbiraka audio signala. Većina sistema vodenog žiga dizajnirana je tako da im se šema za detekcija ne oslanja na apsolutne vrijednosti odbiraka signala, već na međusobne odnose odbiraka. To ove sisteme čini otpornim na skaliranje amplitude. Uvođenjem sloja koji simulira ovu vrstu napada u arhitekturu neuronskih mreža, osigurava se i da rezultujuće šeme za umetanje i detekciju našeg sistema posjeduju ove poželjne karakteristike.

Digitalni filtri su često korišteno sredstvo u obradi signala. Filtri se realizuju tako da istaknu poželjne, a ublaže ili eliminišu loše karakteristike signala. Za audio votermarking sisteme od najvećeg interesovanja su niskopropusni filtri koji propuštaju niskofrekventne, a guše visokfrekventne komponente signala. Otpornost na primjenu ovih filtara je od presudnog značaja za sve sisteme vodenog žiga. U suprotnom se vodeni žigovi koje oni ugrađuju mogu veoma lako obrisati, bez ugrožavanja perceptivnog kvaliteta signala nosioca.

6.1.3.2 Efekti desinhronizacije

Efekti desinhronizacije su najefikasnija vrsta napada na sve sisteme vodenog žiga. Karakteriše ih to što remete poravnanje signala nosioca i vodenog žiga, smanjenjem ili povećanjem broja odbiraka signala ili njihovom permutacijom. Na taj način mijenja se i relativni poredak i lokacija odbiraka u kojima su ugrađeni bitovi

vodenog žiga, što otežava njegovu detekciju. Zbog svoje efikasnosti, ovi napadi su u fokusu mnogih istraživanja [4, 5, 7, 15] u kojima su razvijane tehnike za njihovo suzbijanje. Ove tehnike zasnivaju se na ekstrakciji karakteristika audio signala koje su u izvjesnoj mjeri otporne na ovu vrstu efekata. Bitovi vodenog žiga ugrađuju se u vrijednostima ekstrahovanih koeficijenata za koje se očekuje da neće biti značajno izmijenjeni desihnronizujućim efektima. Međutim, efekti desinhronizacije zavise od vrijednosti više parametara. Njihova primjena rezultuje raznolikim promjenama u audio signalima. Zbog toga se ne može razviti tehnika sa rigidnim skupom pravila kojom se u potpunosti mogu prevazići sve varijacije signala koje ovi efekti mogu proizvesti.

Neuronske mreže mogu modelovati zavisnosti proizvoljne složenosti. Stoga je pretpostavka da se, uz odgovarajuću proceduru obučavanja, mogu modelovati karakteristike ulaza otporne na desinhronizaciju u okviru mreže za umetanje. Takođe, mreža za detekciju može modelovati inverze pojedinih efekata desinhronizacije, estimacijom vrijednosti njihovih parametara. Shodno tome, neuronske mreže mogu se koristiti za neutralisanje ovih napada.

Nekoliko predstavnika grupe efekata desinhronizacije dodato je u slojeve za aproksimaciju napada modela A. Njihova primjena uslovila je i korišćenje dužih segmenata audio signala u ovom modelu, u trajanju od 2 sekunde. Korišćenje kraćih intervala može dovesti do toga da se prilikom desinhronizacije čitav interval izmjesti na drugu poziciju, pa nijedan odbirak originalnog signala ne bi bio dostupan prilikom detekcije. Zbog toga je za analizu i obradu efekata desinhronizacije potrebno koristiti što duže audio segmente, poželjno i kompletan signal. Memorijski kapaciteti onemogućili su korišćenje segmenta dužih od 2 sekunde.

Jedan predstavnik nije dovoljan da se na cjelovit način obuhvate svi efekti desinhronizacije, s obzirom na njihovu heterogenost. U ovoj disertaciji razmatrano je pet tipova ovih efekata: brisanje odbiraka, permutacija odbiraka, pomjeranje u vremenu, ponovno uzorkovanje i skaliranje vremena.

U Prilogu D.4 su pobrojani svi efekti desinhronizacije razmatrani u ovoj studiji. Objašnjeni su postupci za njihovu simulaciju i definisani parametri kojima se kontroliše intenzitet, odnosno ozbiljnost ovih napada.

Simulacije su parametrizovane, a vrijednosti parametara su izabrane tako da se postigne balans između razornosti efekta i očuvanja informacija koje audio signal nosi.

6.2 Procedura obučavanja

Procedura obučavanja ovog sistema mora biti pažljivo sprovedena zbog, u osnovi, različitih i donekle suprotstavljenih zadataka mreža za umetanje i detekciju. Cilj mreže za umetanje je rekonstrukcija originalnog signala nakon dodavanja vodenog žiga. Potpuno brisanje vodenog žiga bio bi najbolji korak ka ispunjenju ovog zadatka. Nasuprot tome, detektor bi u ovom slučaju bio u potpunosti onemogućen da izvršava svoje zadatke, jer na ulazu očekuje signal u kojem se nalazi vodeni žig. Neuspjehom komponente detektora bi čitav sistem vodenog žiga postao neupotrebljiv. Stoga je u procesu obučavanja kompletnog sistema potrebno donekle sputavati mrežu za umetanje kako bi se spriječilo brisanje vodenog žiga. Ograničavanje treba da bude kontrolisano kako mreža za detekciju ne bi prevladala i tako spriječila umetač da izvrši kvalitetnu rekonstrukciju, zadatak gotovo jednako važan kao i detekcija vodenog žiga. Dakle, cilj procedure obučavanja je da pronađe optimalni balans u kojem sve neuronske mreže u sistemu ispunjavaju svoje zadatke.

Sve neuronske mreže koje čine ovaj sistem moraju se obučavati istovremeno. Odvojeno obučavanje neuronskih mreža bilo bi izuzetno komplikovano. U tako dizajniranoj proceduri obučavanja bi teško bilo procijeniti kada zaustaviti obučavanje određene neuronske mreže tako da ne ugrozi ostale, a da postiže zadovoljavajuće performanse na sopstvenom zadatku. Pored toga, ranije je naglašeno da je za ostvarivanje robustnosti neophodno učešće svih komponenti sistema. Iz tog razloga se parametri svih mreža u sistemu istovremeno ažuriraju, kako bi se procedure umetanja i detekcije prilagodile napadima.

Kako se sistem sastoji od više neuronskih mreža, namijenjenih za različite zadatake, tim neuronskim mrežama dodijeljene su različite funkcije gubitka. Neuronske mreže koje vrše regresiju minimizuju srednju kvadratnu ili apsolutnu grešku, dok je za klasifikaciju korišćena unakrsna entropija iz jednakosti (63). Ovim se dodatno komplikuje procedura obučavanja koju je znatno teže održavati stabilnom u odnosu na obučavanje sistema sa jednom neuronskom mrežom.

Funkcije gubitka primjenjivane su u osnovnom obliku i nisu proširivane izrazima za regularizaciju s obzirom na to da je regularizacija izvršena normalizacijom po seriji koja je sastavni dio svih mreža u sistemu. Dodatno je dio korpusa podataka izdvojen za validaciju koja je vršena nakon svakog ciklusa, kako bi se pratile performanse sistema i detektovalo eventualno preprilagođavanje podacima za obučavanje. Sistem je obučavan optimizacionim algoritmom Nadam, jer su u njemu objedinjeni svi značajni aspekti ostalih tehnika optimizacije. Stopa obučavanja postavljena je na 0.0001.

Trajanje procedure obučavanja obično se izražava u epohama. Jedna epoha predstavlja prolazak čitavim skupom za obučavanje. Međutim, u ovom radu, procedura obučavanja podijeljena je na cikluse. Termin „ciklus" uveden je da bi predstavio manji broj iteracija procedure za obučavanje. Ovo je urađeno kako bi se olakšalo razumijevanje faza u proceduri obučavanja, jer su, zbog obima podataka, potrebne intervencije unutar jedne epohe. Dužina ciklusa utvrđena je eksperimentalnim putem.

Na samom početku procedure obučavanja, sistem nije u mogućnosti da vrši niti umetanje niti detekciju vodenih žigova. Stoga je prvih nekoliko ciklusa posvećeno obučavanju sistema za obavljanje osnovnih zadataka, čime se postiže dobra polazna tačka za razvijanje robustnosti, uvođenjem napada. Udio nenapadnutih serija zadržan je i u narednim ciklusima, čime se sprečava potpuno prilagođavanje sistema signalima koji su pod dejstvom napada i zanemarivanje osnovnog scenarija. U obučavanju mreža za detekciju uporedo su korišćeni i votermarkovani signali, kao i njihovi nevotermarkovani ekvivalenti. Napadi su primijenjivani na nevotermarkovanim signalima sa istom stopom kao i na votermarkovanim, kako njihova kasnija pojava ne bi izazvala probleme poput neautorizovane detekcije.

Takođe, nad napadnutim signalima je primjenjivan niskopropusni filtar, kako bi se obezbijedio dodatni nivo robustnosti. Većina audio efekata ekvivalentno tretira sve opsege frekvencija. Ugrađivanje bitova vodenog žiga u visokofrekventnim komponentama signala olakšava proces detekcije kada je signal pod dejstvom nekog od ovih efekata. Na tim frekvencijama nema mnogo audio sadržaja, pogotovo ljudskog govora, pa je lakše dodati nove informacije i kasnije ih izvući. Dodatno, izmjene visokofrekventnih komponenti signala manje utiču na ljudski sluh i utisak o kvalitetu audio zapisa, što ih čini još podesnijim za umetanje. Međutim, ovim načinom umetanja se osigurava robustnost isključivo na nezavisnu primjenu audio efekata. Sistem bi se mogao obučiti da u visokim frekvencijama unosi bitove vodenog žiga potrebne za detekciju u situacijama kada se desi neki od efekata koji jednako mijenjaju sve frekvencije. Time bi votermarking signala, koji nakon primjene ovih efekata prolaze kroz niskopropusni filtar, bio neizvodljiv. Sparivanjem svih realizovanih efekata sa niskopropusnim filtriranjem prilikom obučavanja sistem se primorava da koristi niže frekvencije za ugradnju vodenog žiga, čime se rješava prethodno opisani problem.

Prethodno opisani koraci i karakteristike procedure obučavanja su zajedničke za oba predložena modela. Međutim, procedure obučavanja za model A i model B se i razlikuju u određenim detaljima koji su iznijeti u nastavku.

6.2.1 Procedura obučavanja modela A

Detektor modela A definisan je kao klasifikator, stoga je očigledan izbor za funkciju gubitka mreže za detekciju unakrsna entropija. Preciznije, kao funkcija gubitka korišćena je prosječna binarna unakrsna entropija (*engl. average binary cross-entropy* - ABCE):

$$J_{\text{ABCE}}(\Theta) = \frac{1}{ML_w} \sum_{m=1}^{M} \sum_{k=1}^{L_w} \log\left(P_k^{(m)}(\Theta)\right) v^{(m)}(k) + \log\left(1 - P_k^{(m)}(\Theta)\right)\left(1 - v^{(m)}(k)\right), \tag{88}$$

gdje je $P_k^{(m)}(\Theta)$ izlaz detektora za k-ti bit vodenog žiga u m-tom isječku signala iz date serije, odnosno vjerovatnoća da je vrijednost tog bita jednaka 1. Vrijednost $v^{(m)}(k)$ je vrijednost k-tog bita vodenog žiga ugrađenog u taj isječak.

Pri konstrukciji ove funkcije gubitka, detekcija svakog bita vodenog žiga posmatra se kao zaseban zadatak binarne klasifikacije. Gubitak po jednom bitu može se izračunati unakrsnom entropijom iz jednakosti (63), gdje je $N_c = 2$. Kako u ovom radu nisu definisane prioritetne grupe bitova za ekstrakciju, gubici po pojedinačnim bitovima su uprosječeni kako bi se dobila procjena ukupne greške detektora.

U izboru funkcije gubitka za očuvanje kvaliteta razmatrane su srednja kvadratna i srednja apsolutna greška. Korišćenje srednje kvadratne greške dovodi do superiornijih performansi iz nekoliko razloga. Kako je grafik srednje kvadratne greške glađi od grafika srednje apsolutne greške, manje su šanse da se procedura obučavanja zaglavi u dijelu parametarskog prostora koji ima oblik grebena. Takođe, srednja kvadratna greška uključuje izraz koji predstavlja kvadratnu razliku između dva upoređivana signala koji se nalazi i u SNR mjeri očuvanja kvaliteta signala. Stoga je očekivano da se optimalna vrijednost SNR mjere dostigne korišćenjem funkcije gubitka koja joj je sličnija. Jedini nedostatak srednje kvadratne greške u odnosu na srednju apsolutnu grešku je osjetljivost na izuzetke (*engl. outliers*). Međutim u korpusu podataka koji je korišten u ovom radu nema ovakvih primjeraka, čime je taj nedostatak otklonjen.

Funkcije gubitka za očuvanje kvaliteta signala i detekciju su objedinjene i uvedeni su težinski koeficijenti kako bi se postigao optimalni balans u ispunjenju zahtjeva koji su postavljeni pred sistemom. Očuvanje kvaliteta mora se održavati na određenom nivou ispod optimalnog, kako bi se postigla robustnost i obratno. Koeficijenti w_Q i w_D pridruženi su funkcijama gubitka za kvalitet J_Q i detekciju J_D, respektivno.

$$J(\Theta) = w_Q J_Q(\Theta) + w_D J_D(\Theta). \tag{89}$$

Eksperimentalnim putem je utvrđeno da mreža za detekciju, zbog svoje veličine, ima tendenciju da prevlada i obuči se za savršenu rekonstrukciju, čime se onemo-

gućava uspješna ekstrakcija bitova vodenog žiga. Željeni balans se postiže kada su težinski koeficijenti u razmjeri 3 : 1 na strani umetača.

Obučavanje modela A trajalo je 60 ciklusa. U jednom ciklusu obrađuje se 200 serija podataka za obučavanje. Vodeni žigovi iz predefinisanog skupa ugrađivani su u 40% serija, 40% serija označeno je slučajno generisanim vodenim žigom izvan skupa kako bi se sistem obučio da bude otporan na neautorizovano umetanje. Preostalih 20% serija nije označeno vodenom žigom. Prilikom obrade nevotermarkovanih isječaka, ažurirani su isključivo parametri detektora.

Istovremeno uvođenje svih napada u okviru procedure obučavanja pokazalo se previše zahtjevnim za sistem neuronskih mreža, koji u takvom scenariju nije mogao konvergirati. Stoga su napadi uvođeni postepeno, kao podzadaci. Ovakav dizajn procedure baziran je na ideji transfera učenja iz Sekcije 5.4.3. U okviru jednog ciklusa rješava se jedan novi podzadatak, odnosno napad. Model dobijen na kraju ciklusa koristi se kao polazna tačka za naredni ciklus, odnosno podzadatak.

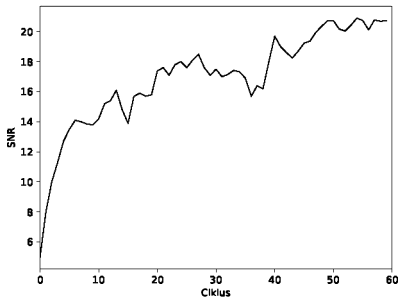
Naša ideja se u izvjesnoj mjeri razlikuje od transfernog učenja u tradicionalnom smislu. Koristimo isti skup podataka za sve podzadatke i ne vršimo promjene u arhitekturi mreža tokom obučavanja. Takođe, obuka za prethodne podzadatke se ne prekida kada se započne novi podzadatak. Nastavlja se, ali u smanjenom obimu.

Dužina ciklusa utvrđena je na osnovu broja serija potrebnih sistemu da riješi najsloženiji podzadatak, odnosno postigne otpornost na najkompleksniji napad. Prva dva ciklusa sprovode se bez primjene napada nad izlazom umetača, kako bi se sistem obučio da obavlja osnovne zadatke. Nakon toga se novi napad uvodi na svaka dva ciklusa. U prvom ciklusu nakon uvođenja novog napada, sistem uči njegove karakteristike. Napad se primjenjuje na odabrane serije, bez ikakvih dopunskih operacija. Počevši od drugog ciklusa po predstavljanju novog napada sistemu, pa nadalje, napad je uparen sa niskopropusnim filtriranjem. Ovo je neophodno kako bi se spriječilo da umetač bitove vodenog žiga dodaje na visokim frekvencijama.

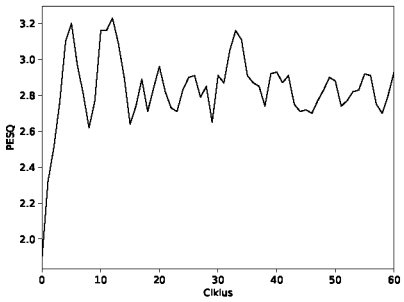
Za svaki napad data je vjerovatnoća njegove primjene nad serijom podataka u jednom ciklusu:

$$P(i_c, i_a) = \begin{cases} 0, & i_c \leq 2 \\ P_a P_{a_0}, & 2i_a + 1 \leq i_c \leq 2i_a + 2 \\ P_a(1 - P_{a_0}) / \lfloor (i_c - 1)/2 \rfloor, & i_c > 2i_a + 2 \ \wedge \ i_c \leq 2N_a + 2 \\ P_a / N_a, & i_c > 2N_a + 2, \end{cases} \tag{90}$$

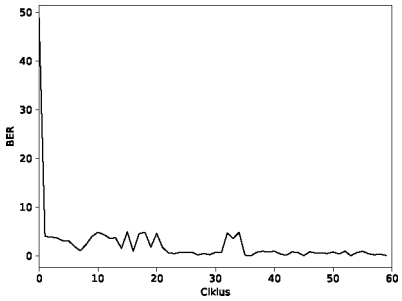
gdje i_a predstavlja redni broj napada ($i_a \in \{1, 2, ..., N_a\}$), a i_c je redni broj ciklusa ($i_c \in \{1, 2, ..., 60\}$).



(a) SNR



(b) PESQ



(c) BER

Slika 22: Trajektorije mjera performansi tokom obučavanja modela A.

Hiperarametar N_a označava ukupan broj napada, dok je vrijednost P_a vjerovatnoća primjene bilo kojeg napada nad serijom podataka. U radu [136] udio napadnutih i nenapadnutih serija je identičan. Nakon uvođenja efekata desinhronizacije u radu [137] udio nenapadnutih serija u ciklusu smanjen je na 25%, jer se tretira znatno veći broj napada $N_a = 9$. Slijedi da je $P_a = 0.75$.

Vjerovatnoće pojave pojedinačnih napada u svakom ciklusu su različito raspodijeljene. Novouvedenom napadu se daje prioritet, tako što mu se dodjeljuje vjerovatnoća pojave $P_{a_0} = 0.5$. Svi prethodno uvedeni napadi moraju ostati prisutni kako bi se osiguralo da naučene tehnike za postizanje robustnosti ne budu isključene uslijed pojave novih napada. Vjerovatnoće pojavljivanja ranije uvedenih napada su ravnomjerno raspodijeljene, kako bi se među njima očuvala ravnoteža.

Nakon što se posljednji napad uvede u ciklusima 19 i 20, sistem se obučava dodatnih 40 ciklusa, pri čemu je svakom napadu dodijeljena ista vjerovatnoća. Obučavanje je zaustavljeno nakon što 10 ciklusa nije bilo poboljšanja neke od mjera za očuvanje kvaliteta audio signala. Detekciona statistika je već nakon prvog ciklusa dosegnula prilično veliku vrijednost, koja je održavana do kraja procedure obučavanja, uz određene oscilacije uglavnom izazvane uvođenjem novih napada, što se može vidjeti na Slici 22.

U ovom radu, robustnost, tj. detekciona statistika, ima prioritet nad očuvanjem signala, pa je u kriterijum za izbor najbolje verzije obučenog modela uključen prag za procenat greški u detekciji od 5%. Verzije modela sa procentom grešaka detektora iznad ovog praga ne smatraju se podesnim, bez obzira na njihove performanse u pogledu očuvanja kvaliteta signala. Ukoliko je procenat grešaka u detekciji ispod definisanog praga, uspješnost različitih modela poredi se mjerama za očuvanje kvaliteta i od njih se, za testiranje, bira najbolji.

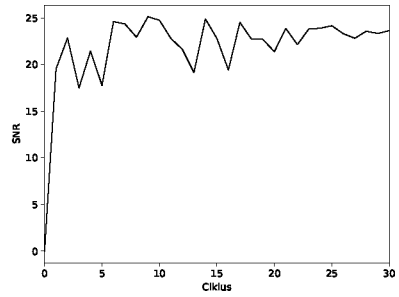
Na Slici 22 su prikazane trajektorije mjera performansi tokom obučavanja modela A, računate na posebnom validacionom skupu. Kod mjera kvaliteta primjećuje se stabilniji rast SNR mjere u odnosu na PESQ, izazvan ranije pomenutom sličnošću te mjere i srednje kvadratne greške koju ova procedura obučavanja minimizuje. Takođe, primjetan je i nesklad u vrijednostima ovih dviju mjera. Ovaj nesklad ogleda se u tome da porast vrijednosti SNR mjere ne implicira porast PESQ mjere i obratno.

Slika 22c prikazuje prosječan procenat pogrešno detektovanih bitova vodenog žiga u signalima iz validacionog skupa. Eksperimenti za izračunavanje ovih vrijednosti izvedeni su i sa napadima uvedenim do tog ciklusa procedure obučavanja. Napadnuti i nenapadnuti signali su na slučajan način uzorkovani, raspodjelom iz (90). Detaljnija diskusija rezultata, ostvarenih modelom A, data je u Poglavlju 8.

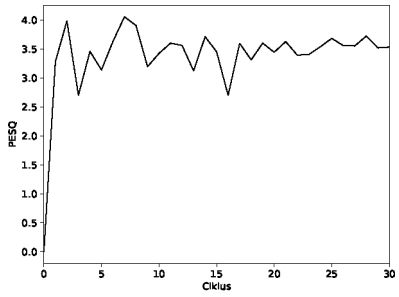
6.2.2 Procedura obučavanja modela B

Modelovanjem cjelokupnog procesa detekcije vodenog žiga kao zadatka klasifikacije sistem je usmjeren na razlikovanje ugrađivanih vodenih žigova umjesto nezavisne procjene vrijednosti njihovih bitova. Na taj način se stvara zavisnost u ekstrakciji bitova vodenog žiga, što za posljedicu može imati smanjenje kapaciteta prisutno u modelu A. Modelom B se proces detekcije razdvaja na dva dijela. Razlikovanje votermarkovanih i nevotermarkovanih signala je prirodno definisano kao klasifikacioni zadatak koji izvršava komponenta diskriminatora, dok se u komponenti za ekstrakciju, umjesto razlikovanja vodenih žigova, vrši procjena vrijednosti njihovih pojedinačnih bitova regresionim modelom. Na taj način se ekstrakcija jednog bita izoluje od ekstrakcije ostalih bitova istog ili drugih vodenih žigova.

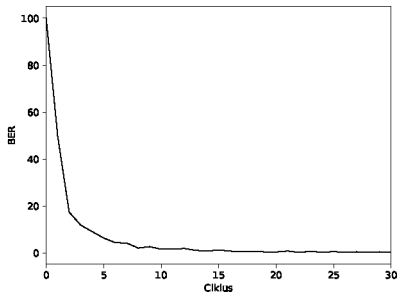
Vrijednosti bitova vodenog žiga kodirane su sa -1 i 1. To je omogućilo primjenu hiperboličkog tangensa, umjesto sigmoida, kao aktivacione funkcije na izlazu ekstraktora. Ova odluka donosi nekoliko prednosti u proceduri obučavanja, koje su detaljno opisane u Sekciji 5.3.3. Dodatno, korišćenjem ovakvog načina kodiranja, reprezentacije signala nosioca i vodenog žiga su svedene na isti opseg, što značajno olakšava inicijalizaciju i obučavanje sistema neuronskih mreža.



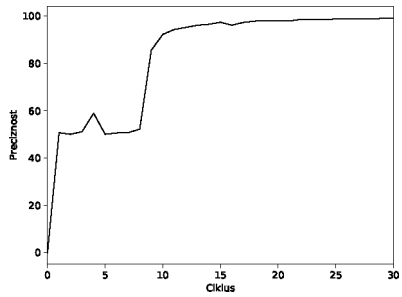
(a) SNR



(b) PESQ



(c) BER



(d) Preciznost diskriminatora

Slika 23: Trajektorije mjera performansi tokom obučavanja modela B.

Neuronskim mrežama su, prema definisanim zadacima, dodijeljene funkcije gubitka. Cilj obučavanja mreže diskriminatora je razlikovanje raspodjela votermarkovanih i nevotermarkovanih signala. U skladu sa tim je za funkciju gubitka odabrana binarna unakrsna entropija. Neuronske mreže za umetanje i ekstrakciju bitova vodenog žiga minimizuju srednju kvadratnu grešku.

Tokom obučavanja modela B pokazala pokazalo se da je neophodno dodatno uvećati stopu učenja mreže za umetanje, povećavanjem pridruženog težinskog koeficijenta. Ukupni gubitak sistema izračunava se analogno modelu A, kao ponderisana suma gubitaka umetača J_Q, diskriminatora J_D i ekstraktora J_E:

$$J(\Theta) = w_Q J_Q(\Theta) + w_D J_D(\Theta) + w_E J_E(\Theta). \tag{91}$$

Rezultati modela B, prijavljeni u Poglavlju 8, ostvareni su uz težinske koeficijente u razmjeri $100 : 1 : 1$.

Model B je zbog manjeg broja napada obučavan ukupno 30 ciklusa. Dužina jednog ciklusa je takođe prepolovljena na 100 serija zbog odsustva efekata desinhronizacije. Po uzoru na model A, prva dva ciklusa procedure obučavanja sprovedena su bez primjene napada nad signalima. Počevši od trećeg ciklusa, 75% serija je izloženo aditivnom šumu, skaliranju amplitude i niskopropusnom filtriranju, uključujući i

njihove različite kombinacije. Sve vrste napada uvedene su istovremeno, s jednakim vjerovatnoćama primjene, koje su zadržane do kraja procedure obučavanja. Neuronskoj mreži diskriminatora se prosljeđuju dvostruko veće serije podataka koje sadrže i votermarkovane i nevotermarkovane oblike isječaka audio signala.

Kriterijum za selekciju najbolje verzije modela zahtijeva postizanje preciznosti diskriminatora od najmanje 90%, stope grešaka u ekstrakciji bitova vodenog žiga ispod 5%, uz istovremeno održavanje maksimalnog kvaliteta signala. Model koji je dostigao zadate pragove u detekciji i ostvario najbolje performanse u očuvanju kvaliteta signala, odabran je za testiranje u Poglavlju 8.

Slika 23 prikazuje grafike kretanja vrijednosti mjera performansi, izračunatih na validacionom skupu prilikom obučavanja modela B. Kao i kod modela A, performanse komponente za umetanje procjenjivane su mjerama SNR (Slika 23a) i PESQ (Slika 23b). Performanse diskriminatora i ekstraktora prate se mjerenjem preciznosti prepoznavanja votermarkovanih i nevotermarkovanih signala, odnosno udjela tačno ekstrahovanih bitova ugrađenog vodenog žiga.

Sa date slike se primjećuje da je model u nekim instancama postizao i veći stepen očuvanja kvaliteta signala. PESQ mjera prelazila je 4.0, a SNR 25 dB. Međutim, eliminatorni kriterijumi u pogledu stopi grešaka diskriminatora i ekstraktora nisu bili u potpunosti zadovoljeni, pa iz tog razloga ove verzije modela nisu uzete za dalju evaluaciju.

7 Korpus podataka

Korpus podataka je od presudnog značaja za ishod obučavanja modela mašinskog učenja, odnosno njegov kvalitet. Ukoliko su obezbijeđeni kvalitetno obrađeni i sveobuhvatni podaci, za očekivati je da algoritam mašinskog učenja postigne bolje performanse.

Korpusi podataka su dokazali svoju vrijednost u svim oblastima u kojima je primjenjivano duboko učenje. Zahvaljujući brzom razvoju interneta, obimni skupovi podataka postali su javno dostupni. Međutim, mnoge grupe istraživača utrošile su značajno vrijeme i resurse kako bi se ti podaci na prikladan način obradili i uspješno primijenili u dubokom učenju. Ovi enormni, relevantni i javno dostupni korpusi podataka nisu bili od značaja samo u rješavanju konkretnih zadataka u određenim oblastima, već i uopšteno za razvoj i nastanak koncepata dubokog učenja. Saznanja dobijena na ovim primjerima prenošena su i primjenjivana na probleme u drugim oblastima. U oblasti kompjuterske vizije je pojava korpusa kao što su CIFAR [155] i ImageNet [156] direktno uticala na stvaranje najpreciznijih algoritama za klasifikaciju slika. Rad na ovim korpusima uzrokovao je ogromne iskorake u razvoju konvolucionih neuronskih mreža. Penn Treebank korpus [157] bio je jedna od prekretnica u razvoju modela za obradu prirodnog jezika, odnosno rekurentnih neuronskih mreža. Postoji nekoliko dobro poznatih skupova podataka koji su bili krucijalni u unapređenju istraživanja u oblasti obrade zvuka. Korpusi UrbanSound [158] i ESC [159] sadrže zvuke iz različitih okruženja i koriste se u obučavanju modela za prepoznavanje zvukova. Korpusi poput TIMIT [160] i LibriSpeech [161] doprinijeli su razvoju algoritama za sintezu govora i automatsku transkripciju.

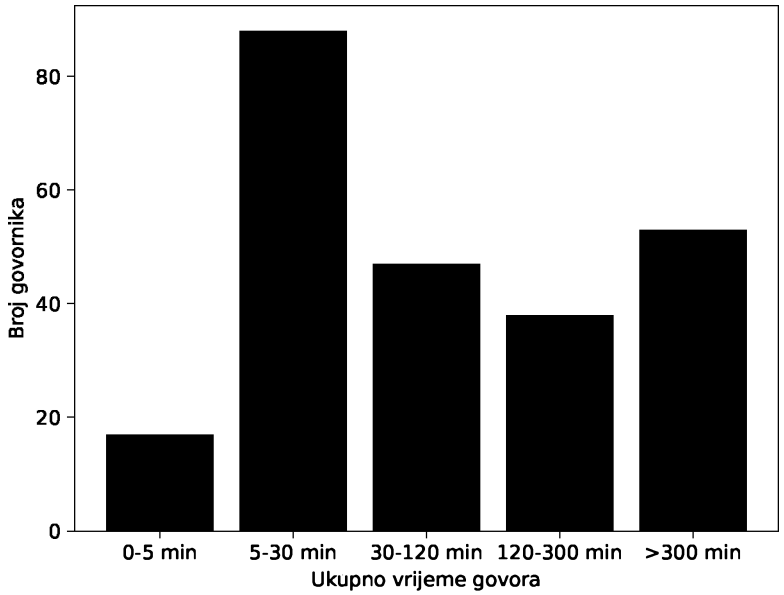
Standardizovan, referentni skup podataka za evaluaciju audio votermarking sistema još uvijek nije usvojen. Autori u svojim radovima koriste različite skupove audio snimaka, izabrane po spostvenom nahođenju, na kojima testiraju i upoređuju performanse votermarking tehnika. Postojanjem jednog sadržajnog, javno dostupnog korpusa podataka obezbjeđuju se identični uslovi za testiranje različitih votermarking i drugih odgovarajućih tehnika u digitalnoj obradi signala, što je jedan od osnovnih preduslova za njihovo pravedno poređenje. Poređenja sprovedena na ovakvom korpusu bila bi daleko temeljnija i mjerodavnija od ustaljenih. Trenutno se u literaturi upotrebljavaju skromni skupovi od svega desetak zapisa, poput SQAM fajlova [162] koji se koriste u [88]. U radu [16] autori koriste mali dio TIMIT baze za evaluaciju svoje tehnike. Međutim čak i kompletan ovaj korpus nije dovoljno obiman da bi mogao biti primijenjen u dubokom učenju jer se sastoji od ukupno 6300 kratkih snimaka u kojima 630 govornika izgovara istih 10 rečenica. Stvaranjem opsežnog kor-

pusa otvara se prostor za dublje istraživanje potencijala primjene tehnika dubokog učenja u sistemima vodenog žiga. Istraživačkim grupama se na taj način omogućava da dizajniraju sopstvene arhitekture i procedure obučavanja za komponente sistema vodenog žiga što će na kraju dovesti do naprednijih i robustnijih modela.

Za potrebe ovog istraživanja, s nastojanjem da se u doglednom periodu učini javno dostupnim, prikupljen je i obrađen korpus snimaka govora poslanika i brojnih gostiju u Skupštini Crne Gore za vrijeme 26. saziva ovog zakonodavnog tijela. Obučavanjem sistema vodenog žiga na ovakvom skupu podataka stavlja se akcenat na zaštitu izjava političara i drugih javnih ličnosti kroz potvrđivanje njihove autentičnosti. Zaštita autentičnosti govora je od izuzetne važnosti za društvo, jer on predstavlja preovlađujući način izražavanja mišljenja i stavova, što je jedno od osnovnih ljudskih prava i temelja demokratije.

Odjeljenje za snimanje i emitovanje pri Skupštini Crne Gore dužno je da obavlja snimanje i arhiviranje audio i video sadržaja sa sjednica, radnih tijela, konferencija i drugih vrsta sastanaka u Skupštini. U našem korpusu nalaze se zapisi sa ukupno 76 redovnih, vanrednih i drugih zasijedanja u 2016, 2017. i 2018. godini, kao i zasijedanjima iz prve polovine 2019. godine. Korpus ukupno sadrži 6199 snimaka u trajanju od po 10 minuta, što čini oko 1033 časa audio materijala. Na snimcima se pojavljuju 242 različita govornika koji pripadaju različitim etničkim i regionalnim grupama u Crnoj Gori, što, pored uobičajenih razlika u glasovnom rasponu, ritmu i dinamici govora, rezultuje i raznovrsnošću dijalekata i naglasaka u ovom korpusu. Na nekim snimcima ima buke, aplaudiranja i dovikivanja, što je uobičajeno za burne skupštinske debate, kao i drugih zvuka iz okruženja, što dodatno doprinosi raznolikosti ovog korpusa.

Svi govornici su odrasle osobe između 22 i 73 godine. Nesklad u broju muških i ženskih glasova je, nažalost, očekivan, zbog nedovoljne prisutnosti žena u crnogorskoj politici. Ukupna trajanja govora su takođe neravnomjerno raspodijeljena, što zbog kratkih gostovanja osoba koje nisu poslanici, što zbog neujednačene aktivnosti samih poslanika. Zbirne dužine audio zapisa po osobi su u rasponu od svega nekoliko sekundi, do preko 5 časova. Predsjedavajućima je u metapodacima pripisano po više od 10 časova audio snimaka. Međutim, ovaj disbalans se ne odražava negativno na kvalitet korpusa zbog velike količine podataka koja je na raspolaganju. Postoji dovoljan broj govornika koji je govorio dovoljno dugo. Prema dostupnim metapodacima, za 226 govornika postoji više od 5 minuta audio zapisa, a 138 osoba imalo je riječ u zbirnom trajanju od preko 30 minuta. Ove vrijednosti ne predstavljaju efektivna trajanja govornih aktivnosti jer u govorima postoje periodi tišine, koji su eliminisani. Pomenuti detalji prikazani su na Slici 24. Govornici su podijeljeni u pet grupa, prema ukupnom trajanju govora, a zatim je izračunat broj govornika po definisanim



Slika 24: Raspodjela ukupnog trajanja govora po grupama govornika.

grupama koja je prikazana na grafiku.

Tokom analize korpusa primijećeno je da u izlaganjima govornika često ima prekida. Takođe, nerijetko su snimani i djelovi sjednice u kojima niko ne govori. Ovo se obično dešava na počecima i krajevima sjednica ili u pauzama između izlaganja. Eksperimentalnim putem je utvrđen prag tišine od -35 dBFS. Svi intervali u audio signalima, sa trajanjem više od 1 sekunde, čiji je nivo snage ispod ove vrijednosti su eliminisani. Kraći intervali nisu uklanjani jer na krajevima riječi signal ima tendenciju da padne ispod definisanog praga tišine, pa bi njihovo brisanje izazvalo gubitak informacija. Zbog toga je zahtijevano da intervali tišine traju barem 1 sekund kako bi bili obrisani. To vrijeme smatra se dovoljnim da osoba dovrši izgovaranje započete riječi. Uostalom i tišina predstavlja validan audio signal i ne treba je u potpunosti zanemariti. Sistem vodenog žiga mora biti u mogućnosti da sakrije bitove i u ovim djelovima signala. Inače bi njegov kapacitet direktno zavisio od broja tihih segmenata u audio signalu, što bi bio očigledan nedostatak. Ipak, duži intervali u kojima nema govora su uklonjeni kako ne bi predstavljali dominantan segment korpusa, jer su od manjeg značaja od intervala u kojima ima govorne aktivnosti. Takođe, postoji šansa da bi se model mogao poremetiti ako bi bio suočen sa neopravdano velikim brojem tihih intervala tokom obučavanja. Fragmenti tišine, sasvim dovoljni za nepristrasno obučavanje modela, nalaze se u kratkim zastojima koje govornici prave u svojim obraćanjima. Operacijom uklanjanja tišine eliminisano je 165 časova audio snimaka, čime je korpus sveden na ukupnu dužinu od 868 časova.

Pored količine i raznolikosti, kvalitet podataka takođe ima veliki uticaj na ishod procedure obučavanja i performanse modela dubokog učenja. Frekvencija odabiranja u svim signalima u korpusu iznosi 44.1 kHz. Ova frekvencija odabiranja je nepotrebno velika kada se radi sa govornim signalima, jer frekvencija ljudskog glasa rijetko kada prelazi 8 kHz [163]. Uglavnom su signali proizvedeni ljudskim govornim aparatom između 1 i 5 kHz [164]. Kako prikupljeni korpus sadrži isključivo govorne signale, originalna frekvencija odabiranja se može značajno smanjiti na nivou čitavog korpusa, bez negativnih posljedica na kvalitet signala. Decimacijom ulaznih signala snižavaju se zahtjevi sistema, kako u pogledu memorije, tako i procesorske moći, jer se smanjivanjem dimenzija ulaza smanjuje broj operacija koje sistem treba da izvrši.

Standardna uskopojasna telefonska infrastruktura ograničava prenos audio signala na opseg širine 3.4 kHz, što je dovoljno da se očuvaju informacije. Ipak, mogući su gubici u kvalitetu s obzirom na to da govor može dosegnuti 8 kHz, što je uzeto kao maksimalna frekvencija u ovom radu. Uzimajući u obzir Nikvist-Šenonovu teoremu odabiranja [165], prema kojoj frekvencija odabiranja mora biti barem dva puta veća od maksimalne frekvencija signala kako bi se iz diskretizovane reprezentacije mogao u potpunosti rekonstruisati analogni signal, decimacija audio signala u korpusu izvršena je na frekvenciju od 16 kHz.

8 Rezultati

U ovom poglavlju izloženi su rezultati predloženog sistema vodenog žiga i data je uporedna analiza sa istaknutim tehnikama u oblasti. Tehnike su testirane po kriterijumima navedenim u Poglavlju 3. Eksperimenti su vršeni nad trećim dijelom korpusa podataka, izdvojenim isključivo za ovu namjenu, čime se osigurava nepristrasnost i objektivnost sprovedene procedure.

Raznovrsnost primjena audio votermarking sistema rezultovala je pristupima koji se različito odnose prema kriterijumima za procjenu performansi. Primarni kriterijumi za većinu tehnika su robustnost i očuvanje kvaliteta. Međutim, ova dva kriterijuma su međusobno suprotstavljeni. Očuvanje vodenog žiga prilikom napada često iziskuje njegovo ugrađivanje u signal nosilac s povećanim intenzitetom kako bi opstao nakon modifikacije, što negativno utiče na kvalitet signala. Dakle, s povećanjem robustnosti opada kvalitet signala i obratno, pa je neophodno napraviti odgovarajući kompromis.

Većina votermarking tehnika, pogotovo one kojima je u fokusu zaštita autorskih prava, veći značaj daju robustnosti, dok je drugima primarno očuvanje kvaliteta signala, a robustnost marginalna. Eksperimenti u ovom radu vršeni su nad govornim signalima. Kada se žigovi umeću u govorne signale, u cilju njihove zaštite, robustnost dobija veći značaj u odnosu na očuvanje kvaliteta signala. Kvalitet nije od presudnog značaja kada se obrađuje govor. Mnogo je važnije sačuvati informaciju koju on nosi, odnosno razumljivost. Dakle, dok god je govor lako razumljiv, određeni pad u kvalitetu može se tolerisati. Stoga, možemo smatrati da robustnost ima prioritet prilikom umetanja vodenih žigova u govorne signale.

Bez obzira na izbor osnovnog kriterijuma performansi, neophodno je održati dovoljan nivo performansi i po ostalim kriterijumima kako bi se osigurala primjenljivost sistema. Na primjer, sistem koji čuva vodeni žig u različitim scenarijima napada nije od koristi ako to čini uništavajući kvalitet signala nosioca. Takođe, maksimizacija kvaliteta votermarkovanog signala ne smije se ostvarivati po cijenu brisanja vodenog žiga jednostavnim efektima i napadima, ukoliko je potrebna robustna šema za detekciju.

Pristup kojim se vodi jedan dio tehnika je optimizacija ključnog kriterijuma performansi, uz održavanje nivoa performansi po ostalim kriterijumima na unaprijed zadatom, fiksiranom nivou. Druge tehnike definišu slobodne parametre kojima se može kontrolisati kompromis između različih kriterijuma performansi. Međutim, prethodno navedene različitosti čine da upoređivanje votermarking tehnika bude veoma delikatno. Na primjer, mnogi pristupi tvrde robustnost na različite napade,

ali se ona postiže pri različitim stepenima očuvanja kvaliteta signala i kapaciteta. Potpuno pravedno poređenje robustnosti moguće je tek ukoliko se tehnike najprije usklade po ostalim mjerilima, prevashodno u očuvanju kvaliteta signala i kapaciteta, što nije jednostavan poduhvat. Stepen očuvanja kvaliteta se nekada može ujednačiti podešavanjem vrijednosti slobodnih parametara, poput intenziteta vodenog žiga. Međutim, u savremenim votermarking tehnikama, ovi parametri nisu uvijek otvoreni za prilagođavanje, a nekada očuvanje kvaliteta može zavisiti od više parametara čije usklađivanje radi postizanja unaprijed definisane vrijednosti nije praktično izvodljivo. Osim toga, ujednačavanje votermarking tehnika po nekom kriterijumu mora se izvesti tako da se one, u najboljem, svedu na nivo najslabije tehnike, čime se zanemaruju napori uloženi u poboljšanja po datom kriterijumu.

Većina uporednih studija pribjegava kvalitativnom poređenju tehnika [166]. U tim studijama se, na osnovu prijavljenih rezultata i analize šema za umetanje i detekciju, identifikuje koji su kriterijumi performansi zadovoljeni. Na primjer, utvrđuje kojim napadima tehnika vodenih žigova ima sposobnost da se suprotstavi, ali ne i sa kojom stopom grešaka u detekciji.

U ovoj studiji sprovedena je kvantitativna evaluacija svih razmatranih tehnika. Prilikom realizacije odabranih tehnika korištene su preporučene vrijednosti slobodnih parametara i vrijednosti mjerila perfomansi dobijene sa tim postavkama su uzete za poređenje. Ovakvim pristupom se obezbjeđuju najpribližniji uslovi onima koji su pretpostavljeni prilikom dizajniranja datih tehnika, iako je jasno da postoje nedostaci. Sprovođenje iscrpne i sveobuhvatne uporedne studije je veoma složen zadatak, a prethodno navedeni razlozi otkrivaju zašto još uvijek nije izložena nijedna studija takve vrste u literaturi.

8.1 Robustnost

Robustnost sistema testirana je prema dvijema grupama efekata, navedenim u Sekciji 6.1.3. Prva grupa su ustaljeni audio efekti na koje gotovo svaki sistem vodenog žiga mora biti otporan. Ova grupa predstavlja efekte kojima audio signali mogu biti izloženi i bez malicioznih namjera, u uobičajenim okolnostima upotrebe sistema. Toj grupi pripadaju: aditivni šum, skaliranje amplitude i niskopropusno filtriranje. Druga grupa su efekti desinhronizacije. Oni se najčešće primjenjuju s namjerom da se onemogući detekcija vodenog žiga u signalu. Zbog svoje efikasnosti, zavrijedili su posebnu pažnju u literaturi. Testiranje otpornosti na ove napade jedan je od ključnih indikatora robustnosti votermarking tehnike.

	Model A	Model B	[16]	[80]	[88]	[7]	[15]
Bez napada	0.00	0.27	0.00	0.00	0.00	0.00	0.00
Aditivni šum	0.01	1.57	4.31	7.13	3.07	2.21	4.20
Skaliranje amplitude	0.03	1.30	0.00	0.00	0.00	0.00	0.00
Niskopropusni filtar	0.00	0.31	0.98	6.94	5.15	1.30	1.00

Tabela 5: Procenat pogrešno detektovanih bitova vodenog žiga pri ustaljenim audio efektima.

8.1.1 Otpornost na ustaljene audio efekte

Otpornost predložene tehnike na ustaljene audio efekte upoređivana je sa predstavnicima svih relevantnih grupa tehnika iz Sekcije 4.2, pri čemu je prioritet dat nedavno objavljenim radovima, afirmisanim u istraživačkoj zajednici. Uslijed nedostataka skorijih značajnijih iskoraka u *ad hoc* audio votermarking tehnikama u vremenskom domenu, sve evaluirane tehnike realizuju se u transformacionim domenima. Za predstavnika QIM tehnika izabrana je tehnika predložena u [16]. Vrednovana je i jedna od najskorije objavljenih LSB tehnika [80]. Tehnika [88] uzeta je kao predstavnik pečvork pristupa. U skup referentnih tehnika za analizu otpornosti na ustaljene efekte dodate su i tehnike [7, 15], koje su prevashodno namijenjene suzbijanju efekata desinhronizacije. Međutim, neophodno je ispitati otpornost ovih metoda na ustaljene audio efekte, kako bi se na valjan način mogle vrednovati. Otpornost na efekte desinhronizacije je suvišna, ako tehnika nije istovremeno otporna na osnovne audio efekte. Tehnika [7] se može svrstati u grupu QIM tehnika, dok je tehnika [15] bazirana na pečvorku. Time je upotpunjen referentni skup tehnika nad kojima su sprovedeni eksperimenti u kojima je izračunavana vrijednost BER mjere.

U Tabeli 5 su prikazane stope pogrešno detektovanih bitova vodenog žiga analiziranih tehnika u različitim scenarijima. Vrijednost BER mjere ispitivana je pri stopi umetanja od 40 bps, kako bi se osiguralo što pravednije poređenje. Model B predloženog sistema ostvario je odlične rezultate na testu robustnosti. Izračunate vrijednosti BER mjere manje su od 1.6% pri svakom od napada, što ukazuje na visoku otpornost. Ove greške mogu se gotovo u potpunosti otkloniti proširivanjem vodenih žigova kodovima za korekciju grešaka. Otpornost sistema na aditivni šum i niskopropusno filtriranje ukazuje na mogućnost njegove primjene u praktičnim okruženjima. U tim okruženjima signal može biti podvrgnut operacijama poput kompresije i prenosa komunikacionim kanalima, koje često podrazumijevaju prisustvo šumova i filtara. Model A ostvaruje superiornije rezultate u pogledu robustnosti na ustaljene audio efekte, ali pri nižem kapacitetu. Zbog toga je model B korišten u komparativnoj analizi, a rezultati za model A su samo evidentirani u tabeli.

Otpornost sistema na posmatranu grupu napada na nivou je najboljih metoda iz literature. Sve posmatrane tehnike gotovo savršeno ekstrahuju bitove vodenog žiga ukoliko signal nije izmijenjen nakon umetanja. Međutim, dodavanje šuma i niskopropusno filtriranje predstavljaju izazov svim votermarking tehnikama. Najizrazitiji pad u performansama primjetan je kod tehnike koja umetanje vrši u najmanje značajnim bitovima reprezentacije signala nosioca [80]. Kod preostalih tehnika također dolazi do povećanja procenta pogrešno detektovanih bitova vodenog žiga pri ovim napadima, u čemu ih predloženi sistem nadmašuje.

Otpornost sistema na niskopropusno filtriranje testirana je i sa idealnim i Batervortovim filtrom. Dobijeni su identični rezultati. Sistem je uspješno obučen da ne ugrađuje bitove vodenog žiga u visokim frekvencijama, što potvrđuju neznatne razlike u vrijednostima BER mjere za nenapadnute signale i signale propuštene kroz filtar. Napadi šumom i skaliranjem amplitude testirani su uz niskopropusni filtar, kako bi se i u tim situacijama osiguralo izbjegavanje visokih frekvencija.

Skaliranje amplitude je efekat koji nema negativan uticaj ni na jednu od uporednih metoda. Ovaj rezultat je očekivan s obzirom na to da sve šeme za detekciju savremenih konvencionalnih tehnika ekstrahuju bitove vodenog žiga analizom relativnog odnosa koeficijenata tranformacije ili izdvojenih karakteristika, a ne na osnovu njihovih apsolutnih vrijednosti. Nasuprot tome, ovaj efekat ispostavio se kao najizazovniji sistemu zasnovanom na dubokom učenju i u ovom aspektu on trenutno zaostaje za tradicionalnim pristupima.

U izvršenim eksperimentima je razmatrana i preciznost prepoznavanja votermarkovanih i nevotermarkovanih signala. Ovo je veoma važan indikator performansi sistema vodenog žiga, koji se u literaturi rijetko analizira. Stoga poređenje tehnika po ovom kriterijumu performansi nije sprovedeno, već je predloženi sistem evaluiran izolovano.

Oba modela predloženog sistema ostvaruju valjane rezultate na ovom zadatku. Diskriminator modela A je sastavni dio mreže za detekciju. Detektor u tom modelu nevotermarkovane signale tretira kao signale označene vodenim žigom sa rednim brojem $N_w + 1$. Ukoliko detektor isporuči ovaj vodeni žig na izlazu, signal se smatra nevotermarkovanim. Shodno tome preciznost u razlikovanju votermarkovanih i nevotermarkovanih signala inkorporirana je u vrijednosti BER mjere za model A, koja je navedena u Tabeli 5. Detektor je posebna neuronska mreža u modelu B, pa je odvojeno evaluiran i ostvareni rezultati su dati u Tabeli 6. Postignuta tačnost predviđanja je iznad 95% u gotovo svim scenarijima. Preciznost prepoznavanja votermarkovanih signala je posebno visoka, što je u većini primjena prioritet.

	Votermarkovani	Nevotermarkovani
Bez napada	99.68	96.57
Aditivni šum	97.17	92.81
Skaliranje amplitude	96.73	97.41
Niskopropusni filtar	99.45	98.73

Tabela 6: Preciznost prepoznavanja votermarkovanih i nevotermarkovanih signala od strane diskriminatora modela B, pri ustaljenim audio efektima, mjerena u procentima.

8.1.2 Otpornost na efekte desinhronizacije

U ovoj sekciji data je analiza robustnosti u pogledu efekata desinhronizacije. Predloženi pristup upoređuje se sa nedavno objavljenom tehnikom [7], dizajniranom posebno za prevazilaženje efekata desinhronizacije u audio signalima. Pored toga, testirana je i tehnika višeslojnog umetanja vodenog žiga predložena u [15]. Model A predloženog sistema korišćen je za testiranje, budući da je taj model obučavan sa ovom vrstom napada.

Rezultati koje su posmatrane tehnike ostvarile u sprovedenim eksperimentima prikazani su u Tabeli 7. Vrijednosti stope pogrešno detektovanih bitova za svaki od efekata date su u dva reda. Svaki efekat testiran je sa i bez uparivanja sa niskopropusnim filtrom, kako bi se ispitalo da li posmatrana tehnika otpornost na efekte desinhronizacije postiže ugrađivanjem bitova u visokim frekvencijama. Dobijene vrijednosti BER mjere ukazuju da je sistem uspješno obučen da izbjegava umetanje vodenog žiga u visokim frekvencijama. Ipak, upotreba visokofrekventnih koeficijenata nije u potpunosti spriječena, što pokazuju neznatna povećanja vrijednosti BER-a kada se napad upari sa niskopropusnim filtrom.

Očekivano, najveći broj grešaka u detekciji izazivaju napadi u kojima se gube odbirci signala, odnosno brisanje odbiraka i ubrzavanje reprodukcije. Razlog tome je što ovi napadi, pored gubitka odbiraka signala, uzrokuju gubitak informacija o ugrađivanju vodenog žiga. Smanjivanje frekvencije odabiranja takođe dovodi do gubitka odbiraka, ali rigidnijim skupom pravila, pa je stoga ovaj efekat lakše prevazići.

Razlika u performansama predloženog metoda u odnosu na uporedne tehnike u pogledu robustnosti je evidentna. Ona još više dolazi do izražaja u slučaju otpornosti na efekte desinhronizacije. Ovo je u potpunosti očekivano, budući da je našem sistemu omogućeno da se prilagodi ovim operacijama obrade signala tokom obuke. Osim toga, napadi desinhronizacije, primjenjivani u ovim eksperimentima, značajno su razorniji od napada na koje su navedene tehnike testirane u radovima u kojima

		Model A	[7]	[15]
Brisanje odbiraka	Bez filtra	1.34	16.37	50.02
	Sa filtrom	1.70	17.39	50.09
Permutacija odbiraka	Bez filtra	0.00	3.46	16.37
	Sa filtrom	0.03	3.50	17.39
Pomjeranje u vremenu	Bez filtra	0.02	48.44	18.22
	Sa filtrom	1.58	49.06	18.39
Decimacija	Bez filtra	0.02	49.80	50.00
	Sa filtrom	0.05	50.02	50.17
Interpolacija	Bez filtra	0.01	49.06	50.31
	Sa filtrom	0.01	50.24	50.59
Ubrzavanje reprodukcije	Bez filtra	3.02	38.33	39.44
	Sa filtrom	7.85	38.62	39.86
Usporavanje reprodukcije	Bez filtra	0.38	36.33	44.06
	Sa filtrom	0.57	36.70	44.57

Tabela 7: Procenat pogrešno detektovanih bitova vodenog žiga pri efektima desinhronizacije.

su prvobitno predstavljene.

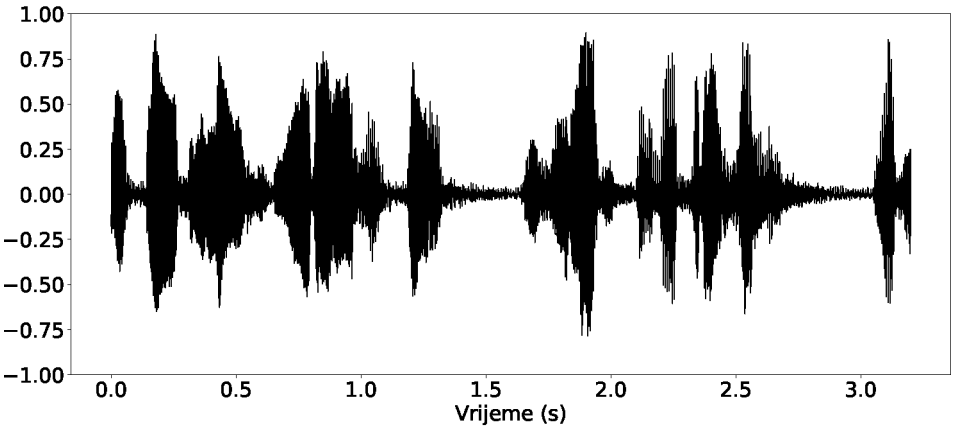
Kako je primarni cilj dizajniranog sistema bio postizanje robustnosti, na osnovu izloženih rezultata on se može smatrati ostvarenim. U nastavku je pokazano da je ostvarivanje ovog cilja postignuto bez ozbiljnog ugrožavanja ostalih mjera performansi.

8.2 Kvalitet signala

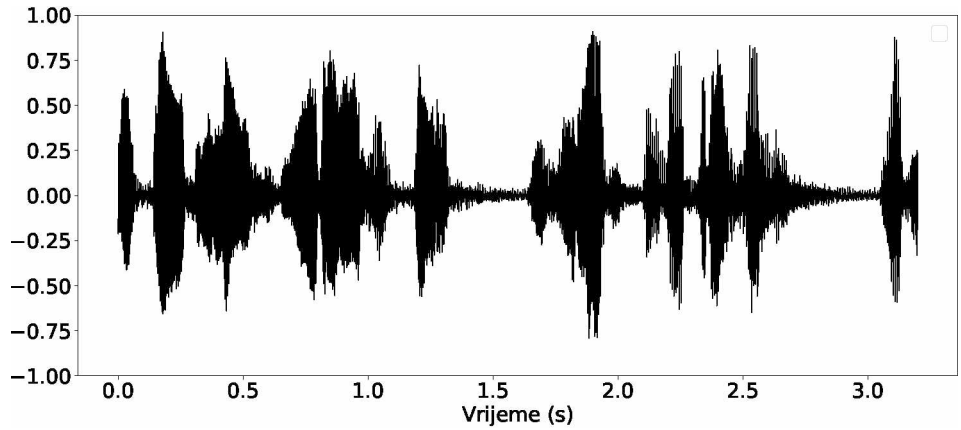
Predloženim modelima, kao i svim referentnim tehnikama iz Sekcije 8.1, utvrđena je sposobnost umetanja vodenog žiga bez značajnog ugrožavanja kvaliteta signala. U skladu sa trenutnim preporukama u oblasti, neprimjetnost vodenog žiga procjenjuje se razlikom kvaliteta originalnog i votermarkovanog signala, odnosno izračunavanjem

	Model A	Model B	[16]	[80]	[88]	[7]	[15]
PESQ	4.33(2.83)	3.68	4.45	4.21	4.08	3.91	4.43
SNR	38.75(21.03)	23.41	27.59	37.95	16.16	21.03	38.60

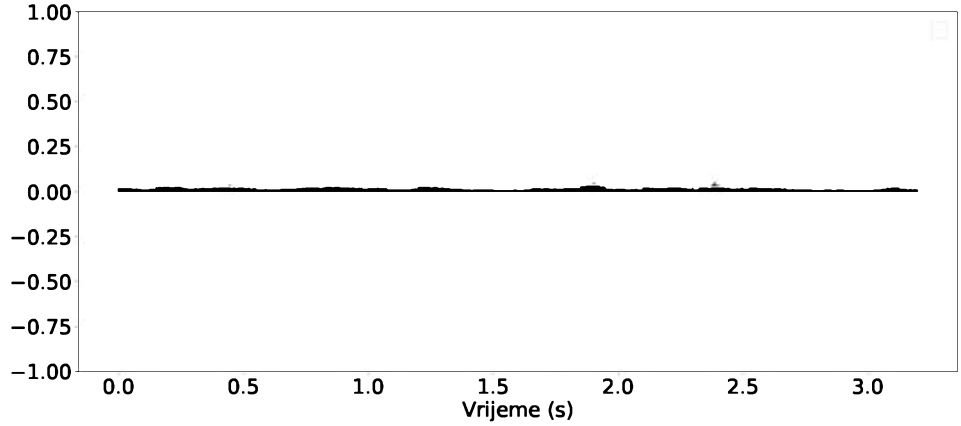
Tabela 8: Rezultati votermarking tehnika u pogledu očuvanja kvaliteta signala.



(a) Originalni signal nosilac.

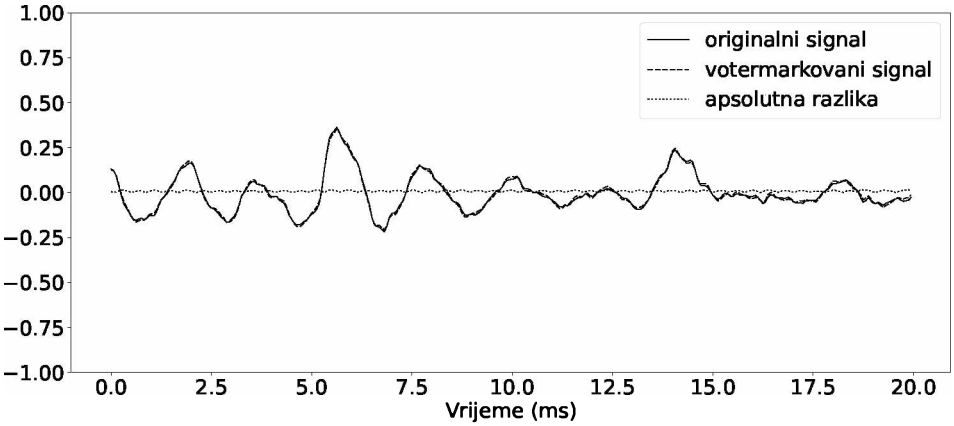


(b) Votermarkovani signal.



(c) Apsolutna razlika originalnog i votermarkovanog signala.

Slika 25: Prikaz isječka originalnog, votermarkovanog signala i njihove apsolutne razlike.



Slika 26: Uveličani prikaz isječka originalnog, votermarkovanog signala i njihove apsolutne razlike.

SNR i PESQ vrijednosti. Eksperimentalni rezultati, dobijeni na skupu za testiranje, dati su u Tabeli 8. Određeni pad u kvalitetu audio signala nakon umetanja vodenog žiga je opravdan i očekivan, s obzirom na to da se u njega ugrađuju dodatni podaci. Štaviše, nekada je umetanjem vodenog žiga potrebno stvoriti čujne artefakte, kako žig ne bi mogao biti uklonjen operacijama poput kompresije sa gubicima i efektima desinhronizacije, Svakako, pad u kvalitetu ne smije biti drastičan kako se signal nosilac ne bi obezvrijedio.

Model A dostigao je PESQ vrijednost preko 4.3 i SNR veći od 38 dB, što sugeriše da je umetnuti vodeni žig u potpunosti nečujan. Model B, iako većeg kapaciteta, održava kvalitet signala na veoma visokom nivou. Slika 25, na kojoj je prikazan primjer jednog isječka signala nosioca i odgovarajući isječak votermarkovanog signala to vizuelno i potvrđuje. Primjer votermarkovanog signala proizveden je modelom B. Između data dva grafika ne mogu se uočiti bitne razlike.

Vrijednosti u zagradama, navedene za model A, predstavljaju instancu modela obučavanu sa efektima desinhronizacije. U ovom slučaju, pad u kvalitetu signala je primjetan. PESQ vrijednosti između 2.5 i 3.5 tumače se kao čujna izobličenja signala koja neznatno utiču na iskustvo slušaoca. Sadržaj signala, odnosno razumljivost govora, u potpunosti je očuvana. Stoga je ovaj kompromis u kvalitetu signala prihvatljiv, budući da su postignuti odlični rezultati u pogledu robustnosti. Postavljeni cilj u pogledu očuvanja razumljivosti govora zasigurno je ispunjen.

Vrijednost PESQ mjere od 3.68 za model B ukazuje na prisustvo čujnih artefakata u signalu. Međutim, prema MOS skali, ukoliko je PESQ vrijednost iznad 3.5, kreirani artefakti ne utiču na utisak slušaoca i na razumljivost govora. Isto se može zaključiti i reprodukcijom ovih signala. Na grafiku sa Slike 25c iscrtana je apsolutna

razlika originalnog i votermarkovanog signala. Na ovoj slici se potvrđuje mala razlika između originalnog signala nosioca i signala označenog vodenim žigom. Radi boljeg prikaza, na Slici 26 dat je uveličani prikaz jednod kratkog isječka, na kojem se jasnije može se jasnije mogu vidjeti promjene u signalu nastale umetanjem vodenog žiga, kao i greška u rekonstrukciji. Amplituda apsolutne razlike je značajno manja od amplitude signala, pa stoga ona neće biti čujna. Prikazana razlika može se posmatrati kao reprezentacija vodenog žiga u vremenu. Shodno tome, data slika ukazuje da je vodeni žig prisutan u intervalima u kojima ima govorne aktivnosti, dok je odsutan u intervalima tišine. Ovakvo ponašanje je poželjno, jer je vodeni žig znatno teže ukloniti iz intervala sa govorom. Razlika originalnog i votermarkovanog signala obuhvata i grešku u rekonstrukciji koju pravi mreža umetača. Međutim, nesumnjivo je da su informacije o vodenom žigu njen primarni činilac.

Sve vrijednosti PESQ mjere veće od 4 kvalifikuju se kao u potpunosti neprimjetne razlike u govornim signalima, koje ljudski slušni sistem ekvivalentno tretira. Na osnovu vrijednosti mjera kvaliteta u Tabeli 8, može se zaključiti da su performanse svih uporednih tehnika po ovom kriterijumu slične, kao i da njihova primjena ne utiče osjetno na ljudski sluh. Kod modela A i B primjetna je određena nesrazmjernost u vrijednostima SNR i PESQ mjera, u odnosu na uporedne tehnike. Vrijednosti SNR mjere na nivou su najboljih tehnika u oblasti, dok su vrijednosti PESQ nesrazmjerno niže u poređenju sa ostalim tehnikama. Pretpostavka je da je razlog ovoga upotreba srednje kvadratne greške u proceduri obučavanja, koja se bliže podudara sa SNR nego sa PESQ mjerom.

8.3 Kapacitet

Kako bi ustanovili kapacitet sistema vodenog žiga, realizovanog modelom A, potrebno je odrediti količinu informacija (*engl. information gain*) koju ovaj sistem može prenijeti u jednoj sekundi audio signala. U teoriji informacija se za mjerenje količine informacija koje donosi saznanje o ishodu slučajnog događaja koristi entropija.

Modelom A predviđeno je ugrađivanje N_w vodenih žigova u intervalima dugim 2 sekunde. Ugrađivanje vodenog žiga ovim modelom možemo posmatrati kao slučajni događaj sa N_w mogućih ishoda, pri čemu svaki od ishoda ima jednaku vjerovatnoću javljanja, jednaku $\frac{1}{N_w}$. Tada se količina informacija koju dobijamo saznanjem koji je žig ugrađen u dati interval audio signala izračunava sa:

$$H = -\sum_{i=1}^{N_w} \frac{1}{N_w} \log_2 \left(\frac{1}{N_w} \right). \tag{92}$$

S obzirom na to da se vodeni žig ugrađuje u dvije sekunde audio signala, vrijednost dobijenu u prethodnoj jednakosti treba podijeliti sa 2 kako bi se dobio kapacitet posmatranog sistema. Ispostavlja da je kapacitet modela A, obučavanog bez efekata desinhronizacije, 1.5 bps, dok verzija modela A, otporna na desinhronizaciju, može ugrađivati informacije u audio signal stopom od 1 bps. Promjene u arhitekturi sistema i proceduri obučavanja, uvedene modelom B, rezultovale su značajnim povećanjem kapaciteta na vrijednost od 40 bps.

Poređenje predložene tehnike sa drugim metodama u pogledu kapaciteta nije moguće izvesti na potpuno korektan način, zbog fundamentalnih razlika u dizajnu šema umetanja. Tradicionalnim metodama se u signal može ugraditi vodeni žig proizvoljne dužine. Međutim, ugrađivanje većeg broja bitova u jedinici vremena negativno utiče na kvalitet votermarkovanog signala i preciznost detekcije. Nasuprot tome, predloženi sistem neuronskih mreža obučavan je da za vodene žigove fiksne dužine ostvari što bolje rezultate u pogledu efikasnosti, robustnosti i očuvanja kvaliteta signala. Ove suprotnosti u dizajnu šeme umetanja onemogućavaju direktno poređenje kapaciteta između ovih sistema.

8.4 Računska složenost

Vremenska složenost predloženog pristupa razmatrana je s teorijskog aspekta, jer vrijeme izvršavanja bilo koje operacije sistema vodenog žiga zavisi od raspoloživih hardverskih resursa. Osim toga, votermarking šeme, koje uključuju operacije nad realnim brojevima, prikladnije je izvršavati na grafičkoj procesorskoj jedinici (GPU), dok votermarking operacije, zasnovane na cjelobrojnoj aritmetici, bolje odgovaraju centralnoj procesorskoj jedinici (CPU). Takođe, postojeća literatura audio votermarkinga ne sadrži poređenja po ovom kriterijumu.

Sa striktno teorijskog stanovišta može se tvrditi da je vremenska složenost predloženog sistema $\mathcal{O}(N_I)$, gdje je N_I broj intervala na koji se dijeli signal nosilac pri votermarkovanju. Dužina intervala za model A iznosi 2 sekunde, dok je za model B 25 milisekundi.

Pojedinačni intervali obrađuju se neuronskim mrežama, odnosno konačnim skupom slojeva, s ograničenom veličinom ulaza i izlaza koji se primjenjuju sekvencijalno. Vremenska složenost ovog dijela sistema može se smatrati konstantnom, s obzirom na to da parametri dizajna arhitekture neuronskih mreža imaju konstantne vrednosti. Međutim, ako dozvolimo da broj slojeva i dimenzije ulaza budu promjenljivi, vremenska složenost bila bi jednaka vremenskoj složenosti najkompleksnijeg sloja pomnoženoj sa brojem takvih slojeva.

Najsloženije operacije u predloženom sistemu neuronskih mreža su Furijeova transformacija i konvolucija. Vremenska složenost procedure za efikasno računanje Furijeove transformacije (*engl. fast Fourier transform* - FFT) je $\mathcal{O}(N \log_2 N)$ [167], gdje je N broj odbiraka signala u vremenskom domenu. Vremenska složenost kratkotrajne Furijeove transformacije, koja se izračunava na ulazu umetača i detektora u model A, je $\mathcal{O}(N_\xi L_\xi \log_2 L_\xi)$, gdje je N_ξ broj pozicija prozora koje su uzete prilikom izračunavanja, L_ξ njegova dužina. Konvolucija bi se mogla izračunati kao proizvod u domenu Furijeove transformacije, čime bi se njena vremenska složenost mogla svesti na vremensku složenost FFT. Međutim, u realizaciji predloženog sistema konvolucija se izračunava direktno, jer su dimenzije konvolucionih filtara znatno manje od dimenzija signala. Može se smatrati da je $K \lesssim \log_2 N \iff \exists c(c > 0) \wedge (K < cN)$. Rezultujuća vremenska složenost je tada $\mathcal{O}(NK)$, za 1D konvoluciju u modelu B, odnosno $\mathcal{O}(N_1 N_2 K_1 K_2)$ za 2D konvoluciju u modelu A. Simboli N, N_1, N_2 predstavljaju dimenzije ulaza, a simboli K, K_1 i K_2 dimenzije konvolucionih filtara.

Uporedne votermarking tehnike takođe najprije dijele signal na više intervala u kojima se zatim sukcesivno ugrađuju bitovi vodenog žiga. Najsloženija operacija u okviru tehnika u transformacionom domenu je najčešće upravo prevođenje signala u drugi domen. Ovo preslikavanje obavlja se za $\mathcal{O}(L_I \log_2 L_I)$ vremena, ukoliko se primjenjuju transformacije poput diskretne Furijeove ili kosinusne, a L_I je dužina posmatranog intervala signala. Vremenska složenost diskretne vejvlet transformacije, realizovane Malatovim algoritmom [168] pomoću banke filtara, je $\mathcal{O}(L_I)$, što ovu transformaciju čini efikasnijom od drugih i, u tom pogledu, favorizuje tehnike koje je koriste.

Nakon konverzije u odabrani domen, dalji koraci najčešće uključuju računanje neke statistike ili norme vektora koeficijenata, čija je vremenska složenost $O(L_I)$. Eho kernel metode koriste konvoluciju. Tehnike poput [7], vrše dekompoziciju vektora koeficijenata na singularne vrijednosti, koje zatim koriste za ugrađivanje bitova vodenog žiga. Ova operacija se Golub-Rajnšovom metodom [169] može realizovati sa vremenskom složenošću $O(L_I)$, jer je, u svim razmatranim tehnikama, jedna dimenzija matrice koja se dekomponuje konstantna.

Date ocjene vremenske složenosti odnose se i na proceduru umetanja i na proceduru detekcije, budući da su ove dvije procedure uglavnom jednako kompleksne. Procedura detekcije se često svodi na ponavljanje koraka umetanja i detektovanje izmjena koje su tom procedurom unijete u signal. U nekim slučajevima, procedura detekcije je jednostavnija od procedure umetanja, ali nikada složenija.

Analizom uporednih votermarking tehnika zaključeno je da je najveći broj ovih tehnika vremenske složenosti $\mathcal{O}(N_I L_I \log_2 L_I)$, odnosno $\mathcal{O}(N \log_2 L_I)$, gdje je N uku-

	Model A	Model B
Umetač	16130818	5606529
Diskriminator	–	553793
Detektor	39208192	553793

Tabela 9: Broj parametara u dubokim neuronskim mrežama modela A i modela B.

pan broj odbiraka u signalu, $N = N_I L_I$. Neke tehnike koje koriste isključivo vejvlet transformaciju, poput [5], ili vrše umetanje i detekciju u vremenskom domenu [72] mogu imati složenost $\mathcal{O}(N)$.

S obzirom na to da su sve neuronske mreže u predloženom sistemu fiksne veličine, njegova prostorna složenost je $\mathcal{O}(1)$. Neuronskim mrežama se vezano za memorijsku složenost obično navodi broj parametara, pa su oni navedeni u Tabeli 9. Iz date tabele evidentna je razlika u veličinama modela A i modela B, čime se dodatno naglašavaju performanse ostvarene manjim modelom, većeg kapaciteta.

9 Zaključak

Votermarking je jedna od osnovnih tehnika zaštite digitalnog sadržaja, koja obilježavanjem digitalnih podataka nizom bitova nazvanim vodeni žig čuva autorska prava i autentičnost. Na taj način sprečavaju se krađa intelektualne svojine, povrede integriteta informacija, ugleda ličnosti, ustanova i organizacija. Vodeni žig može se primjenjivati i u digitalnoj forenzici, praćenju emitovanja, metapodacima i drugim kontekstima.

U ovoj disertaciji predložena je i realizovana nova paradigma za razvoj sistema vodenog žiga za digitalne audio signale, zasnovana na dubokom učenju. Duboke neuronske mreže korištene su kao osnovna tehnologija za obavljanje svih ključnih zadataka jednog sistema vodenog žiga, što predstavlja prvi takav pristup u literaturi. Sposobnosti generalizacije i univerzalne aproksimacije, koje su ključna svojstva dubokih neuronskih mreža, čine ih veoma pogodnim za modelovanje šema za umetanje i detekciju vodenih žigova.

Prednost predloženog pristupa je i u tome što predstavlja temelj za razvoj audio votemarking sistema koji se mogu prilagođavati specifičnim potrebama korisnika i na taj način poboljšavati performanse. Upotreba dubokih neuronskih mreža donosi veću fleksibilnost i efikasnost u razvoju i primjeni vodenih žigova u različitim aplikacijama. Ovaj novi obrazac razvoja omogućava istraživačima i inženjerima da koriste sopstvene skupove podataka i dodaju sopstvene implementacije efekata kako bi kreirali votermarking sisteme usklađene sa zahtjevima njihovih projekata. Tradicionalne tehnike dizajnirane su imajući u vidu potencijalne efekte i vrstu signala nad kojima se primjenjuju. Shodno tome, zahtijevale bi značajne metodološke izmjene u cilju prilagođavanja novim vrstama efekata ili signala. Stoga smatramo da će u budućnosti predloženi obrazac biti preovlađujući način za kreiranje votermarking sistema, ali i osnova za dalje inovacije u ovoj oblasti.

Predloženi sistem neuronskih mreža je obučavan u striktno kontrolisanoj proceduri, s osnovnim ciljem efikasnog i robustnog ugrađivanja vodenog žiga, uz očuvanje kvaliteta signala nosioca. Performanse sistema ispitane su po svim glavnim kriterijumima i upoređene sa najistaknutijim pristupima u oblasti. Posebna pažnja je posvećena analizi različitih vrsta efekata koji mogu uticati na signal koji sadrži vodeni žig i koji mogu rezultovati gubitkom ili oštećenjem vodenog žiga. Sistem je ostvario visoku otpornost na testirane efekte i time nadmašio uporedne pristupe. Pokazano je da efekti desinhronizacije predstavljaju najozbiljniju prepreku i za ovu vrstu sistema, ali da se neuronskim mrežama mogu uspješnije savladavati nego tradicionalnim pristupima.

Iako je predloženo rješenje ostvarilo zavidne rezultate u primarnim kriterijumima performansi, postoji nekoliko aspekata po kojima su moguća dalja unapređenja.

Sve poznate tehnike, kao i one predložene u ovoj disertaciji, ugrađuju vodene žigove u isječke signala ograničene dužine. Desinhronizacija ovih blokova može značajno poremetiti proces detekcije jer je neophodno precizno odrediti gdje je početak, a gdje kraj svakog isječka. Jedno od mogućih rješenja ovog problema je primjena sekvenca-u-sekvencu (*engl. sequence-to-sequence*) modela, koji su dizajnirani tako da mapiraju ulazne na izlazne sekvence, bez obzira na njihovu dužinu. Ovi modeli mogu se koristiti za votermarkovanje čitavih audio sekvenci, čime bi se izbjegao problem identifikovanja tačne pozicije početka i kraja isječaka signala.

Dizajn novih, kompleksnijih i robustnijih arhitektura neuronskih mreža i sofisticiranih procedura obučavanja svakako predstavljaju stalni osnov za poboljšavanje performansi sistema. Ipak, postoje i drugi potencijalni pravci za dalja istraživanja.

Otpornost na efekte desinhronizacije je u ovoj disertaciji ostvarena uz značajne kompromise u kapacitetu. Jedan od prvih ciljeva u nastavku ovog istraživanja biće stvaranje modela koji će s većom stopom ugrađivati bitove vodenog žiga a istovremeno biti otporan na efekte desinhronizacije. Prvi korak u tom pravcu je napravljen, obučavanjem modela B, koji dostiže veći kapacitet od modela A pri ustaljenim audio efektima.

Različita sredstva mogu se razmotriti kako bi potreba za smanjenjem kapaciteta bila minimizovana. Jedan od mogućih smjerova razvoja je unapređenje šeme umetanja tako da formira karakteristike signala nosioca invarijantne na desinhronizaciju. Ove karakteristike ostaju nepromijenjene pri efektima desinhronizacije i ukoliko bi se mogle sintetisati, predstavljale bi idealan domen za ugrađivanje bitova vodenog žiga.

Sistem se eventualno može proširiti i komponentom za identifikaciju efekata desinhronizacije. Zatim bi se interpolacijom prvobitnog votermarkovanog signala, odnosno inverzijom prepoznatog efekta, mogla nesmetano izvršiti detekcija vodenog žiga. Neuronske mreže se mogu, zajedno sa tradicionalnim tehnikama obrade signala, iskoristiti za obavljanje zadataka klasifikacije i invertovanja efekata desinhronizacije. Integrisanjem ovih komponenti u arhitekturu sistema postigao bi se sličan efekat ugrađivanju sinhronizujućih kodova kod drugih votermarking tehnika.

Stepen očuvanja kvaliteta signala, ostvaren predloženim sistemom, u ravni je sa najboljim tehnikama u oblasti. Međutim, zasigurno da se performanse po ovom kriterijumu mogu dalje poboljšavati. To bi se moglo postići dizajniranjem funkcije gubitka kojom se preciznije ocjenjuje preceptivno rastojanje votermarkovanog i originalnog audio signala. Jedan od mogućih način je i dodavanje člana u funkciju

gubitka umetača kojim se ta neuronska mreža obučava tako da maksimizuje grešku diskriminatora originalnih i votermarkovanih signala, odnosno tretiranjem umetača i diskriminatora kao generativnih suprotstavljenih neuronskih mreža (*engl. generative adversarial networks* - GANs).

Skup razmatranih napada može se proširiti svim efektima iz StirMark skupa. Takođe, neophodno je analizirati otpornost sistema na efekat presnimavanja. Ovim efektom simulira se reprodukcija votermarkovanog signala sa jednog uređaja, odnosno zvučnika, i njegovo snimanje na drugom uređaju, tj. mikrofonu. Ovaj efekat je veoma važan jer predstavlja jedan veoma čest scenario u kojem se dešava povreda autorskih prava.

Neuronska mreža se može iskoristiti za dizajniranje napada na sisteme vodenog žiga, kada napadač u posjedu ima detektor ili kada je šema za detekciju javna. Ovo je takođe jedan od pravaca koji vrijedi ispitati. Cilj takve mreže bilo bi modifikovanje votermarkovanog signala tako da detektor u njemu ne pronalazi vodeni žig ili detektuje pogrešan vodeni žig. Naravno, ovo je potrebno postići bez osjetnog pada u kvalitetu signala.

Dalja praktična proširenja dizajniranog sistema mogu uključiti višestruko votermarkovanje audio signala. Ovaj vid votermarkinga je potreban u situacijama kada digitalni podaci imaju više vlasnika i svaki od njih želi da označi svoje vlasništvo. Takođe, ugrađivanje više vodenih žigova može biti potrebno i prilikom distribuiranja digitalnog signala putem lanca posrednika. Svaki posrednik u lancu može ugraditi svoj vodeni žig kako bi identifikovao potencijalna curenja informacija ili pratio emitovanje.

Hiperparametri, poput intenziteta vodenog žiga, mogu se osloboditi kako bi korisnik mogao da ih podešava i postigne željeni kompromis robustnosti i očuvanja kvaliteta. Proširenja korpusa podataka, prije svega uključivanje različitih muzičkih žanrova, su takođe poželjna. Ova dodatna raznolikost podataka povećala bi upotrebljivost sistema kroz cjelovitiju analizu performansi na više tipova medija, sa različitim kvalitetom signala. Konačni cilj ovog istraživanja je razviti rješenje koje je primjenljivo u različitim scenarijima, a na tom putu postoje mnogi izazovi.

Literatura

[1] Z. Liu and A. Inoue. Audio watermarking techniques using sinusoidal patterns based on pseudo-random sequences. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(8):801–812, 2003.

[2] M. Arnold, X.-M. Chen, P. Baum, U. Gries, and G. Doërr. A phase-based audio watermarking system robust to acoustic path propagation. *IEEE Transactions on Information Forensics and Security*, 9(3):411–425, 2014.

[3] Y. Xiang, I. Natgunanathan, D. Peng, G. Hua, and B. Liu. Spread spectrum audio watermarking using multiple orthogonal PN sequences and variable embedding strengths and polarities. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(3):529–539, 2018.

[4] Z. Liu, Y. Huang, and J. Huang. Patchwork-based audio watermarking robust against de-synchronization and recapturing attacks. *IEEE Transactions on Information Forensics and Security*, 14(5):1171–1180, 2019.

[5] W. Jiang, X. Huang, and Y. Quan. Audio watermarking algorithm against synchronization attacks using global characteristics and adaptive frame division. *Signal Processing*, 162:153–160, 2019.

[6] M.-J. Hwang, J.-S. Lee, M.-S. Lee, and H.-G. Kang. SVD-based adaptive QIM watermarking on stereo audio signals. *IEEE Transactions on Multimedia*, 20(1):45–54, 2018.

[7] J. Zhao, T. Zong, Y. Xiang, L. Gao, W. Zhou, and G. Beliakov. Desynchronization attacks resilient watermarking method based on frequency singular value coefficient modification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:2282–2295, 2021.

[8] T. Furon. A constructive and unifying framework for zero-bit watermarking. *IEEE Transactions on Information Forensics and Security*, 2(2):149–163, 2007.

[9] A. Dwivedi, A. Kumar, M.K. Dutta, R. Burget, and V. Myska. An efficient and robust zero-bit watermarking technique for biometric image protection. In *Proc. of the International Conference on Telecommunications and Signal Processing*, pages 236–240, 2019.

[10] M. Chen, Y. He, and R.L. Lagendijk. A fragile watermark error detection scheme for wireless video communications. *IEEE Transactions on Multimedia*, 7(2):201–211, 2005.

[11] X. Zhang and S. Wang. Fragile watermarking with error-free restoration capability. *IEEE Transactions on Multimedia*, 10(8):1490–1499, 2008.

[12] A. Shehab, M. Elhoseny, K. Muhammad, A.K. Sangaiah, P. Yang, H. Huang, and G. Hou. Secure and robust fragile watermarking scheme for medical images. *IEEE Access*, 6:10269–10278, 2018.

[13] X. Qi and X. Xin. A quantization-based semi-fragile watermarking scheme for image content authentication. *Journal of Visual Communication and Image Representation*, 22(2):187–200, 2011.

[14] B. Widrow. Statistical analysis of amplitude-quantized sampled-data systems. *Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry*, 79(6):555–568, 1961.

[15] I. Natgunanathan, Y. Xiang, G. Hua, G. Beliakov, and J. Yearwood. Patchwork-based multilayer audio watermarking. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(11):2176–2187, 2017.

[16] H.-T. Hu and T.-T. Lee. Frame-synchronized blind speech watermarking via improved adaptive mean modulation and perceptual-based additive modulation in dwt domain. *Digital Signal Processing*, 87:75–85, 2019.

[17] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2):113–120, 1979.

[18] Y. Lu and P.C. Loizou. A geometric approach to spectral subtraction. *Speech Communication*, 50(6):453–466, 2008.

[19] Y. Ephraim and H.L. Van Trees. A signal subspace approach for speech enhancement. *IEEE Transactions on Speech and Audio Processing*, 3(4):251–266, 1995.

[20] N. Nguyen, P. Milanfar, and G.H. Golub. A computationally efficient superresolution image reconstruction algorithm. *IEEE Transactions on Image Processing*, 10(4):573–583, 2001.

[21] W.T. Freeman, T.R. Jones, and E.C. Pasztor. Example-based superresolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.

[22] C. Dong, C.C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *Proc. of the European Conference on Computer Vision*, pages 184–199, 2014.

[23] B. Lim, S. Son, H. Kim, S. Nah, and K.-M. Lee. Enhanced deep residual networks for single image super-resolution. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017.

[24] H. Fletcher and W.A. Munson. Loudness, its definition, measurement and calculation. *The Journal of the Acoustical Society of America*, 5(2):82–108, 1933.

[25] Y. Suzuki and H. Takeshima. Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America*, 116(2):918–933, 2004.

[26] T. Painter and A. Spanias. Perceptual coding of digital audio. *Proc. of the IEEE*, 88(4):451–515, 2000.

[27] P. Bas and T. Furon. A new measure of watermarking security: The effective key length. *IEEE Transactions on Information Forensics and Security*, 8(8):1306–1317, 2013.

[28] E.H. Rothouser. IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, 17(3):225–246, 1969.

[29] Y. Hu and P.C. Loizou. Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(1):229–238, 2008.

[30] J.H.L. Hansen and B.L. Pellom. An effective quality evaluation protocol for speech enhancement algorithms. In *Proc. of the International Conference on Spoken Language Processing*, 1998.

[31] J. Tribolet, P. Noll, B. McDermott, and R. Crochiere. A study of complexity and quality of speech waveform coders. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 586–590, 1978.

[32] S. Quackenbush, T. Barnwell, and M. Clements. *Objective measures of speech quality.* Prentice Hall, 1988.

[33] P.C. Loizou. *Speech Enhancement.* CRC Press, 2007.

[34] P. Ladefoged and K. Johnson. *A Course in Phonetics*. Wadsworth Publishing, 2014.

[35] D. Klatt. Prediction of perceived phonetic distance from critical-band spectra: A first step. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1278–1281, 1982.

[36] S. Wang, A. Sekey, and A. Gersho. An objective measure for predicting subjective quality of speech coders. *IEEE Journal on Selected Areas in Communications*, 10(5):819–829, 1992.

[37] J.D. Johnston. Transform coding of audio signals using perceptual noise criteria. *IEEE Journal on Selected Areas in Communications*, 6(2):314–323, 1988.

[38] A. Rix, J. Beerends, M. Hollier, and A. Hekstra. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 749–752, 2001.

[39] J.G. Beerends and J.A. Stemerdink. A perceptual speech-quality measure based on a psychoacoustic sound representation. *Journal of The Audio Engineering Society*, 42(3):115–123, 1994.

[40] T. Thiede, W. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten. PEAQ–the ITU standard for objective measurement of perceived audio quality. *Journal of the Audio Engineering Society*, 48(1.2):3–29, 2000.

[41] A. Neubauer, J. Freudenberger, and V. Kuhn. *Coding theory*. Wiley-Blackwell, 2007.

[42] L. Boney, A.H. Tewfik, and K.N. Hamdy. Digital watermarks for audio signals. In *Proc. of the IEEE International Conference on Multimedia Computing and Systems*, pages 473–480, 1996.

[43] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, 1997.

[44] N. Nikolaidis and I. Pitas. Robust image watermarking in the spatial domain. *Signal Processing*, 66(3):385–403, 1998.

[45] W. Zeng and B. Liu. A statistical watermark detection technique without using original images for resolving rightful ownerships of digital images. *IEEE Transactions on Image Processing*, 8(11):1534–1548, 1999.

[46] I. Djurović, S. Stanković, and I. Pitas. Digital watermarking in the fractional Fourier transformation domain. *Journal of Network and Computer Applications*, 24(2):167–173, 2001.

[47] S. Stanković, I. Djurović, and I. Pitas. Watermarking in the space/spatial-frequency domain using two-dimensional Radon-Wigner distribution. *IEEE Transactions on Image Processing*, 10(4):650–658, 2001.

[48] C.-W. Tang and H.-M. Hang. A feature-based robust digital image watermarking scheme. *IEEE Transactions on Signal Processing*, 51(4):950–959, 2003.

[49] C.-C. Lai and C.-C. Tsai. Digital image watermarking using discrete wavelet transform and singular value decomposition. *IEEE Transactions on Instrumentation and Measurement*, 59(11):3060–3063, 2010.

[50] L. Luo, Z. Chen, M. Chen, X. Zeng, and Z. Xiong. Reversible image watermarking using interpolation technique. *IEEE Transactions on Information Forensics and Security*, 5(1):187–193, 2010.

[51] S. Kirbiz and B. Gunsel. Robust audio watermark decoding by supervised learning. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages V–V, 2006.

[52] X. Xu, H. Peng, and C. He. DWT-based audio watermarking using support vector regression and subsampling. In *Proc. of the International Workshop on Fuzzy Logic and Applications*, pages 136–144, 2007.

[53] A.M. Abdelhakim and M. Abdelhakim. A time-efficient optimization for robust image watermarking using machine learning. *Expert Systems with Applications*, 100:197–210, 2018.

[54] H. Kandi, D. Mishra, and S.R.K.S. Gorthi. Exploring the learning capabilities of convolutional neural networks for robust image watermarking. *Computers & Security*, 65:247–268, 2017.

[55] S.-M. Mun, S.-H. Nam, H. Jang, D. Kim, and H.-K. Lee. Finding robust domain from attacks: A learning framework for blind watermarking. *Neurocomputing*, 337:191–202, 2019.

[56] R. Sinhal, D.K. Jain, and I.A. Ansari. Machine learning based blind color image watermarking scheme for copyright protection. *Pattern Recognition Letters*, 145:171–177, 2021.

[57] J. Zhu, R. Kaplan, J. Johnson, and L. Fei-Fei. HiDDeN: Hiding data with deep networks. In *Proc. of the European Conference on Computer Vision*, pages 682–697, 2018.

[58] M. Ahmadi, A. Norouzi, N. Karimi, S. Samavi, and A. Emami. ReDMark: Framework for residual diffusion watermarking based on deep networks. *Expert Systems with Applications*, page 113157, 2019.

[59] Y. Liu, M. Guo, J. Zhang, Y. Zhu, and X. Xie. A novel two-stage separable deep learning framework for practical blind watermarking. In *Proc. of the ACM International Conference on Multimedia*, page 1509–1517, 2019.

[60] X. Zhong, P.-C. Huang, S. Mastorakis, and F.Y. Shih. An automated and robust image watermarking scheme based on deep neural networks. *IEEE Transactions on Multimedia*, 23:1951–1961, 2021.

[61] W. Ding, Y. Ming, Z. Cao, and C.-T. Lin. A generalized deep neural network approach for digital watermarking analysis. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(3):613–627, 2022.

[62] Y. Uchida, Y. Nagai, S. Sakazawa, and S. Satoh. Embedding watermarks into deep neural networks. In *Proc. of the ACM on International Conference on Multimedia Retrieval*, 2017.

[63] Y. Nagai, Y. Uchida, S. Sakazawa, and S. Satoh. Digital watermarking for deep neural networks. *International Journal of Multimedia Information Retrieval*, 7(1):3–16, 2018.

[64] J. Zhang, Z. Gu, J. Jang, H. Wu, M.P. Stoecklin, H. Huang, and I. Molloy. Protecting intellectual property of deep neural networks with watermarking. In *Proc. of the Asia Conference on Computer and Communications Security*, page 159–172, 2018.

[65] B.D. Rouhani, H. Chen, and F. Koushanfar. DeepSigns: An end-to-end watermarking framework for ownership protection of deep neural networks. In *Proc. of the International Conference on Architectural Support for Programming Languages and Operating Systems*, page 485–497, 2019.

[66] J. Zhang, D. Chen, J. Liao, W. Zhang, H. Feng, G. Hua, and N. Yu. Deep model intellectual property protection via deep watermarking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4005–4020, 2022.

[67] N. Cvejic and T. Seppanen. Increasing the capacity of LSB-based audio steganography. In *Proc. of the IEEE Workshop on Multimedia Signal Processing*, pages 336–338, 2002.

[68] K. Gopalan. Audio steganography using bit modification. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages II–421, 2003.

[69] P. Bassia, I. Pitas, and N. Nikolaidis. Robust audio watermarking in the time domain. *IEEE Transactions on Multimedia*, 3(2):232–241, 2001.

[70] W.-N. Lie and L.-C. Chang. Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification. *IEEE Transactions on Multimedia*, 8(1):46–59, 2006.

[71] S. Xiang and J. Huang. Histogram-based audio watermarking against time-scale modification and cropping attacks. *IEEE Transactions on Multimedia*, 9(7):1357–1372, 2007.

[72] X. Liang and S. Xiang. Robust reversible audio watermarking based on high-order difference statistics. *Signal Processing*, 173:107584, 2020.

[73] D. Gruhl, A. Lu, and W. Bender. Echo hiding. In *Information Hiding*, pages 295–315. Springer Berlin Heidelberg, 1996.

[74] H.-O Oh, J.-W. Seok, J.-W. Hong, and D.-H. Youn. New echo embedding technique for robust and imperceptible audio watermarking. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1341–1344. IEEE, 2001.

[75] H.-J. Kim and Y.-H. Choi. A novel echo-hiding scheme with backward and forward kernels. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(8):885–889, 2003.

[76] B.-S. Ko, R. Nishimura, and Y. Suzuki. Time-spread echo method for digital audio watermarking. *IEEE Transactions on Multimedia*, 7(2):212–221, 2005.

[77] S. Wang, W. Yuan, and M. Unoki. Multi-subspace echo hiding based on time-frequency similarities of audio signals. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:2349–2363, 2020.

[78] S.-K. Lee and Y.-S. Ho. Digital audio watermarking in the cepstrum domain. *IEEE Transactions on Consumer Electronics*, 46(3):744–750, 2000.

[79] N. Cvejic and T. Seppanen. A wavelet domain LSB insertion algorithm for high capacity audio steganography. In *Proc. of IEEE Digital Signal Processing Workshop*, pages 53–55, 2002.

[80] E. Salah, K. Amine, K. Redouane, and K. Fares. A fourier transform based audio watermarking algorithm. *Applied Acoustics*, 172:107652, 2021.

[81] D. Kirovski and H.S. Malvar. Spread-spectrum watermarking of audio signals. *IEEE Transactions on Signal Processing*, 51(4):1020–1033, 2003.

[82] H.S. Malvar and D.A. Florencio. Improved spread spectrum: a new modulation technique for robust watermarking. *IEEE Transactions on Signal Processing*, 51(4):898–905, 2003.

[83] A. Valizadeh and Z.J. Wang. An improved multiplicative spread spectrum embedding scheme for data hiding. *IEEE Transactions on Information Forensics and Security*, 7(4):1127–1143, 2012.

[84] B. Chen and G.W. Wornell. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. *IEEE Transactions on Information Theory*, 47(4):1423–1443, 2001.

[85] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35(3.4):313–336, 1996.

[86] M. Arnold. Audio watermarking: features, applications and algorithms. In *Proc. of the IEEE International Conference on Multimedia and Expo*, pages 1013–1016, 2000.

[87] I.-K. Yeo and H.-J. Kim. Modified patchwork algorithm: a novel audio watermarking scheme. *IEEE Transactions on Speech and Audio Processing*, 11(4):381–386, 2003.

[88] S. Saadi, A. Merrad, and A. Benziane. Novel secured scheme for blind audio/speech norm-space watermarking by arnold algorithm. *Signal Processing*, 154:74–86, 2019.

[89] W.S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, 1943.

[90] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[91] D. Rumelhart, G.E. Hinton, and R. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[92] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2(4):303–314, 1989.

[93] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.

[94] V. Kurkova. Kolmogorov's theorem and multilayer neural networks. *Neural Networks*, 5(3):501–506, 1992.

[95] D.-X. Zhou. Universality of deep convolutional neural networks. *Applied and Computational Harmonic Analysis*, 48(2):787–794, 2020.

[96] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.

[97] G.W. Lindsay. Convolutional neural networks as a model of the visual system: Past, present, and future. *Journal of Cognitive Neuroscience*, 33(10):2017–2031, 2021.

[98] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[99] A. Krizhevsky, I. Sutskever, and G.E. Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[100] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28, 2015.

[101] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788, 2016.

[102] O. Abdel-Hamid, A. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(10):1533–1545, 2014.

[103] A. Baevski, H. Zhou, A. Mohamed, and M. Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. *Advances in Neural Information Processing Systems*, 33:12449–12460, 2020.

[104] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1904–1916, 2015.

[105] K. Fukushima. Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, 20(3.4):121–136, 1975.

[106] A. Maas, A. Hannun, and A. Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. of the International Conference on Machine Learning*, 2013.

[107] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proc. of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[108] P. Ramachandran, B. Zoph, and Q.V. Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941v2*, 2017.

[109] P. Ramachandran, B. Zoph, and Q.V. Le. Swish: a self-gated activation function. *arXiv preprint arXiv:1710.05941v1*, 2017.

[110] D.R. Wilson and T.R. Martinez. The general inefficiency of batch training for gradient descent learning. *Neural Networks*, 16(10):1429–1451, 2003.

[111] B.T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17, 1964.

[112] Y. Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Proc. of the USSR Academy of Sciences*, 269:543–547, 1983.

[113] I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In *Proc. of the International Conference on Machine Learning*, pages 1139–1147, 2013.

[114] J. Duchi and E. Hazan. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159, 2011.

[115] T. Tieleman and G.E. Hinton. Lecture 6.5—rmsprop: Divide the gradient by a running average of its recent magnitude. *Coursera: Neural networks for machine learning*, 4:26–31, 2012.

[116] D.P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[117] T. Dozat. Incorporating Nesterov momentum into Adam. In *Proc. of the International Conference on Learning Representations*, 2016.

[118] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proc. of the International Conference on Artificial Intelligence and Statistics*, pages 249–256, 2010.

[119] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.

[120] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proc. of the International Conference on Machine Learning*, pages 448–456, 2015.

[121] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry. How does batch normalization help optimization? In *Proc. of the International Conference on Neural Information Processing Systems*, page 2488–2498, 2018.

[122] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[123] A. Halevy, P. Norvig, and F. Pereira. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2):8–12, 2009.

[124] S. Stevens, J. Volkmann, and E. Newman. A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, 8(3):185–190, 1937.

[125] B. Bogert, M.R. Healy, and J. Tukey. The quefrency analysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking. In *Proc. of the Symposium on Time Series Analysis*, pages 209–243, 1963.

[126] P. Mermelstein. Distance measures for speech recognition, psychological and instrumental. *Pattern Recognition and Artificial Intelligence*, 116:374–388, 1976.

[127] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely. The Kaldi speech recognition toolkit. In *Proc. of the IEEE Workshop on Automatic Speech Recognition and Understanding*, 2011.

[128] A. Radford, J.-W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever. Robust speech recognition via large-scale weak supervision. In *Proc. of the International Conference on Machine Learning*, pages 28492–28518. PMLR, 2023.

[129] D. Reynolds, T. Quatieri, and R. Dunn. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing*, 10(1):19–41, 2000.

[130] J. Salamon and J.P. Bello. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3):279–283, 2017.

[131] M. Deng, T. Meng, J. Cao, S. Wang, J. Zhang, and H. Fan. Heart sound classification based on improved mfcc features and convolutional recurrent neural networks. *Neural Networks*, 130:22–32, 2020.

[132] K. Choi, G. Fazekas, M. Sandler, and K. Cho. Convolutional recurrent neural networks for music classification. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2392–2396, 2017.

[133] L. Muda, M. Begam, and I. Elamvazuthi. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *arXiv preprint arXiv:1003.4083*, 2010.

[134] X.-C. Yuan, C.-M. Pun, and C.L.P. Chen. Robust mel-frequency cepstral coefficients feature detection and dual-tree complex wavelet transform for digital audio watermarking. *Information Sciences*, 298:159–179, 2015.

[135] K. Pavlović, S. Kovačević, and I. Djurović. Speech watermarking using deep neural networks. In *Proc. of the Telecommunications Forum*, pages 1–4, 2020.

[136] K. Pavlović, S. Kovačević, I. Djurović, and A. Wojciechowski. Robust speech watermarking by a jointly trained embedder and detector using a DNN. *Digital Signal Processing*, 122:103381, 2021.

[137] K. Pavlović, S. Kovačević, I Djurović, and A. Wojciechowski. Dnn-based speech watermarking resistant to desynchronization attacks. *International Journal of Wavelets, Multiresolution and Information Processing*, page 2350009, 2022.

[138] S. Kovačević, K Pavlović, and I. Djurović. A novel DNN-based image watermarking algorithm. *Proc. in Polish Artificial Intelligence Research 4, Seria: Monografie Politechniki Łódzkiej Nr. 2437*, 2023.

[139] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[140] P. Isola, J.-Y. Zhu, T. Zhou, and A.A. Efros. Image-to-image translation with conditional adversarial networks. In *Proc. of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1125–1134, 2017.

[141] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen. PointCNN: Convolution on X-transformed points. *Advances in Neural Information Processing Systems*, 31, 2018.

[142] Y. Tian, D. Krishnan, and P. Isola. Contrastive multiview coding. In *Proc. of the European Conference on Computer Vision*, pages 776–794, 2020.

[143] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[144] C. Saharia, W. Chan, S. Saxena, L. Li, J. Whang, E.L. Denton, K. Ghasemipour, R.G. Lopes, B.K. Ayan, T. Salimans, J. Ho, D.J. Fleet, and M. Norouzi. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in Neural Information Processing Systems*, 35:36479–36494, 2022.

[145] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proc. of the IEEE-/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

[146] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[147] M. Steinebach, F.A.P. Petitcolas, F. Raynal, J. Dittmann, C. Fontaine, C. Seibel, N. Fates, and L. Ferri. Stirmark benchmark: audio watermarking attacks. In *Proc. of the International Conference on Information Technology: Coding and Computing*, pages 49 – 54, 2001.

[148] F. Hartung, J. Su, and B. Girod. Spread spectrum watermarking: Malicious attacks and counterattacks. In *Proc. of the Security and Watermarking of Multimedia Contents*, pages 147–158, 1999.

[149] S. Voloshynovskiy, S. Pereira, T. Pun, J. Eggers, and J. Su. Attacks on digital watermarks: classification, estimation based attacks, and benchmarks. *IEEE communications Magazine*, 39(8):118–126, 2001.

[150] A. Lang, J. Dittmann, R. Spring, and C. Vielhauer. Audio watermark attacks: from single to profile attacks. In *Proc. of the Workshop on Multimedia and Security*, pages 39–50, 2005.

[151] M. Tanha, S.D.S. Torshizi, M. Taufik, and F.H. Abdullah. An overview of attacks against digital watermarking and their respective countermeasures. In *Proc. of the International Conference on Cyber Security, Cyber Warfare and Digital Forensic*, pages 265–270, 2012.

[152] P. Singh and R.S. Chadha. A survey of digital watermarking techniques, applications and attacks. *International Journal of Engineering and Innovative Technology*, 2(9):165–175, 2013.

[153] D. Kirovski, F.A.P. Petitcolas, and Z. Landau. The replacement attack. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(6):1922–1931, 2007.

[154] A. Robert and J. Picard. On the use of masking models for image and audio watermarking. *IEEE Transactions on Multimedia*, 7(4):727–739, 2005.

[155] A. Krizhevsky and G. Hinton. Learning multiple layers of features from tiny images. *Master's thesis, Department of Computer Science, University of Toronto*, 2009.

[156] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[157] M. Marcus, M.A. Marcinkiewicz, and B. Santorini. Building a large annotated corpus of english: The penn treebank. *Computational Linguistcs*, 19(2):313–330, 1993.

[158] J. Salamon, C. Jacoby, and J.P. Bello. A dataset and taxonomy for urban sound research. In *Proc. of the ACM International Conference on Multimedia*, pages 1041–1044, 2014.

[159] K.J. Piczak. ESC: Dataset for environmental sound classification. In *Proc. of the ACM International Conference on Multimedia*, pages 1015–1018, 2015.

[160] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N.L. Dahlgren, and V. Zue. TIMIT acoustic-phonetic continuous speech corpus LDC93S1, 1993.

[161] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur. Librispeech: An ASR corpus based on public domain audio books. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5206–5210, 2015.

[162] The European Broadcasting Union. Sound quality assessment material recordings for subjective tests, 2008.

[163] T. Baer, B. Moore, and K. Kluk. Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. *The Journal of the Acoustical Society of America*, 112:1133–44, 2002.

[164] B. Hornsby and T. Ricketts. The effects of hearing loss on the contribution of high- and low-frequency speech information to speech understanding. *The Journal of the Acoustical Society of America*, 113:1706–17, 2003.

[165] C. Shannon. Communication in the presence of noise. *Proc. of the Institute of Radio Engineers*, 37(1):10–21, 1949.

[166] G. Hua, J. Huang, Y.-Q. Shi, J. Goh, and V.L.L. Thing. Twenty years of digital audio watermarking - a comprehensive review. *Signal Processing*, 128:222–242, 2016.

[167] J.W. Cooley and J.W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

[168] S.G. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989.

[169] G.H. Golub and C. Reinsch. Singular value decomposition and least squares solutions. In *Handbook for Automatic Computation: Volume II: Linear Algebra*, pages 134–151. Springer, 1971.

[170] R. Crochiere. A weighted overlap-add method of short-time Fourier analysis/synthesis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(1):99–102, 1980.

[171] A. Oppenheim. *Discrete-time signal processing*. Prentice Hall, 1999.

[172] J. Flanagan and R. Golden. Phase vocoder. *The Bell System Technical Journal*, 45(9):1493–1509, 1966.

[173] L. Rabiner and R. Schafer. *Digital Processing of Speech Signals*. Prentice Hall, 1978.

[174] D. Malah. Time-domain algorithms for harmonic bandwidth reduction and time scaling of speech signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 27(2):121–133, 1979.

[175] F. Charpentier and M. Stella. Diphone synthesis using an overlap-add technique for speech waveforms concatenation. In *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2015–2018, 1986.

A Diskretna Furijeova transformacija

Fourijeova transformacija je matematički koncept koji se koristi za analizu periodičnih signala. Ona dekomponuje signal u frekvencijskom domenu, odnosno izdvaja komponente različitih frekvencija koje su prisutne u signalu. Za digitalni signal x u vremenskom domenu koristi se diskretna Furijeova transformacija (DFT), koja se izračunava na sljedeći način:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}, \tag{93}$$

gdje $k \in \{0, ..., K-1\}$, a N je broj odbiraka posmatranog signala [3]. Vrijednosti $X(k)$ su kompleksni koeficijenti koji opisuju amplitudu i fazu sinusnih i kosinusnih komponenti različitih frekvencija koje čine originalni signal.

Odbirci signala x mogu se rekonstruisati iz K odbiraka DFT, inverznom diskretnom Furijeovom transformacijom (IDFT).

$$x(n) = \frac{1}{K} \sum_{k=0}^{K-1} X(k) e^{j2\pi kn/N}. \tag{94}$$

[3] Simbol j predstavlja imaginarnu jedinicu.

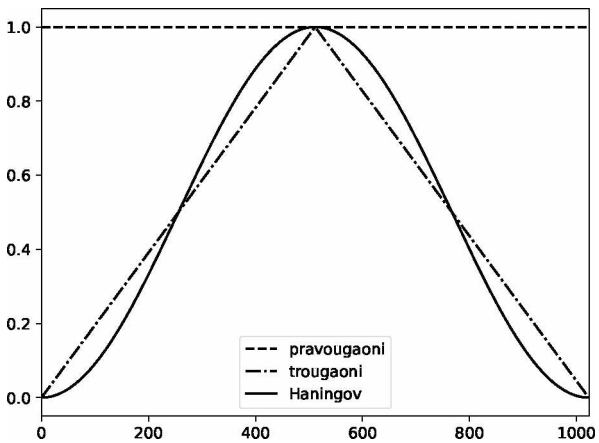
B Kratkotrajna Furijeova transformacija

Primjenom DFT dobija se spektralni sadržaj cjelokupnog signala, ali DFT ne pruža informacije o promjenama u frekvencijskim komponentama signala tokom vremena. STFT je proširenje DFT koje proizvodi vremensko-frekvencijsku reprezentaciju signala, odnosno vremenski lokalizuje pojavu frekvencijskih komponenti. STFT se računa primjenom diskretne Furijeove transformacije iz jednakosti (93) na kratke djelove signala, lokalizovane prozorskom funkcijom ξ:

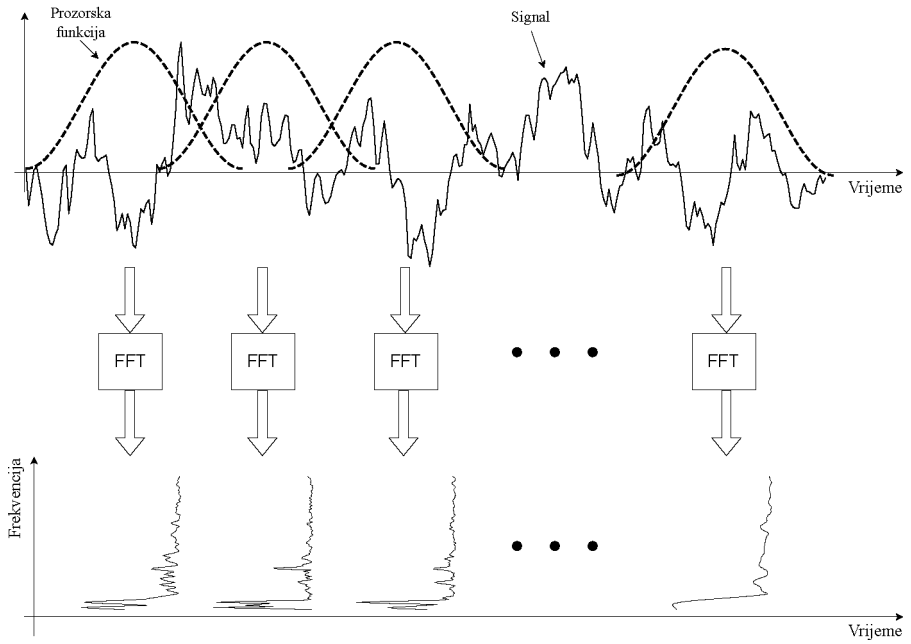
$$\mathrm{STFT}\,(n,k) = e^{-j2\pi nk/L_\xi} \sum_{m=-L_\xi/2}^{L_\xi/2-1} \xi\,(m)\,x\,(n-m)\,e^{j2\pi mk/L_\xi}, \tag{95}$$

gdje je L_ξ broj tačaka u kojima funkcija ξ ima nenulte vrijednosti, odnosno tzv. širina prozora. Pomjeranjem odabranog prozora signalom dodaje se vremenska dimenzija ovoj reprezentaciji, odnosno dobija se spektralni sadržaj u različitim djelovima signala.

U računanju STFT mogu se koristiti različite prozorske funkcije. Predloženo je mnoštvo ovih funkcija, a grafici nekolicine predstavnika su prikazani na Slici 27. Neki prozori su izuzetno jednostavni, poput pravougaonog ili trougaonog (Bartletovog) prozora, ali njihova primjena izaziva pojavu artefakata ili razlivanja u spektru, što vodi ka lošijoj estimaciji frekvencijskih komponenti. Kompleksnijim (glatkijim) prozorima se poboljšava lokalizacija frekvencijskih komponenti. U ovom radu je korišten



Slika 27: Različite prozorske funkcije.



Slika 28: Ilustracija izračunavanja kratkotrajne Furijeove transformacije.

Haningov prozor. U diskretnom domenu on je definisan sa:

$$\xi(n) = \begin{cases} 0.5\left(1+\cos\left(\frac{2\pi n}{L_\xi}\right)\right), & -L_\xi/2 \leq n < L_\xi/2 \\ 0, & n < -L_\xi/2 \vee n \geq L_\xi/2. \end{cases} \tag{96}$$

Vrijednost parametra L_ξ, odnosno širina prozora je od presudne važnosti za preciznost informacija koje daje STFT. Široki prozori obuhvataju veće djelove signala, pa se tada sa većom preciznošću mogu izračunati frekvencijske komponente koje se javljaju u datom dijelu signala. Međutim, tačan trenutak pojave tih komponenti se, zbog „razmazivanja" prozora na većem vremenskom intervalu, ne može precizno procijeniti. Nasuprot tome, upotrebom užeg prozora dobija se više detalja po vremenskoj osi, odnosno imamo veću vremensku rezoluciju, ali to dovodi do manje frekvencijske rezolucije. U suštini, izbor širine prozora predstavlja kompromis između vremenske i frekvencijske rezolucije i zavisi od specifičnih potreba analize signala. U ovom radu je empirijskim postupkom utvrđena optimalna širina prozora $L_\xi = 1024$. Prozor je signalom pomjeran sa korakom od 512 odbiraka. Postupak izračunavanja kratkotrajne Furijeove transformacije ilustrovan je na Slici 28.

Činjenica da je vrijednost $\mathrm{STFT}\,(n,k)$ zapravo DFT sekvence $\xi\,(m)\,x\,(n-m)$, implicira da se ta sekvenca može rekonstruisati iz STFT reprezentacije primjenom inverzne DFT. Preklapanjem i sabiranjem djelova ove sekvence tzv. *overlap-add* metodom [170] sintetiše se originalni audio signal $x(n)$.

C Konvolucija

Konvolucija je matematička operacija koja se često koristi u obradi signala. Sprovodi se nad dvijema funkcijama f i g (dva signala ili dva skupa informacija) i kao rezultat daje funkciju koja odražava na koji način oblik jedne funkcije modfikuje drugu. U diskretnom domenu konvolucija je suma:

$$(f * g)(n) = \sum_{k=-\infty}^{k=+\infty} f(n-k)g(k). \tag{97}$$

Digitalni signali imaju konačan skup odbiraka, pa se suma iz prethodne jednakosti može se ograničiti:

$$(f * g)(n) = \sum_{k=0}^{k=K} f(n-k)g(k). \tag{98}$$

Konvolucija se može definisati i u prostorima sa više dimenzija. Često se sreće konvolucija u dvodimenzionom prostoru, jer se ona koristi za slike [54, 55, 57, 58, 60, 61, 138], spektrograme audio signala [135] i sl. 2D konvolucija definisana je na sljedeći način:

$$(f * g)(n_1, n_2) = \sum_{k_1=0}^{k_1=K_1} \sum_{k_2=0}^{k_2=K_2} f(n_1 - k_1, n_2 - k_2)g(k_1, k_2) \tag{99}$$

U kontekstu konvolucionih slojeva neuronskih mreža f je ulaz konvolucionog sloja, a g je filtar, odnosno skup težinskih koeficijenata. U kontekstu eho kernel votermarking metoda f predstavlja audio signal, a g je eho kernel.

D Audio efekti

Ovaj prilog sadrži detaljan pregled i opis audio efekata koji su primjenjivani tokom obučavanja i testiranja sistema vodenog žiga u okviru ove studije.

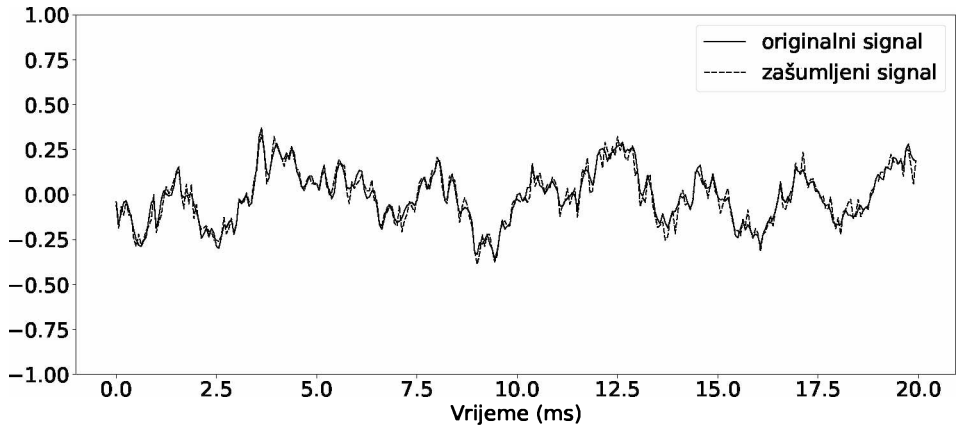
D.1 Aditivni šum

Postoje različiti modeli kojima se može simulirati šum, a najjednostavniji i najopštiji je aditivni model predstavljen u jednakosti (8). Modeli šuma se koriste za razvijanje i testiranje različitih tehnika obrade signala namijenjenih analizi, otklanjanju ili smanjenju šuma. Za potrebe ovog istraživanja, izabran je model šuma iz [147]:

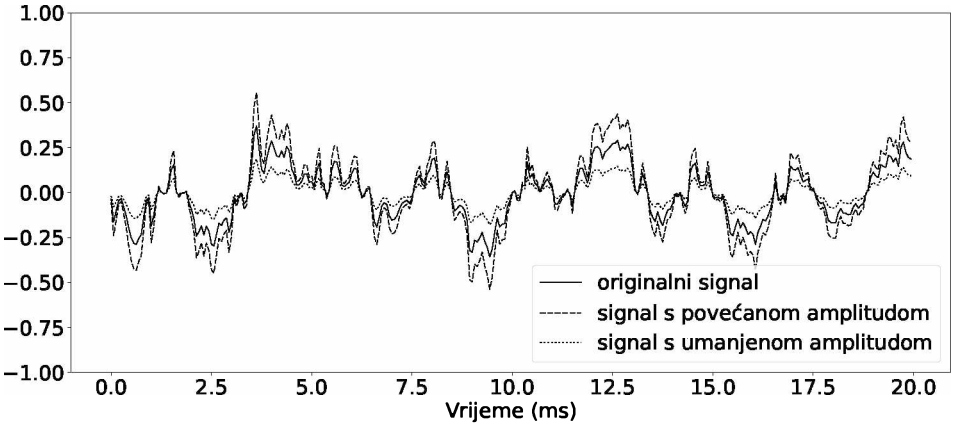
$$z(n) = x(n) + a_\epsilon \epsilon(n), \tag{100}$$

gdje su x, z i ϵ, originalni signal, zašumljeni signal i šum, respektivno, a a_ϵ je parametar kojim se kontroliše snaga šuma. Odbirci šuma slučajno se uzorkuju iz normalne (Gausove) raspodjele $\mathcal{N}(0, 1)$. Vrijednost parametra a_ϵ postavljena je na 0.02. Nakon dodavanja šuma sa ovim karakteristikama, dobijeni nivo odnosa signal-šum u rezultujućem signalu je oko 23 dB, a vrijednost PESQ mjere je iznad 3, što govori da je signal značajno degradiran ali da su informacije koje on nosi sačuvane. Izmjene koje unosi ovaj tip šuma na kratkom isječku signala prikazane su na Slici 29.

Votermarkovani signali mogu biti pogođeni šumom tokom prenosa komunikacionim kanalom ili usljed namjernog dodavanja šuma od strane napadača radi ometanja



Slika 29: Prikaz isječka govornog signala prije i nakon primjene aditivnog Gausovog šuma.



Slika 30: Isječak govornog signala prije i nakon skaliranja amplitude sa faktorom 1.5 i faktorom 0.5.

detekcije. Aditivnim šumom se mogu modelovati i pojednostavljeni oblici različitih efekata koji se koriste za obradu signala, uključujući i votermarkovane. U tu grupu spadaju efekti poput kvantizacije, kao i pojedine metode za poboljšanje kvaliteta audio signala.

D.2 Skaliranje amplitude

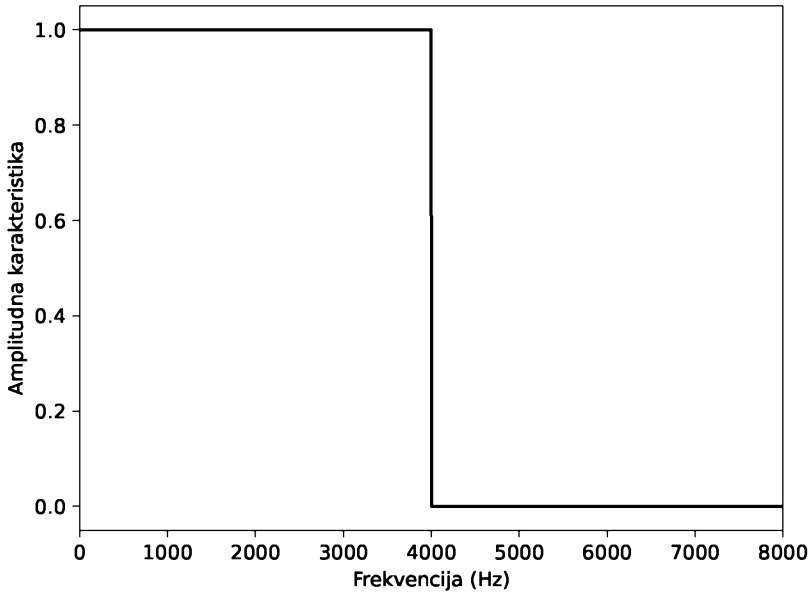
Skaliranje amplitude realizovano je prema modelu iz jednakosti (9), sa slučajno uzorkovanim faktorom skaliranja a iz segmenta $[0.5, 1.5]$. Primjeri signala sa skaliranim amplitudama prikazani su na Slici 30. Skaliranje amplitude ne utiče na PESQ vrijednosti, pošto je prvi korak u izračunavanju te mjere skaliranje oba signala na standardni nivo za slušanje. Međutim, SNR vrijednosti padaju čak do 6 dB kada se izabere jedan od faktora skaliranja sa krajeva datog segmenta.

D.3 Niskopropusni filtri

Ključni parametar pri definisanju niskopropusnih filtara je granična frekvencija (*engl. cutoff frequency*). To je vrijednost frekvencije ispod koje filtar dozvoljava prolazak komponenti signala bez značajnog prigušenja, dok se frekvencijske komponente signala koje su iznad tog praga frekvencije prigušuju.

D.3.1 Idealni filtar

Idealni niskopropusni filtar propušta sve komponente ispod granične frekvencije, dok ostale komponente u potpunosti eliminiše. Amplitudna karakteristika idealnog



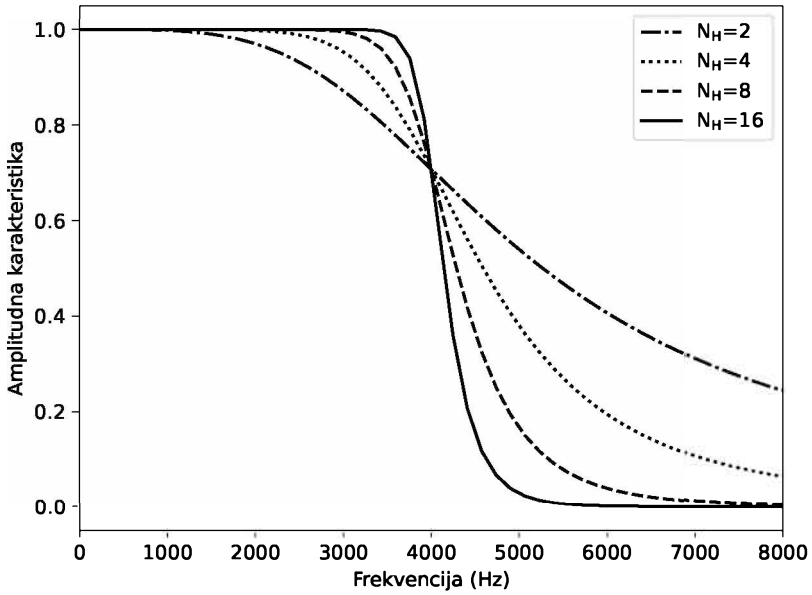
Slika 31: Amplitudna karakteristika idealnog filtra sa graničnom frekvencijom 4 kHz.

filtra ima vrijednost 1 za frekvencije koje filtar propušta, a vrijednost 0 za frekvencije koje se ne propuštaju:

$$|H(j\omega)| = \begin{cases} 1, & \omega \leq 2\pi f_c \\ 0, & \omega > 2\pi f_c, \end{cases} \tag{101}$$

gdje je f_c granična frekvencija data u hercima (Hz). Grafik amplitudne karakteristike idealnog niskopropusnog filtra, sa graničnom frekvencijom 4 kHz, prikazan je na Slici 31.

Idealni filtar se u računaru može jednostavno realizovati transformacijom signala u frekvencijski domen i postavljanjem odgovarajućih odbiraka transformacije na 0, a zadržavanjem ostalih. Ta realizacija uključena je u slojeve za aproksimaciju napada sistema vodenog žiga. Međutim, neka svojstva idealnog filtra onemogućavaju njegovu primjenu u realnom vremenu. Pomenuta svojstva uključuju beskonačno oštar prelaz između propusnog i nepropusnog opsega frekvencija, kao i beskonačno dug impulsni odziv. Stoga je neophodno kreirati filtre koji se u realnom vremenu mogu primjenjivati nad signalom. U praksi su dizajnirani brojni filtri ovog tipa, s težnjom da se približe idealnim performansama. Kod svakog od njih postoje odstupanja od idealnog, poput određenog nivoa prigušivanja u propusnom opsegu ili postojanja frekvencijskih komponenti male amplitude u nepropusnom opsegu.



Slika 32: Amplitudne karakteristike Batervortovih filtara različitog reda, sa graničnom frekvencijom 4 kHz.

D.3.2 Batervortov filtar

S obzirom na realnu mogućnost da signal prođe kroz razne niskopropusne filtre na svom putu do detektora vodenog žiga, procijenjeno je da je u sistem potrebno uključiti i jednog predstavnika grupe filtara koji se mogu primijeniti u realnom vremenu. Za ovu namjenu, odabran je Batervortov filtar. Batervotrov filtar ima sljedeću amplitudnu karakteristiku:

$$|H(j\omega)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_c}\right)^{2N_H}}}, \tag{102}$$

gdje je $\omega_c = 2\pi f_c$ granična ugaona frekvencija, a N_H je red filtra. Ovo su dva definišuća parametra Batervortovog filtra. Na Slici 32 prikazane su amplitudne karakteristike Batervortovih filtara različitog reda. Sa slike se može vidjeti da se, s povećanjem reda filtra, njegov frekvencijski odziv približava idealnom. Kada $N_H \to \infty$, Batervortov filtar dostiže idealan frekvencijski odziv.

U sistem neuronskih mreža za umetanje i detekciju vodenog žiga, uključen je sloj koji simulira Batervortov filtar reda $N_H = 16$. Ovaj napad realizovan kao konvolucioni sloj sa predefinisanim skupom parametara. Vrijednosti ovih parametara predstavljaju odbirke impulsnog odziva filtra. Impulsni odziv se konvoluira sa ulaznim signalom kako bi se dobio izlaz filtra, pa je izbor vrste sloja očekivan.

Impulsni odziv Batervortovog filtra ima sljedeći oblik:

$$h(t) = \sum_{k=1}^{N_H} r_k e^{s_k t} u(t), \tag{103}$$

gdje su s_k polovi, odnosno korijeni imenioca funkcije prenosa Batervortovog filtra $H(s)$. Funkcija prenosa Batervortovog filtra, definisana je u domenu Laplasove transformacije:

$$H(s) = \prod_{k=1}^{N_H} \frac{1}{(s - s_k)}. \tag{104}$$

Laplasova transformacija je uopštenje Furijeove transformacije gdje je $s = \sigma + j\omega$ kompleksna frekvencija.

Koeficijenti r_k iz jednakosti (103) su koeficijenti dobijeni parcijalnom dekompozicijom funkcije prenosa, odnosno njenim predstavljanjem u sljedećem obliku:

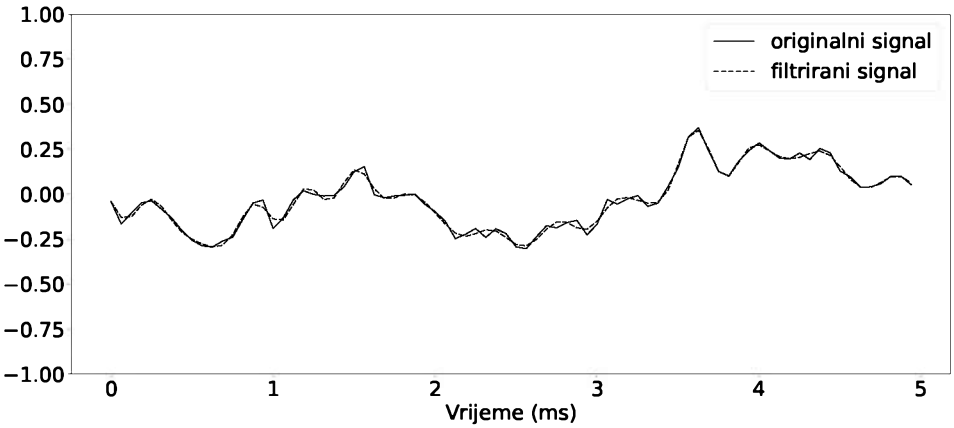
$$H(s) = \sum_{k=1}^{N_H} \frac{r_k}{s - s_k}. \tag{105}$$

Izlazi idealnog i Batervortovog filtra ne mogu se jasno razlikovati na grafiku. Zbog toga su na Slici 33 prikazane promjene koje u signal uvodi samo jedan od ovih efekata. Primjena niskopropusnog filtra na govornom signalu, sa graničnom frekvencijom od 4 kHz gotovo ne utiče na vrijednost PESQ mjere, dok SNR pada na 15dB. Stoga je za očekivati da votermarking tehnike koje ne koriste visoke frekvencije za ugrađivanje bitova budu otporne na ove napade, jer je sadržaj skoro u potpunosti sačuvan. Međutim, primjena Batervotovog filtra nekada uvodi, uglavnom nepoželjni, fazni pomjeraj u izlazni signal. S povećanjem reda filtra i smanjenjem granične frekvencije, ovaj pomjeraj se povećava. Na Slici 33 je prikazan filtrirani signal bez faznog pomjeraja. Ovaj fazni pomjeraj se može izbjeći samo ukoliko je kompletan signal poznat unaprijed, što u nekim situacijama nije moguće. Kada se ovaj pomjeraj desi, Batervortov filtar postaje znatno izazovniji napad za sisteme vodenog žiga, jer se tada može svrstati u kategoriju efekata desinhronizacije.

D.4 Efekti desinhronizacije

U ovoj sekciji izloženo je pet efekata desinhronizacije koji su korišćeni prilikom testiranja predloženog sistema vodenog žiga, kao i uporednih tehnika.

Vrijednosti parametara kojima se određuje razornost efekta se slučajno uzorkuju iz predefinisanog opsega. Granice ovih opsega su birane tako da ne dođe do drastičnog smanjenja PESQ/PEAQ vrijednosti nakon desinhronizacije. Za govorne



Slika 33: Isječak govornog signala prije i nakon primjene niskopropusnog filtra sa graničnom frekvencijom od 4 kHz, bez faznog pomjeraja.

signale PESQ iznad vrijednosti 3 ukazuje na očuvanje informacija koje signal nosi. Nasuprot tome, mimoilaženje odbiraka originalnog i rezultujućeg signala, izazvano ovim efektima, u potpunosti remeti izračunavanje SNR vrijednosti, koja čak i za neke suptilne efekte uzima negativne vrijednosti.

D.4.1 Brisanje odbiraka

Ovim napadom se nasumično odabran skup uzoraka uklanja iz audio signala. Može se predstaviti sljedećom jednakošću:

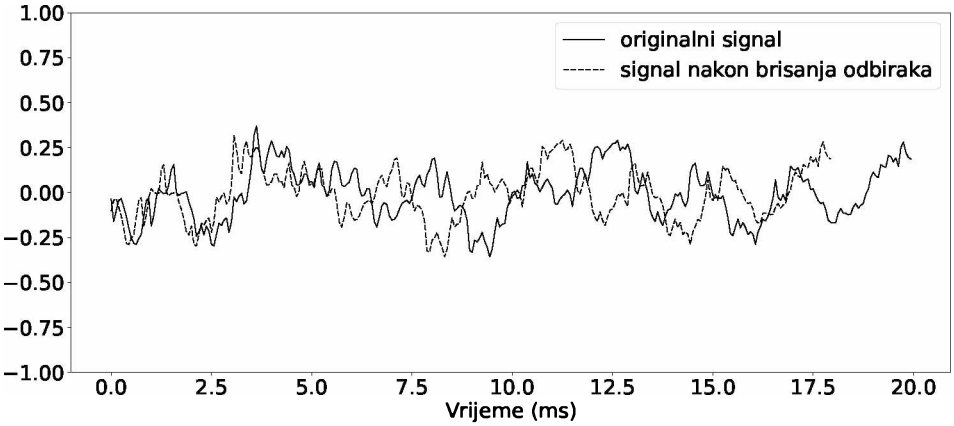
$$z(n) = x(n_m), \tag{106}$$

gdje je $x(n_m), n_m \in \mathcal{S}_B \subset \{0, 1, ..., N-1\}$ podsekvenca ulaznog signala $x(n)$. $\mathcal{S}_B$ je skup indeksa odbiraka koje treba sačuvati. Destruktivnost ovog napada može se regulisati promjenom kardinalnosti skupa $\mathcal{S}_B$. U predloženi sistem vodenog žiga dodat je efekat koji briše do 10% odbiraka signala. Slika 34 ilustruje primjenu ovog efekta na govornom signalu.

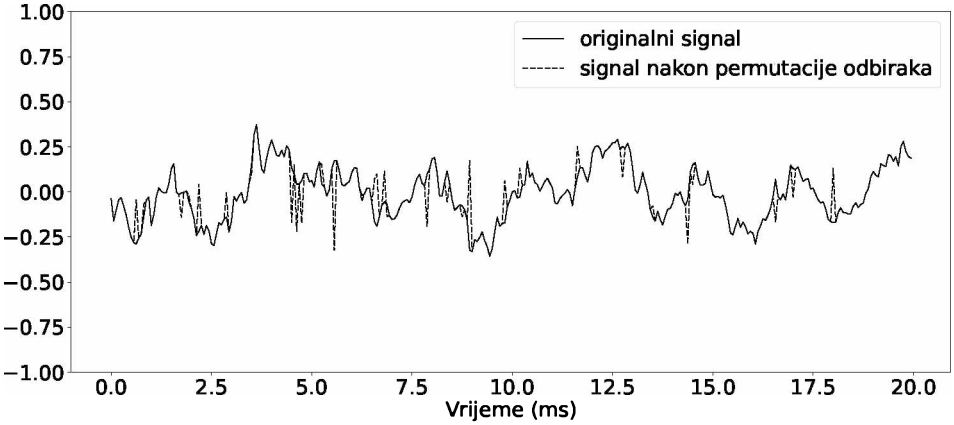
D.4.2 Permutacija odbiraka

Ovim napadom se slučajno odabranoj grupi odbiraka mijenjaju pozicije u signalu. Napad ima samo jedan hiperparametar, broj odbiraka koji se biraju za permutaciju. Može se predstaviti sljedećom jednakošću:

$$z(n) = \begin{cases} x(n), & n \notin \mathcal{S}_P \\ \mathcal{P}\left(x(n)\right), & n \in \mathcal{S}_P, \end{cases} \tag{107}$$



Slika 34: Isječak govornog signala prije i nakon brisanja 10% odbiraka.



Slika 35: Isječak govornog signala prije i nakon permutacije 10% odbiraka.

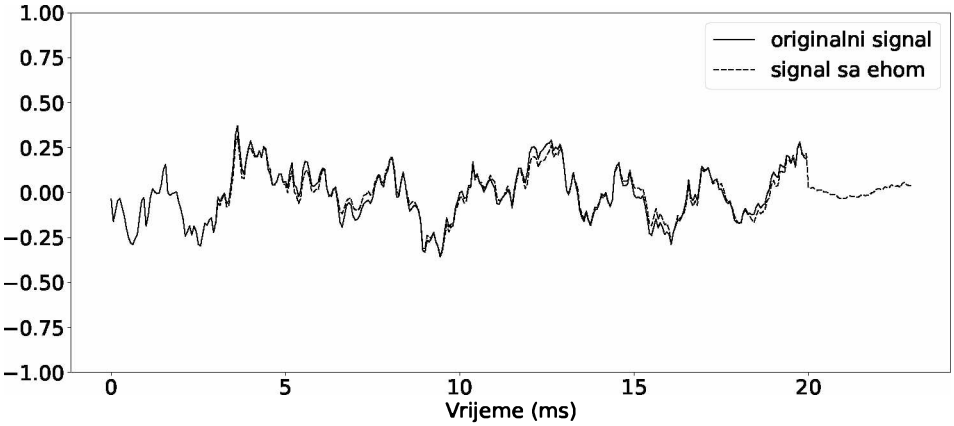
gdje je $\mathcal{S}_P \subset \{0, 1, ..., N-1\}$ skup odbiraka koji se permutuju, $\mathcal{P}$ je funkcija permutacije. Maksimalan procenat permutovanih odbiraka, razmatran u ovom radu, je 10%. Rezultat primjene ovog efekta prikazan je na Slici 35.

D.4.3 Pomjeranje u vremenu

Pomjeranje u vremenu je jedan od osnovnih audio efekata, ali se takođe može smatrati napadom desinhronizacije. Ovim efektom signal može biti dodato kašnjenje ili pomjeren unaprijed po vremenskoj osi sljedećim pravilom:

$$z(n) = \begin{cases} x(n \pm d), & 0 \leq n \pm d < N \\ 0, & \text{inače,} \end{cases} \tag{108}$$

gdje N dužina ulaznog signala, a d je faktor pomjeranja signala, odnosno broj odbiraka za koji je signal pomjeren.



Slika 36: Isječak govornog signala prije i nakon dodavanja eha sa kašnjenjem od 15% trajanja originalnog signala.

Uvođenje kašnjenja uključuje dodavanje nula na početku signala. Pomjeranje unaprijed sastoji se iz brisanja jednog broja odbiraka sa početka signala i popunjavanja ostatka signala nulama. Aspekat brisanja odbiraka pokriven je napadom iz Sekcije D.4.1. Dodavanje nula na početku ili kraju signala obuhvaćeno je opštijim audio efektom, a to je dodavanje odjeka (eha) signalu, što je prikazano na Slici 36. Ovaj efekat može se predstaviti sa:

$$z(n) = \begin{cases} x(n), & n < d_e \\ x(n) + a_e x(n - d_e), & d_e \leq n < N, \end{cases} \tag{109}$$

gdje je d_e kašnjenje odjeka, a_e je parametar kojim se kontoliše jačina eha. Njegova vrijednost postavljena je na 0.2 kako bi se kreirao realističan eho efekat. Najveće unešeno kašnjenje eha iznosi 15% trajanja originalnog signala. Očekuje se da sistem vodenog žiga, otporan na brisanje odbiraka i dodavanje eha, bude takođe otporan na pomjeranje u vremenu.

D.4.4 Ponovno uzorkovanje

Ponovnim uzorkovanjem se mijenja frekvencija odabiranja signala, odnosno broj odbiraka signala u jedinici vremena. Na taj način dolazi do neslaganja originalnog i modifikovanog signala. Ovaj proces može rezultovati pomeranjem ili promjenom položaja uzoraka, što otežava usklađivanje vodenog žiga s originalnim signalom tokom faze detekcije. To ovu vrstu operacije nad signalom kvalifikuje kao efekat desinhronizacije. Postoje dvije vrste ponovnog uzorkovanja, decimacija i interpolacija. Decimacijom se smanjuje frekvencija odabiranja, odnosno broj odbiraka signala. Ova operacija se često koristi u različitim audio sistemima za smanjenje propusnog

opsega i kompresiju signala. S obzirom na to da je ova operacija tako uobičajena, važno je da sistemi audio vodenih žigova budu otporni na nju. Može se predstaviti sa:

$$z(n) = x(nq_d), \tag{110}$$

gdje je q_d faktor decimacije. Ovaj način decimacije bi mogao da dovede do pojave alijasing efekta ukoliko je $\omega_N \geq \pi/q_d$, gdje je ω_N normalizovana Nikvistova frekvencija signala. Alijasing je efekat koji se može desiti kao posljedica obrade signala kada dođe do preklapanja više frekvencijskih komponenti originalnog signala i one postanu nerazlučive. S obzirom na to da u ovom radu koristimo govorne signale koji imaju relativno nisku frekvenciju oko 4 kHz, možemo izvršiti decimaciju na 50% originalne stope uzorkovanja od 16 kHz, bez kreiranja alijasing artefakata.

Realizovana je i interpolacija signala koja povećava frekvenciju odabiranja za faktor q_u. Ovaj napad se sastoji iz dva koraka. Najprije se od originalnog signala x kreira prošireni signal x_e, dodavanjem $q_u - 1$ nula između svaka dva odbirka. Ovaj korak može se predstaviti na sljedeći način:

$$x_e(n) = \begin{cases} x(n/q_u), & n = 0, q_u, 2q_u, ... \\ 0, & \text{inače,} \end{cases} \tag{111}$$

ili ekvivalentno:

$$x_e(n) = \sum_{m=0}^{N-1} x(m)\delta(n - mq_u), \tag{112}$$

gdje je δ jedinična delta funkcija.

Operacija proširivanja signala ima veoma jednostavan oblik u frekvencijskom domenu [171]. Primjenom diskretne Furijeove transformacije na jednakost (112) dobija se:

$$\begin{aligned} X_e(k) &= \sum_{n=0}^{N-1} x_e(n) e^{\frac{-j2\pi}{N}kn} \\ &= \sum_{n=0}^{N-1} \left(\sum_{m=0}^{N-1} x(m)\delta(n - mq_u) \right) e^{\frac{-j2\pi}{N}kn} \\ &= \sum_{m=0}^{N-1} x(m) e^{\frac{-j2\pi}{N}kmq_u} = X(kq_u). \end{aligned} \tag{113}$$

Dakle, DFT proširenog signala je frekvencijski skalirana DFT originalnog signala.

U drugom koraku je potrebno izvršiti interpolaciju odbiraka proširenog signala. Postoji veliki broj metoda za interpolaciju. U našem slučaju, linearna interpolacija

je adekvatna. Ona se vrši primjenom linearnog interpolacionog filtra, čiji je impulsni odziv:

$$h_{lin} = \begin{cases} 1 - \frac{|m|}{q_u}, & |m| \leq q_u \\ 0, & \text{inače.} \end{cases} \tag{114}$$

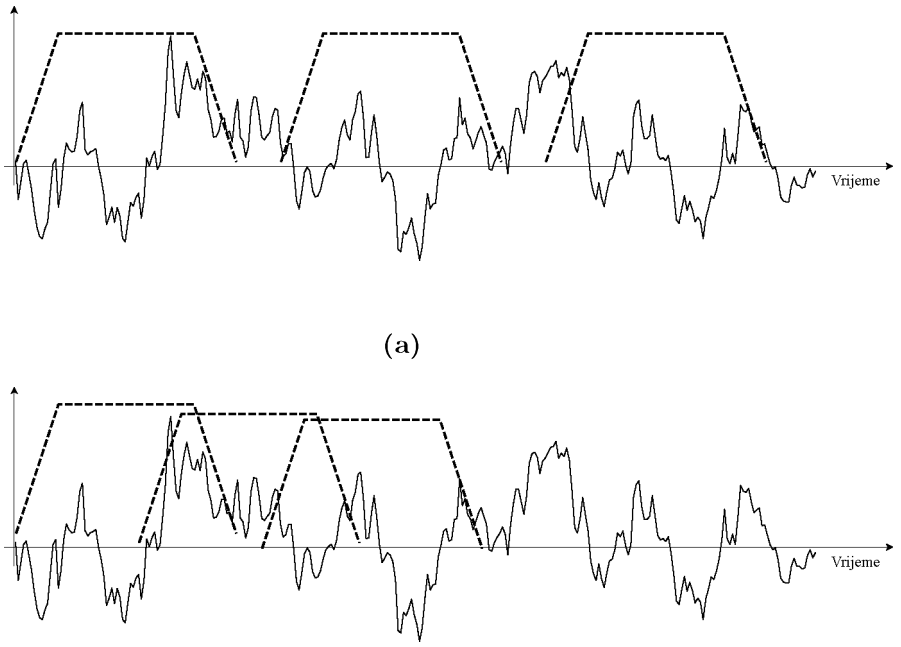
Interpolirani signal dobija se konvolucijom x_e i h_{lin}.

D.4.5 Skaliranje vremena

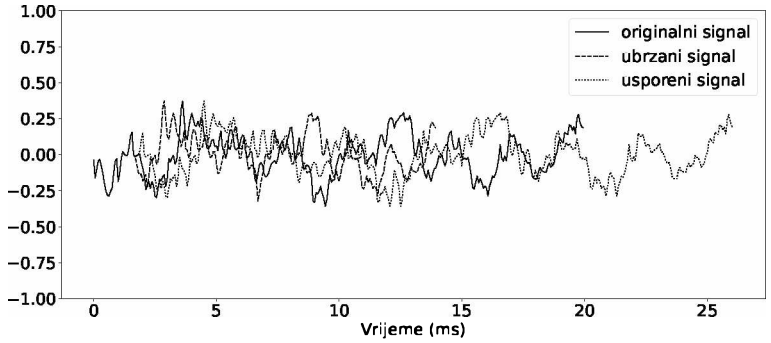
Usporena ili ubrzana reprodukcija audio snimka potrebna je u različitim situacijama. Na primjer, usporavanje snimka govora olakšava njegovu transkripciju. Nasuprot tome, reprodukcija se može ubrzati, ako je potrebno uštedjeti vrijeme. Takođe, promjena stope frejmova (*engl. frame rate*) u reprodukciji videa iziskuje prilagođavanje trajanja pratećeg audio snimka.

Efekat kojim se mijenja trajanje reprodukcije audio snimka naziva se skaliranje vremena (*engl. time scaling*). Skaliranje vremena se može realizovati ponovnim uzorkovanjem, praćenim reprodukcijom snimka sa originalnom frekvencijom odabiranja. Međutim, ponovno uzorkovanje kao nuspojavu ima promjenu visine tonova u audio signalu. Decimacija povećava visinu tona, a interpolacija je smanjuje. Ovakve promjene audio signala obično nisu poželjne prilikom skaliranja. Međutim, postoje i tehnike koje modifikuju trajanje reprodukcije, uz očuvanje visine tona. Kako bi se skaliranje vremena analiziralo kao nezavisan efekat desinhronizacije, u ovom radu realizovana je tehnika iz ove grupe.

Postoje dva opšta pristupa skaliranju vremena koji čuvaju visine tonova u signalu. Prvom grupom pristupa, baziranoj na algoritmu faznog vokodera (*engl. phase vocoder*) [172], vrše se modifikacije audio signala u frekvencijskom domenu. Primjena ovih tehnika za skaliranje vremena održava kvalitet signala na visokom nivou, ali su prilično neefikasne. Njihova implementacija zahtijeva izvođenje velikog broja operacija, pa ih ne bi bilo praktično uključiti u proceduru obučavanja neuronskih mreža u okviru koje se obrađuju velike količine podataka. Druga grupa tehnika za skaliranje vremena sprovodi se u vremenskom domenu [173–175]. Ova grupa tehnika sklonija je stvaranju artefakata u signalu. Međutim, kako kvalitet skaliranog signala u našem slučaju nije presudan, jer se ovi efekti tretiraju kao napadi, pojava artefakata čini tehnike u vremenskom domenu razornijim napadima i time pogodnijim za primjenu u proceduri obučavanja. Obučavanjem sistema s ovom realizacijom vremenskog skaliranja veće su šanse da se postignuti nivo otpornosti prenese na suptilnije skaliranje faznim vokoderom, nego kada bi se obučavanje sprovodilo u inverznom scenariju. Dodatno, tehnike u vremenskom domenu su zbog svoje memorijske i vremenske efikasnosti pogodnije za integrisanje u arhitekturu neuronskih mreža.



Slika 37: Ilustracija postupaka za smanjenje (a) i povećanje (b) trajanja audio signala, bez promjena u visinama tonova.



Slika 38: Isječak govornog signala prije i nakon vremenskog skaliranja za 30% (ubrzavanje i usporavanje za 30%).

Svi pristupi u vremenskom domenu u suštini dijele sličnu ideju. Signal se najprije dijeli na skup kratkih segmenata fiksne dužine. Nakon toga se, da bi se promijenilo trajanje signala, jedan broj odbiraka između svaka dva segmenta može preskočiti, što rezultuje skraćenjem signala, ili se segmenti mogu djelimično preklapati, što za rezultat ima produžavanje signala. Izdvojeni segmenti se zatim spajaju *overlap-add* metodom [173] kako bi se proizveo novi zvuk, drugačijeg trajanja od originalnog. Ova metoda vrši postepeno stapanje susjednih segmenata kako bi se spriječila pojava očiglednih prekida u signalu na pozicijama gdje se spajaju dva segmenta. Postupak stapanja susjednih segmenata podrazumijeva množenje odbiraka na krajevima segmenata odgovarajućim koeficijentima, a zatim i njihovo sabiranje kako bi se dobili odbirci rezultujućeg signala. Na taj način se ublažava prelaz između dva segmenta. Koeficijenti kojima se množe odbirci na kraju prvog segmenta postepeno se smanjuju do vrijednosti 0, dok se koeficijenti za odbirke na početku sljedećeg segmenta postepeno povećavaju do vrijednosti 1. Opisani postupak ilustrovan je na Slici 37. Vrijednosti koeficijenata kojima se množe odbirci signala prikazani su kao prozorska funkcija koja se pomjera signalom. Na Slici 38 prikazan je rezultat primjene vremenskog skaliranja na isječak govornog signala. Prilikom simulacije vremenskog skaliranja u proceduri obučavanja i prilikom testiranja sistema signali su produžavani i smanjivani najviše za 30%.

U praktičnim implementacijama ove metode, segmenti na koje se dijeli signal se biraju uzimajući u obzir njihovu sličnost. Segmenti se ne uzimaju na fiksnoj udaljenosti, već se dopušta određeno odstupanje i pozicija segmenata se bira tako da njihovo spajanje proizvede što manje artefakata. Na ovaj način poboljšava se kvalitet rezultujućeg signala. Podudarnost dva segmenta može se ocjenjivati mjerama poput kros-korelacije. Međutim, budući da maksimizacija kvaliteta nije u fokusu realizacije ovih efekata, a uvođenje ovih koraka značajno usporava njihovu primjenu, pomenuti detalji nisu razmatrani u ovom istraživanju.

Biografija

Kosta Pavlović je rođen 8. maja 1994. godine u Beranama. Osnovnu školu i gimnaziju završio je u Kolašinu, kao dobitnik diplome „Luča“ i đak generacije. Tokom srednje škole, na državnim takmičenjima iz programiranja, ostvarivao je zapažene rezultate i predstavljao Crnu Goru na međunarodnim informatičkim olimpijadama.

Studijske 2012/13. godine upisao je Prirodno-matematički fakultet Univerziteta Crne Gore, smjer Računarske nauke. Osnovne i specijalističke studije završio je u roku sa prosječnom ocjenom 10.00.

Magistarske studije završio je na istom fakultetu 2018. godine. Magistarski rad pod nazivom „Primjena genetičkog algoritma za optimizaciju parametara algoritma izvlačenja informacija iz administrativnih dokumenata“ odbranio je 19.10.2018. sa ocjenom „A“. Iste godine upisao je doktorske studije na Prirodno-matematičkom fakultetu, smjer Računarske nauke. Dana 02.04.2021. godine odbranio je polazna istraživanja na temu „Umetanje vodenih žigova u digitalne audio signale primjenom dubokih neuronskih mreža“.

Kao uspješan student višestruko je nagrađivan. Dobitnik je godišnje studentske nagrade Univerziteta Crne Gore za 2014. godinu, nagrade Opštine Kolašin za najboljeg studenta 2015. godine, nagrade 19. decembar za studente koju dodjeljuje Glavni grad 2015. godine, kao i Plakete Univerziteta Crne Gore za 2016. godinu.

Od oktobra 2016. godine angažovan je kao saradnik u nastavi na Prirodno-matematičkom fakultetu Univerziteta Crne Gore na predmetima: Vještačka inteligencija, Programiranje 1, Programiranje 2, Strukture podataka, Programski prevodioci, Uvod u informacione sisteme, Softversko inženjerstvo, Računarske mreže i komunikacije, Distribuirani računarski sistemi, Bioinformatika, kao i na predmetu Primjena računara na Građevinskom fakultetu.

Izjava o autorstvu

Ime i prezime autora: Kosta Pavlović
Broj indeksa/upisa: 1/2018

U skladu sa članom 22 Zakona o akademskom integritetu, pod krivičnom i materijalnom odgovornošću, izjavljujem da je doktorska disertacija pod naslovom:

„Umetanje vodenih žigova u digitalne audio signale korišćenjem dubokih neuronskih mreža"

rezultat sopstvenog istraživačkog rada.

Istovremeno potvrđujem da:

- predložena disertacija ni u cjelini ni u djelovima nije bila predložena za dobijanje bilo koje diplome prema studijskim programima drugih ustanova visokog obrazovanja.
- su ostvareni rezultati korektno navedeni,
- nisam povrijedio autorska i druga prava intelektualne svojine koja pripadaju trećim licima.

Podnosilac izjave:

Kosta Pavlović

Podgorica, 6. februar 2024.

Izjava o istovjetnosti štampane i elektronske verzije doktorske disertacije

Ime i prezime autora: Kosta Pavlović
Broj indeksa/upisa: 1/2018
Studijski program: Računarske nauke
Naslov disertacije: „Umetanje vodenih žigova u digitalne audio signale korišćenjem dubokih neuronskih mreža"
Mentor: Prof. dr Igor Đurović

Izjavljujem da je štampana verzija moje doktorske disertacije istovjetna elektronskoj verziji koju sam predao za objavljivanje u Digitalni arhiv Univerziteta Crne Gore.

Istovremeno izjavljujem da dozvoljavam objavljivanje mojih ličnih podataka u vezi sa dobijanjem akademskog naziva doktora nauka, odnosno zvanja doktora umjetnosti, kao što su ime i prezime, godina i mjesto rođenja, naslov disertacije i datum odbrane.

Podnosilac izjave:

Kosta Pavlović

Podgorica, 6. februar 2024.

Izjava o korišćenju

Ime i prezime autora: Kosta Pavlović
Broj indeksa/upisa: 1/2018

Ovlašćujem Univerzitetsku biblioteku da u Digitalni arhiv Univerziteta Crne Gore pohrani moju doktorsku disertaciju pod naslovom: „Umetanje vodenih žigova u digitalne audio signale korišćenjem dubokih neuronskih mreža", koja je moje autorsko djelo.

Disertaciju sa svim prilozima predao sam u elektronskom formatu pogodnom za trajno arhiviranje.

Moju doktorsku disertaciju pohranjenu u Digitalni arhiv Univerziteta Crne Gore mogu da koriste svi koji poštuju odredbe sadržane u odabranom tipu licence Kreativne zajednice (Creative Commons) za koju sam se odlučio.

1. Autorstvo
2. Autorstvo – nekomercijalno
3. Autorstvo – nekomercijalno – bez prerade
4. Autorstvo – nekomercijalno – dijeliti pod istim uslovima
5. Autorstvo – bez prerade
6. Autorstvo – dijeliti pod istim uslovima

Podnosilac izjave:

Kosta Pavlović

Podgorica, 6. februar 2024.