UNIVERZITET SINGIDUNUM

BEOGRAD

DEPARTMAN ZA POSLEDIPLOMSKE STUDIJE

DOKTORSKA DISERTACIJA

DETEKTOVANJE MANIPULACIJE U VIDEO SNIMCIMA STVORENIH „DEEPFAKE“ TEHNIKOM SISTEMOM UČENJA PROSTORNO VREMENSKIH KARAKTERISTIKA

Mentor:

Prof. dr Marko Šarac

Student: Dušan Marković

Broj indeksa: 2016460025

Beograd, 2022.

Sažetak

U ovoj disertaciji analizirali smo i radili komparaciju metoda za preciznije i tačnije detektovanje manipulisanih video materijala uz pomoć Deepfake tehnike. Istraživanje je započeto analizom prethodnih modela predviđenih za detekciju manipulacije video materijala kroz Deepfake tehniku. Analizirani su prethodno obučeni modli i njihovi parametri. Analizirani prethodno obučeni modeli su *XceptionNet, EfficientNetB* i *EfficientNetV*. Parametri ovih modela koji su menjani u procesu preobučavanja su konfiguracija mreže *SingleDLCNN*, broj *fold-ova* kao i vrednost za *Hold-out* tehniku. Korišćen je DataSet sa preko 6000 datoteka od kojih je većina datoteka korišćena za treniranje neuronske mreže a ostale datoteke su korišćene za testiranje i validaciju. Za izdvajanje najboljih rezultata korišćen je *CV (Cross-Validation),* a tačnost istih je uvećana tehnikom težinskog usrednjavanja tj. optimizacijom težine. Prikazani su rezultati za sva tri obučena modela, a najbolji rezultat je ostvaren uz pomoć *EfficientNetbB4.* 96.8% *(FAR = 5.97%).* Smatramo da je postignutim rezultatom dokazan kvalitet metode učenja. Za budući rad planiramo unapređenje modela i eventualnu komercijalizaciju.

Abstract

This dissertation analyses the methods for more accurate detection of video materials manipulated using the Deepfake technique. Previous models intended for detecting video manipulation through the Deepfake technique, and their models and parameters were analysed. These models are XceptionNet, EfficientNetB and EfficientNetV. The SingleDLCNN network, the fold number and the value for the Hold-out technique parameters were changed in the retraining process. A DataSet with over 6000 files was used, with the majority used to train the neural network and the rest for testing and validation. Cross-Validation was used to extract the best results, and the accuracy was increased by weight averaging, i.e., by weight optimization. The results for all trained models are presented. The best result was achieved using the EfficientNetbB4. 96.8% (FAR = 5.97%). We believe that the achieved result proves the quality of this training model. For future work, we plan to improve the model and to eventually commercialise it.

Zahvalnost

Zahvaljujem se svojoj porodici, ocu Miodragu, majci Verici, bratu Milošu i supruzi Maji koji su mi uz puno razumevanja dali punu podršku tokom celih posle diplomskih studija.

Zahvaljujem se Univerzitetu Singidunum na čelu sa predsednikom prof. dr Milovanom Stanišićem koji je obezbedio usove za izradu ove disertacije kao i odlične uslove na posle diplomskim studijama.

Zahvaljujem se mentoru prof. dr Marku Šarcu na neizmernoj podršci prilikom izrade disertacije. Nesebičnom deljenju znanja i razumevanja. Takođe bih naglasio veliku požrtvovanost mentora u toku izrade disertacije na čemu sam posebno zahvalan.

Na kraju ali ne i najmanje važno, zahvaljujem se svim kolegama sa Univerziteta Singidunum koji su na bilo koji način doprineli ovoj disertaciji.

Sadržaj

Indeks slika

Indeks tabela

Indeks grafikona

1. Uvod

Većina algoritama za detekciju „Deepfake“ snimaka radi na osnovu analize individualnih kadrova. Pomenuti algoritmi često ne obraćaju pažnju ili u veoma malom obimu na Spatiotemporal, drugim rečima na prostorno vremenske karakteristike. Tačnost ovakvih algoritama nije potpuna, disertacija se bavi mogućim rešenjem povećanja tačnosti detekcije manipulacije video materijalom. U disertaciji se razrađuje 3D model neuronske mreže iz kojeg je moguće naučiti prostorno vremenske karakteristike susednih okvira (kadrova) u određenom video materijalu.

Napredak tehnologije omogućio je realizaciju kompleksnih algoritama u realnom vremenu. Kao posledica toga javili su se i neželjeni efekti u vidu zloupotrebe a u smislu manipulacije video i foto materijala. S obzirom na to, u prethodnom periodu u sve većem broju pojavljuju se video zapisi sa „sintetičkim“ licima. Napretkom algoritama za manipulaciju slikom ovi materijali postaju sve realističniji, a krivotvoreni video snimci zasnovani na „deepfake“ tehnici sve teži za detekciju. Navedeno dovodi do sve manjeg poverenja u validnost video snimaka, a često i zlonamernu distribuciju istih putem interneta. Zlonamerna distribucija se najčešće koristi u cilju da se diskredituje određena javna ličnost a vrlo je česta i u nameri da se diskredituju osobe iz političkog sveta [1]. Koliko je ovo ozbiljan problem možemo zaključiti ako samo zamislimo da neko kreira video gde se pojavljuje sa licem nekoga od svetski najmoćnih i najuticajnih političara i daje izjave umesto istog.

Predložena je metoda za izdvajanje lica iz kadrova video materijala u jednu sekvencu koja se kasnije koristi kao input u predloženom modelu. Predložena metoda je jednostavna, efikasna i pristupačna za upotrebu. Duboke neuronske mreže koje će kasnije biti opisane u tekstu koriste već obučene modele a formirane obučavanjem na osnovu miliona slika velikog broja različitih objekata (ImageNet). Usvojeni metod vrši preobučavanje ovakvih modela (Transfer Learning), uz upotrebu pretprocesiranja (Facial Extraction),

1.1. Motivacija i opšta razmatranja

Unapređenje tehnologije doprinosi mnogim korisnim stvarima, ali takođe moramo biti svesni da se moderna tehnologija može i zloupotrebiti. Kada govorimo u multimedijalnim sadržajima u smislu video materijala i fotografija napredak tehnologije nam je omogućio korisne opcije koje se koriste u svetu filma. Glumcima tehnologija pomaže u mnogim

slučajevima, ukoliko su bolesni glas se može izmeniti na njihov prirodni glas, ukoliko treba da izgledaju mlađe ili starije u toj ulozi takođe se može uticati kroz Deepfake tehnologiju. Kako je i navedeno pored pozitivne strane postoji i negativna. Negativna strana se ogleda u tome da se moderna tehnologija kao što je deepfake može uz pomoć komercijalnih, ali i profesionalnih alata u kratkom vremenskom periodu i bez mnogo truda zloupotrebiti. U najvećem broju slučajeva zloupotreba se svodi na neovlašćeno korišćenje nečijeg lika ili glasa i o tome će biti naknadno detaljnije pisano u poglavlju „Deepfake tehnika“.

S obzirom na lakoću manipulacije video snimaka zloupotreba je postala česta. Kao odgovor na pomenuto mnogi istraživači u svetu se trude da algoritmima prepoznaju manipulisani video snimak i da na taj način spreče zloupotrebu. Istraživači takođe imaju punu podršku velikih IT kompanija koje se takođe bore protiv zloupotrebe. Kao takav primer možemo navesti i kompaniju Google koja je obezbedila bazu podataka sa hiljadama manipulisanih video snimaka. U toj bazi koju u daljem tekstu nazivamo „DataSet“ se nalaze svi potrebni fajlovi za testiranje novih algoritama. To podrazumeva originalni video, masku i izmenjeni video. Jedna takva baza jeste korišćena i u ovoj disertaciji o čemu govorimo u poglavlju „Novi model učenja prethodno obučenih modela“. Česta zloupotreba i manipulacija video snimcima uz pomoć velikih kompanija jeste bila dovoljno velika motivacija za pisanje disertacije i istraživanje novih načina da se spreči ovakav vid korišćenja novih tehnologija.

1.2. Hipoteze doktorske disertacije

Opšta hipoteza od koje bi se krenulo u istraživanje u disertaciji je: *„Manipulacija video materijalima predstavlja realni problem, pouzdana detekcija i forenzika video materijala je moguća primenom algoritama prostorno vremenskih karakteristika”*

Posebna hipoteza koja proizilazi iz opšte je: *„Modeli i algoritmi spatiotemporal su dovoljno sofisticirani kako bi sa velikom pouzdanošću detektovali manipulaciju u video materijalu”*

Pojedinačne hipoteze koje su korišćene u disertaciji su:

1. Metode dubokog učenja su značajne za automatizaciju i pouzdanost modela detekcije manipulacije u video materijalima
2. Neuronske mreže pomažu povećavaju pouzdanost detekcije
3. Ekspanzija deepfake tehnike predstavlja problem koji je potrebno pravilno adresirati

4. 3D konvulucione mreže omogućavaju bolje i preciznije profilisanje i detekciju

1.3. Cilj doktorske disertacije

Cilj disertacije je definisanje, implementacija i analiza radnog okvira koji predlaže studiju izrade 3D slika i 3D modela CNN reportera, koji mogu da prilagode svoje osobine, prostorne i vremenske dimenzije na osnovu 3D slike izvora. Eksperimentalno je testiran model i dao je izvanredne rezultate u otkrivanju Deepfake video snimaka iz izvornih primeraka FaceForensics++ i VidTIMIT. Model je postigao tačnost veći od 99% primenom oba izvora Deepfake video snimaka za istraživanje. Postignuti rezultati prevazilaze tačnost najsavremenijih metoda koje se koriste u praksi.

Naučni cilj istraživanja se može opisati kao potreba i želja za definisanjem novog pouzdanijeg metoda detektovanja manipulacijom u savremenom video materijalu današnjice.

Praktični cilj ovog istraživanja je razvoj sopstvenog modela detekcije deepfake video materijala, verifikovan sa aspekta teorijsko informacione analize, opisanim procesom razvoja, upotrebe, analize performansi i poređenja sa ostalim srodnim rešenjima.

Društveni cilj ovog istraživanja je pomoć organizacijama različitih delatnosti (vlade zemalja, vojska, bezbednosne službe, veliki privredni subjekti) koje imaju neizostavnu potrebu za zaštitom važnih informacija.

1.4. Prikupljanje i izvori podataka

Istraživanje je sprovedeno upotrebom dostupnih saznanja i informacija iz navedene oblasti, koristile su se informacije prikupljene putem Interneta, dostupne literature, časopisa i radova. Prikupljeni sadržaj je analiziran i ustanovljeno je postojeće stanje. Nakon toga teži se novim metodama i eksperimentima koji dokazuju da se može doći do boljih rezultata u otkrivanju manipulisanih video snimaka Deepfake tehnikama.

Od naučnih metoda koristila se analitičko-deduktivna metoda, od opšte-naučnih hipotetičko-deduktivna, uporedna i komparativna metoda, a od metoda i tehnika koristila se eksperimentalna metoda ispitivanja.

Prikupljanje i analiza podataka izvršena je:

- postavljanjem kriterijuma za poređenje i klasifikaciju,

- analizom prikupljenih podataka
- upoređivanjem prikupljenih podataka,
- utvrđivanjem relevantnih činjenica i veza među podacima,
- preispitivanjem hipoteza,
- testiranjem i proverom zaključaka do kojih smo došli
- postavljanjem budućih ciljeva.

1.5. Struktura disertacije

Proces naučnog istraživanja je podeljen u nekoliko koraka

- Uvod
- Pregled u oblasti istraživanja
- Neuronske mreže
- Duboko učenje
- Novi metod konfiguracije učenja prethodno obučenih modela
- Zaključak sa sumarnim rezultatima i predlog daljeg rada

U nastavku disertacije, drugom poglavlju urađen je pregled u oblasti istraživanja. Opisana je sama Deepfake tehnika, način funkcionisanja iste kao i njena upotreba. Pored opštih metoda objašnjene su i metode koje su zasnovane na vremenskoj doslednosti, vizuelnim artefaktima, otiscima prstiju kamere i biološkim signalima

U trećem poglavlju disertacije opisane su neuronske mreže. Preciznije su definisane jednoslojne neuronske mreže kao i višeslojne neuronske mreže. Takođe opisuju se i vrste veza između neurona i smerovi kretanja informacija, kako povratni tako i nepovratni smer.

U četvrtom poglavlju disertacije objašnjeno je duboko učenje i postojeće metode na vremenskim odlikama. Objašnjene su tri postojeće metode ovih karakteristika.

U petom poglavlju disertacije opisan je novi metod konfiguracije učenja prethodno obučenih modela. Opisana su tri prethodno obučena modela koja su korišćena u eksperimentu. Detaljno je opisan postupak konfiguracije DataSeta koji je korišćen kao i konfiguracija

neuronske mreže sa konvolucijom. Takođe opisani su i modeli koji su korišćeni. Nakon toga predstavljeno je istraživanje i rezultati istog. U istom poglavlju predočeno je i poređenje sa drugim radovima.

Šesto poglavlje predstavlja zaključak. Prikazan je rezultat i doprinos disertacije. U istom poglavlju predstavljen je i predlog daljeg rada.

2. Pregled u oblasti istraživanja

Brzim napretkom metoda za krivotvorenje lica u prethodnim godinama došlo je do sve veće i veće zloupotrebe što je izazvalo masovnu zabrinutost javnosti. Samo delimično falsifikovana lica kao i kompletna zamena lica se danas mogu izvesti mnogo lakše nego ikada u istoriji modernog doba. Alati kojima se može izvršiti zamena lica danas su lako dostupni, mogu se lako proizvesti falsifikovana lica uz pomoć alata kao što su *FaceSwap*, *Deepfakes*, itd. Ovi alati više ne zahtevaju sofisticirane računare, to dodatno pogoršava stvari i samim tim falsifikovanje postaje još lakše i pristupačnije. Sve ovo ozbiljno utiče na internet, društvenu, ali čak i na političku sigurnost [2]. Zbog navedenog ključna je potreba da se razviju efikasne tehnologije za otkrivanje falsifikovanog materijala. Možemo zaključiti da je problem binarne klasifikacije i postojeće metode se mogu podeliti na metode zasnovane na slikama ili na određenim okvirima ekstrahovanim iz video materijala. Metode zasnovane na okvirima u većem broju slučajeva se fokusiraju na pretraživanje falsifikovanih obrazaca na svakoj slici koristeći informacije o bojama [3], statistiku frekvencija [4], pomoćnih maski [5] ili informacije o granicama mešanja [6]. S obzirom da se falsifikovani video materijal generiše sliku po sliku, kadar po kadar, često postoji neprirodan prelaz između slika ili vremenske nedoslednosti kao što je podrhtavanja položaja lica. Naprednije i novije tehnologije falsifikovanja mogu stvoriti snimak jako blizu originalnom izgledu lic,a ali ne mogu eliminisati ovu vremensku nedoslednost. Mnogi istraživači su počeli da rade na modelima i različitim metodama u otkrivanju falsifikovanih video materijala. Generalni modeli za analizu koji su korišćeni u ovoj oblasti su C3D [7], I3D [8] i LSTM [9]. Ove metode se koriste u oblasti otkrivanja falsifikata, ali one nisu kreirane iz tog razloga, bez obzira na to, metode su dosta pomogle, ali su i iziskivale velike troškove u računarskoj opremi. Pored visokih troškova mana im je i što nisu mogle da se fokusiraju na uočavanje vremenskih nedoslednosti.

2.1. Deepfake tehnika

Nove digitalne tehnologije sve teže prikazuju razliku između pravih i lažnih medija. Jedna od novijih problema je pojava deepfake tehnologije koji čine video snimke hiper-realističnim i koji primenjuju veštačku inteligenciju da bi prikazali određenu osobu koja govori i radi stvari koje se nikada nisu dogodile [10]. Deepfake predstavlja proces u kome se vrši zamena postojeće osobe iz video snimka ili slike sa drugom osobom, pri čemu se, nakon izvršene zamene, novo lice na video snimku ponaša i izgovara isti tekst identično osobi koja se nalazi na originalnom snimku. Zamena lica, odnosno manipulacija izrazom lica na slici i video snimku naziva se Deepfake tehnikom [11]. Problemi videa u kojima se manipuliše licem postaju sve rasprostranjeniji nakon pojave deepfake tehnologije koja manipuliše slikama i video zapisima pomoću dubokog učenja. Duboko učenje je omogućilo kreiranje lažnih tekstova, glasova i video snimaka, kao i lažnih fotografija, koji na prvi pogled izgledaju autentično i realistično, iako zapravo nisu . Algoritam deepfake tehnologije može da zameni lica u video snimku sa licima u izvornom video snimku pomoću autoenkodera ili generativnih suparničkih mreža [12]. Sa ovom tehnologijom, video snimci kojima se manipuliše su izuzetno jednostavni za generisanje usled dostupnosti velike količine podataka javnosti. Deepfake korišćen u svrhe lažnih vesti je postao tema koja je pretnja javnosti, društvu i demokratiji [13]. Lažne vesti se odnose na fiktivne sadržaje koje za cilj imaju obmanu javnosti [14]. Pomoću društvenih mreža, ovakve vesti se šire veoma brzo i mogu doći do velikog broja korisnika [15]. Najveći broj korisnika dolazi do vesti pomoću youtube.com veb sajta. Takav način prikazivanja informacija pomoću videa povećava potrebu za alatima koji potvrđuju autentičnost medijskog i novinskog sadržaja [16]. Razlog tome je taj što nove tehnologije omogućavaju ubedljivu manipulaciju video snimcima. S obzirom na jednostavno i brzo širenje neistinitih informacija putem platformi društvenih mreža, sve je teže prepoznati originalnu i validnu vest, što ostavlja negativne posledice pri donošenju odluka i sagledavanja istine [13]. Današnja tehnološka otkrića kao što je deepfake, omogućavaju da se u obrađenim video snimcima i slikama, u kojima se koristi zamena lica, ostavlja veoma malo traga manipulacije [17].

Kao što je već napomenuto, deepfake su proizvodi veštačke inteligencije, koji spajaju, kombinuju, zamenjuju i preklapaju slike i video snimke kako bi kreirali lažne slike i video snimke koji izgledaju autentično [18]. Deepfake tehnologija može da generiše, na primer, humorističan, politički ili čak neprimeren video sadržaj u kome se, bez pristanka nalazi neka osoba i njen glas, ali ne i njene originalne izjave. Faktori koji utiču na deepfake su obim,

razmera i sofisticiranost uključene tehnologije, i omogućuju da skoro svaki računar može da napravi lažne video snimke koji se teško razlikuju od autentičnih medija [19]. Pored toga što se deepfake tehnologija može koristiti u mnoge pozitivne svrhe, kao što su snimanje filmova i virtuelna stvarnost, ona se učestalo i u velikoj meri zloupotrebljava. Deepfake sadržaj može biti kao što smo rekli političkog, humorističkog ili čak neprimerenog sadržaja ili se koristiti za ucenjivanje pojedinca korišćenjem njenog ili njegovog lika i glasa bez pristanka [19]. Primer za pozitivan način korišćenja Deepfake tehnologije u filmu jeste *CGI* (Computer-generated imagery). *CGI* je model računarskog generisanja slike i primenjuje se u oblasti računarske grafike, da budemo precizniji najčešće se koristi u svrhe 3D računarske grafike. Model je primenu pronašao u filmovima, televizijskim programima i reklamama u najvećem broju slučajeva. *CGI* se koristi za vizualne efekte koji se ovim modelom lakše kontrolišu od drugih zasnovanih na fizičkim procesima [20]. Prednost modela takođe jeste u tome što dozvoljava jednom samostalnom umetniku da proizvede sadržaj bez upotrebe skupih scenografija i skupih rekvizita. Kao neki od primera gde je *CGI* korišćen u svetu filma možemo nvesti *Jurassic Park, Avatar, Toy Story* [21]*, District 9, Transformers, Tron, Gladiator, The Matrix, The Last Strafighter, Spider-Man, Star Trek, Star Wars, The Lord of the Rings, Terminator*, itd. [22]

Pored već navedenih mogućih primera zloupotrebe u politici problem može da nastane i kod građanstva u svakodnevnom životu, primer može biti da se pomoću ove tehnologije, mogu se kreirati lažni avatari za platforme sa mogućnošću video poziva *(Microsoft Teams, Google meet, Viber, Zoom, Skype,..)*, kao i kreiranje lažnih biografija. Navedeni slučaj može predstavljati opasnost za velike kompanije koje imaju veliki broj prijavljenih kandidata za zapošljavanje ili lica koji poslove obavljaju na daljinu, a koji se lažno predstavljaju u realnom vremenu [23]. Dok je tehnologija koja se koristi složena, njena složenost je skrivena iza uobičajenih alata koji su jednostavni za korišćenje i dostupni su široj javnosti [24]. Alati za kreiranje Deepfake-a imaju niske tehničke zahteve koji obično zahtevaju konvencionalne kućne računare sa opremljenim grafičkim karticama za video igre. Zbog široke dostupnosti ovakvoj vrsti tehnologije, ovakve alate mogu koristiti u kućnim uslovima na svojim računarima bez potrebe za duboku tehničku ekspertizu [23]. Ovo može dovesti do stvaranja realističnog lažnog sadržaja za koji je teško proveriti njegovu autentičnost, a koja putem društvenih medija može dospeti u široku javnost. Digitalni mediji pružaju mogućnost kopiranja i prenosa informacija veoma brzo i lako, a ciljana publika može biti bilo ko, bez vremenskih ograničenja, lokacijskih ograničenja i veličine sadržaja. Korišćenje digitalnih medija kao sredstva za komunikaciju, preko internet mreže, omogućuje trenutnu komunikaciju i interakciju na

globalnom nivou [25]. Deepfake tehnologija je snažno povezana sa digitalnim medijima, a posebno sa društvenim medijima preko kojih dopiru do široke publike. Tekst, zvuk, slike i video zapisi su ključni elementi interakcije i komunikacije u javnoj sferi, mogućnost manipulacije njima može biti veoma zloupotrebljena. Ukrštanje digitalnih medija i veštačke inteligencije je od velikog interesa i predstavlja osnovu za medijsku komunikaciju u savremenom veku, a velika tehnološka dostignuća omogućavaju da se pređe granica stvarnosti i širenja informacija [23]. Rezultat može biti pojačan uticaj digitalnih medija u neviđenim razmerama, što može imati pozitivne uticaje na društvo, ali može doneti i negativne uticaje, odnosno zloupotreba pri usmeravanju javnosti. Najčešće su javne ličnosti glavne žrtve deepfake tehnologije. Za širenje lažnih izjava svetskih lidera, ovaj vid tehnologije je korišćen nekoliko puta, pri čemu takva manipulacija videima može izazvati pretnju po mir u svetu [26]. Takođe, može se koristiti za zavaravanje vojnih lica pružanjem lažnih mapa, što može izazvati ozbiljne štete. Danas, deepfake tehnologija postaje sve popularnija zbog svije jednostavnosti i jeftine dostupnosti [27]. Širok spektar primena deepfake tehnika stvara posebno interesovanje među profesionalcima, a takođe i među početnicima. Prvi uspeh deepfake tehnike je kreiranje *FakeApp*-a koji se koristi za zamenu lica osobe sa drugom osobom. Softver *FakeApp* koristi veliki broj podataka kako bi se postigli bolji rezultati. Kreiranje lažnih video snimaka u *FakeApp*-u podrazumeva ekstrakciju svih slika iz videa, zatim se pravilno obrade, nakon čega se vrši spajanje lica što za produkt ima finalni video [28].

Jedan od prvih i verovatno najpoznatijih deepfake-a je video iz 2018. godine gde Barak Obama upozorava na opasnost od deepfake-a, odnosno prikazuje video koji Barak Obama nikada nije snimio. Video je u kratkom vremenskom periodu pogledalo vise od 9.000.000 ljudi [29]. Kadar iz ovog videa se prikazuje na slici 1.

Slika 1 – Kadar iz manipulisanog video materijala

Takođe, jedana od zanimljivijih situacija jeste kada popularni Deepfake, uključuju u alternativnu verziju nekih delova popularnog američkog triler filma „*Taxi driver*“ iz 1976. godine, gde je lice glavnog glumca, odnosno Roberta De Nira, je zamenjeno sa licem Al Paćina. „*Taxi driver*“ deepfake je lako kreiran uz komercijalni računar i javno dostupnog softvera *DeepFaceLab* koji koristi veštački inteligenciju za zamenu lica u video snimku. Deepfake tehnologija je poprimila novi nivo sofisticiranosti, koja omogućuje neovlašćeno umetanje audio i vizuelnih slika na platforme društvenih medija, a razlikovanje stvarnih i lažnih informacija postalo je veoma teže nego ikada ranije. Pretnja od zloupotrebe korišćenja analizirane tehnologije nije upućena samo pojedincima ili grupi građana, već i institucijama, kao i nacionalnoj bezbednosti. [30]

2.2. Način funkcionisanja deepfake tehnike

Kombinacija „dubokog učenja“ i „lažnog“ predstavljaju deepfake tehniku koji se odnose na hiper-realistične video snimke i slike kojima se digitalno manipuliše prikazujući osobe koje govore i rade stvari koje se nikada nisu zapravo dogodile. Deepfake se oslanja na neuronske mreže koje analiziraju velike skupove uzoraka podataka i omogućavaju im da oponašaju izraze lica, pokrete, ali i glas. Ideja o zameni lica na fotografiji nije tako nova kao što se može pretpostaviti. Može se pronaći primer fotografija napravljenih u 19. veku, nastala kao ilustracija, napravljena ručno, na kojoj je glava predsednika Linkolna zamenjena glavom političara Džona Kalhuna [31]. Međutim, kada je ideja o neuronskim mrežama postala popularna i kada su unapređene mogućnosti računara, ova tehnologija počela je da se koristi u svakodnevnom životu. U današnje vreme je moguće preuzeti i pokretati takve programe, a njihovim eksperimentisanjima moguće je omogućiti sticanje iskustava u ovoj oblasti [31].

Prema različitim ciljevi algoritama za manipulaciju licem, postojeći deepfake algoritmi mogu se podeliti u dve kategorije: zamena lica i rekonstrukcija lica. Postoje dva uobičajena načina za kreiranje deepfake videa, a to su autoendkoderi i *GAN*-ovi.

2.2.1. Razvoj deepfake tehnike

Video snimci i zamene lica, odnosno zamene identiteta osoba u video snimku početkom dvehiljaditih godina su počeli da privlače veliku pažnju. Od 2017. godine počinju aktuelna istraživanja na ovu temu, pa je tako u studiji Koršunove [32], neuronske mreže obučene su da snime izgled ciljnog identiteta iz kolekcije fotografija, što omogućava generisanje slike visokog kvaliteta za zamenu lica. Međutim, ovaj pristup se ne može primetiti generisanje video

zapisa visokog kvaliteta [12]. Iste godine Ozlevski [33] je predložio nov pristup za generisanje video zapisao sa jednom *RGB*-om *(Red Green Blue color channels)* i izvornom video sekvencom. Duboka generativna mreža je korišćena kako bi se zaključile deformacije teksture po kadru ciljnog identiteta koristeći izvorne teksture i jednu ciljnu teksturu [12]. Zasnovan na ovoj metodi, novo izvedeno lice bi se moglo komponovati na izvorni video, zamenjujući originalno lice. U jesen 2017. godine *Reddit (Društvena mreža Reddit)* je objavio prvi video o zameni lica generisan *Deepfake* tehnologijom [34]. Nakon spomenutog događaja, otpočinje talas kreiranja videa sa zamenom lica širom sveta, bez obzira da li je uticaj bio pozitivan ili negativan [12]. Predložena je *Facewap-GAN (GAN – Generative Adversarial Networks)*, poboljšana verzija originalnog deepfake algoritma. Da bi se generisala realističnija lica, otklonjen je gubitak percepcije, kako bi se poboljšale performanse autodekodera koji je implemetirao *VGGFace*. *VGGFace* je baza od 2.6 miliona slika lica koju je objavila Kancelarija direktora nacionalne obaveštajne službe SAD *(Sjedinjenje Američke Države)* u cilju suzbijanja zloupotrebe. Skraćen naziv za ovu organizaciju je *ODNI* [35]. *DeepFaceLab* kreirao je okvir za duboko lažiranje, dizajniran tako da bude lako dostupan i jednostavan za korišćenje ljudima koji nisu stručno obučeni za ovakav vid manipulisanja videima.

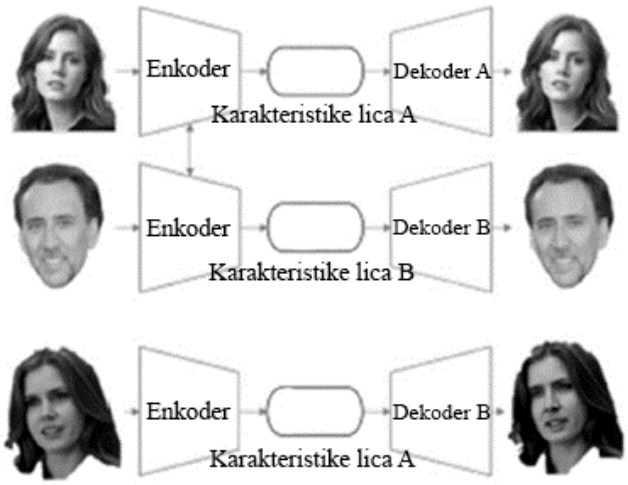
Za razliku od zamene lica, rekonstrukcija lica se odnosi na pokušaj kontrole izraza lica ljudi u video snimku, što znači da je moguće generisati video na kome se nalazi osoba koja vrši radnju koju nikada nije u stvarnosti. Prvi algoritam za rekonstrukciju lica može se reći da datira iz 2006. godine kada je Vlašik [36] predložio da se izvrši rekonstrukcija lica na osnovu šablona lica, koji je modifikovan pod različitim parametrima ekspresije. Većina narednih radova zasnovana je na takvim šemama, gde se parametarski model koristi za prilagođavanje izraza lica. Da bi se izvršila monokularna rekonstrukcija lica, predlaže se Face2Face. U pomenutoj studiji, novi globalni pristup zasnovan na modelu je primenjen da bi se rekonstruisale crte lica ciljnih i izvornih aktera. U isto vreme, dizajnirana tehnika prenosa deformacije vrši prenos izraza između izvornih i ciljnih aktera [12]. Pored ovih doprinosa, ova studija je takođe predložila novi metod u sintezi regiona usta, gde se najbolja podudarna slika preuzima iz ciljne sekvence. Međutim, sa razvojem tehnika dubokog učenja, grubi detalji koji su se uočavali na snimcima, a koji su mogli lako da otkriju lažiran snimak, postepeno rešavaju. Sledeći cilj je bilo mapiranje sekvenci sa audio na video kako bi manipulisali glumcima da govore iste rečenice. Karakteristike izvučene iz glasovnih sekvenci koriste se kao ulaz u *RNN (ponavljajuće neuronske mreže)* koji daje oblik usta koji odgovara svakom kadru u videu. Kako bi se izvršila rekonstrukcija lica sa boljim performansama, predložen je novi metod za foto-realističnu reanimaciju video zapisa u portretu. Predložena generativna neuronska mreža sa

novom arhitekturom prostor-vreme se koristi za transformaciju grubih prikaza modela lica u potpuno foto-realističan portret video izlaz.

2.2.2. Autoendkoderi

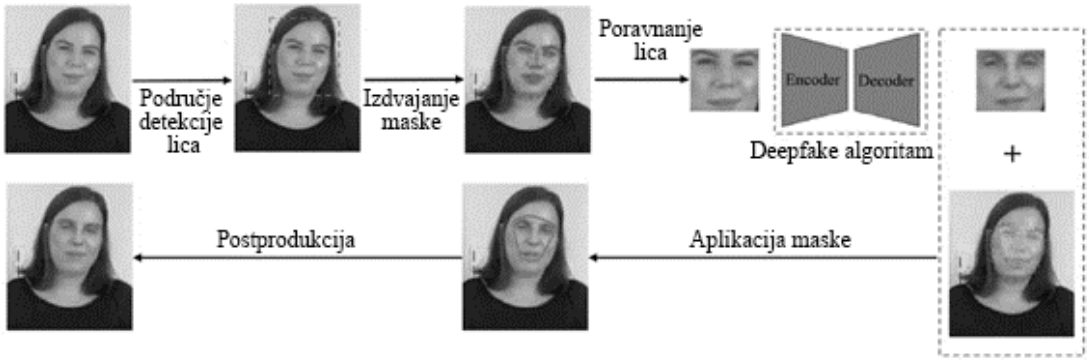
Autoendkoderi su široko korišćen model tehnika dubokog učenja. Sadrže dve uniformne duboke mreže gde 4 ili 5 slojeva predstavljaju polovinu kodiranja, a ostatak polovinu za dekodiranje [27]. Duboko kodiranje se koristi za smanjenje dimenzije i kompresiju slika [37]. Kao što je već spomenuto, prvi uspeh deepfake tehnike bilo je kreiranje *FakeApp* softvera koji se koristio za zamenu lica sa drugom osobom. Autoenkoderi se obično spominju kada je u pitanju smanjenje dimenzija ili generativnog modela učenja. Takođe se mogu koristiti za uzimanje kompresovanih reprezentacija slika, kako bi se nadmašili postojeći standardi za kompresiju slika. Mogu se koristiti takozvani latentni vektori autoenkodera prvog enkodera sa delom dekodera drugog autoenkodera i kao rezultat se dobija slika sa zamenjenim licem [31]. Neke od platformi sa deepfake tehnikom su *Faceswap, DFaker, DeepFakeLab*, itd.

Dve mreže koje dele isti koder, a koriste dva različita dekodera prikazani su na slici 2. Kada je potrebno izvršiti novu zamenu lica, kodiramo ulazno lice i dekodiramo ga pomoću dekodera ciljanog lica. Da biste generisali video sa zamenom lica, svi okviri ciljnog videa moraju biti obrađeni pomoću generativne metode. Na slici (slika 2) prikazan je opšti proces generisanja video zapisa. Očigledno je da je deepfake algoritam, koji implementira zamenu lica dok čuva izvorne izraze, ključni deo generisanja video zapisa [12]. Duboki lažni algoritmi koji se koriste u zameni lica uglavnom su razvijeni na osnovu autoenkodera, koji se široko koristi za zadatke rekonstrukcije podataka. Latentne karakteristike se prvo izdvajaju iz slike od strane enkodera, a zatim se unose u dekoder da bi se rekonstruisala originalna slika. U *Deepfake-u* algoritam, dva autoenkodera su obučena da menjaju lica između izvornih video okvira i ciljnih video okvira.



Slika 2 – Dve mreže koje dele isti koder

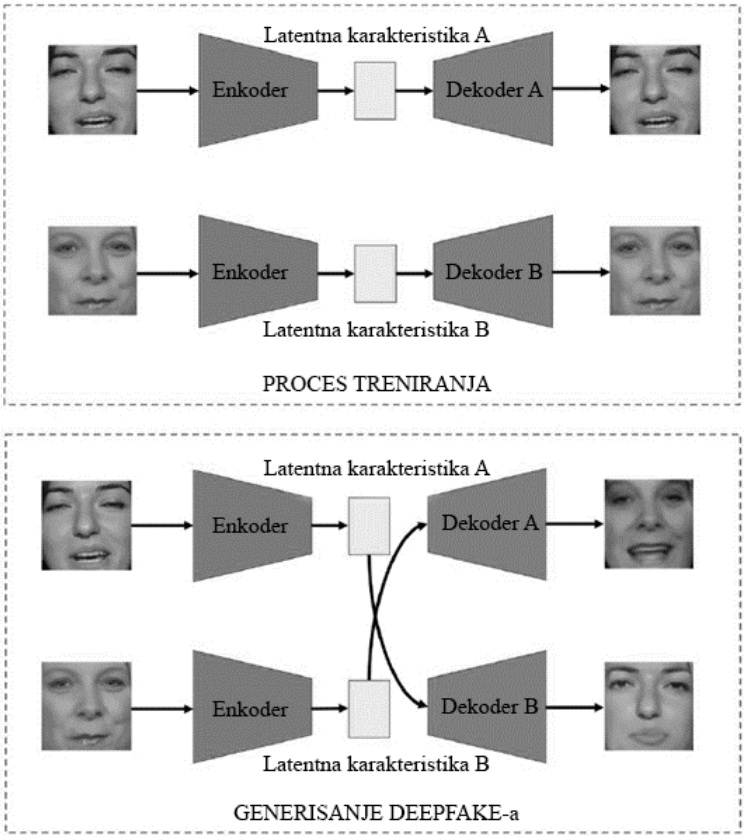
Oblast lica se prvo detektuje u svakom video kadru (Slika 3). Zatim se izdvajaju oznake lica da izvrši poravnavanje lica. Nakon toga, deepfake algoritam *(autoencoder ili GAN – Generative Adversarial Networks)* se primenjuje za generisanje sintetičkog lica unosom slike poravnate sa licem. Da bi se smanjili artefakti uzrokovani mešanjem, orijentiri leve i desne obrve i donjih usta se koriste za generisanje specifične maske. Na ovaj način, nakon spajanja sintetičkog lica sa originalnom slikom, zadržava se samo sadržaj unutar maske. Konačno, da bi se generisana slika dodatno učinila realističnom, operacija post-procesiranja je dopunjena za obradu generisane slike. Konkretno, Gausovo zamućenje se primenjuje na granicu maske dok se boja algoritam korekcije se primenjuje da bi se obezbedila konzistentnost sintetičkog lica i pozadinske slike.



Slika 3 - Proces generisanja video kadrova zamene lica [12]

Tokom procesa obuke, dva enkodera sa istim težinama su obučena da izdvajaju zajedničke karakteristike u izvornom i ciljnom licu [12]. Zatim se ekstrahovane karakteristike unose u dva dekodera da bi se rekonstruisala lica, respektivno. Potrebno je napomenuti da je dekoder A obučen samo sa licima A, dok je dekoder B obučen samo sa licima B. Kada se proces obuke završi, latentno lice generisano iz lica A će biti prosleđeno na dekoder B. Dekoder B bi pokušao da rekonstruiše lice B iz karakteristika u odnosu na lice A. Ako je autokoder dobro obučen, latentni prostor će predstavljati izraze lica. Drugim rečima, lice koje generiše dekoder B imaće isti izraz kao lice A. Prvo, monokularna rekonstrukcija lica se izvodi na izvornom i ciljnom licu da bi se dobili njihovi odgovarajući parametri lica. Nakon toga, parametri se modifikuju očuvanjem parametara osvetljenja i identifikuju uz promenu parametara poze, izraza i pogleda. Sintetičke slike se zatim generišu korišćenjem modifikovanih parametara. Konačno, mreža za prevođenje rendera u video se primenjuje za generisanje video snimaka za rekonstrukciju lica.

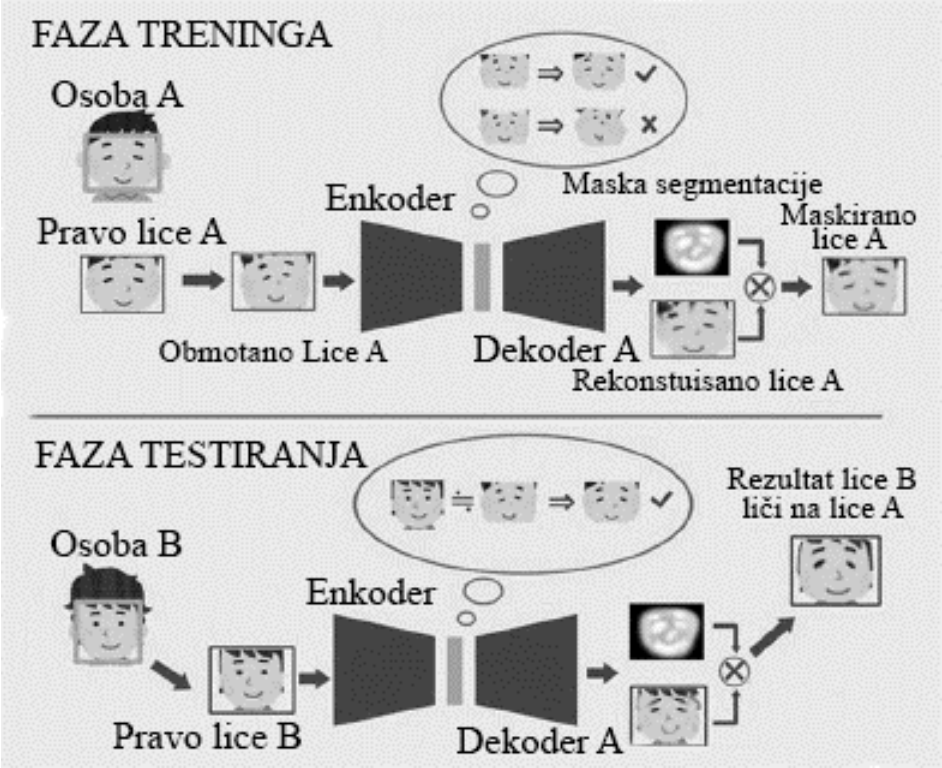
Zadatak rekonstrukcije lica ima za cilj da izvrši migraciju izraza lica. Prikazuje opšti proces izvođenja rekonstrukcije lica. Prvo se koristi niskodimenzionalni prikaz parametara izvora i cilja, zatim se video snimci dobijaju korišćenjem metode monokularne rekonstrukcije lica. Takođe, poza glave i izraz mogu se preneti u prostor parametara. Tokom izvođenja rekonstrukcije lica, scene osvetljenja i parametri identiteta su sačuvani, dok su parametri poze glave, izraza i pogleda oka promenjeni [12]. Nakon toga, sintetičke slike ciljnog aktera se regenerišu na osnovu modifikovanih parametara. Ove slike se zatim služe kao uslovni ulaz naše nove mreže za konverziju videa za renderovanje, koja se zatim obučava da konvertuje sintetizovani ulaz u realističan izlaz. Da bi se dobio kompletan video sa boljom vremenskom konzistentnošću, prostorno-vremenski volumeni kondicioniranja se unose u mrežu na način kliznog prozora. Na ovaj način se može dobiti video za rekonstrukciju lica. Kada se generišu duboko lažna lica, dekoderi se zamenjuju kao što je prikazano na slici 4.



Slika 4 - Autoenkoderi koji se koriste za zamenu lica [12]

2.2.3. GAN (Generative Adversarial Networks)

Jedan od najtežih modela za obuku i upotrebu za duboko učenje u računarskoj tehnici je *GAN* (**Error! Reference source not found.**). Ovaj model sadrži 2 neuronske mreže – generator i diskriminator (klasifikator). Mreža generatora izgleda slično kao i mreža autoenkodera, ali može postići bolje rezultate jer diskriminatorska mreža uklanja loše primere [31]. Ova tehnika kreiranja deepfake-a ima za cilj da prevari diskriminatora i na taj način formira lažne snimke koji su veoma slični originalnim i kao takve ih čini veoma teško prepoznatljivim za ljudsko oko. Odnosno, generator pokušava da stvori nove uzorke koji su dovoljno verodostojni da prevare drugu mrežu, koja radi na utvrđivanju da li su novi mediji koje vidi stvarni i originalni [10]. Na taj način dolazi do poboljšanja videa u slučaju da diskriminator oceni da snimak nije originalan. *GAN* može da analizira hiljade fotografija jedne osobe i da na osnovu analize stvori novi portret koji je približan tim fotografijama, a pritom da ne napravi nijednu kopiju od analiziranih fotografija [10]. Iako je za deepfake obično potreban veliki broj slika da bi se napravio realističan falsifikat, istraživači su već razvili tehniku za generisanje lažnog videa tako što će mu dati samo jednu fotografiju, kao što je selfi.



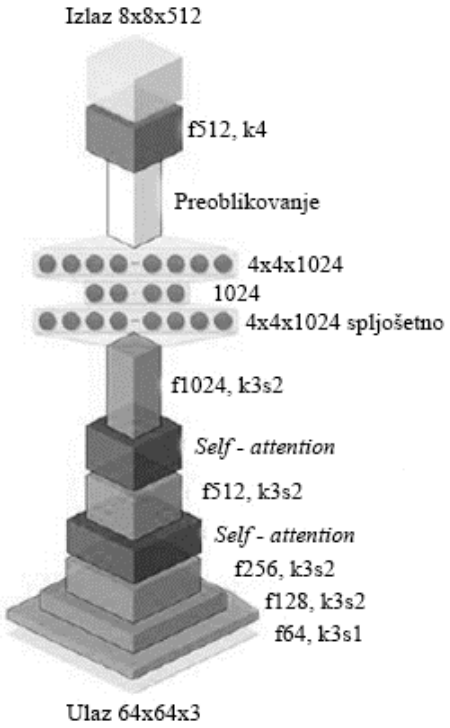
Slika 5 - Prikaz GAN metode [12]

FaceSwap-GAN podrazumeva automatskog kodiranja *VGGface* Deepfake-*a*. Podržava nekoliko izlaznih verzija kao što su 64x64, 128x128 i 256x256 i ima sposobnost da izradi realistične i dosledne pokrete očiju. Koristi se *VGGFace* koncept gubitka percepcije koji pomaže da se poboljša pravac očnih jabučica da bude precizniji i u skladu sa ulaznim licem. Ovaj alat uključuje *MTCNN (Multi-Task cascaded Convolutional Neural Networks)* tehniku detekcije lica i kalman filter za detekciju stabilnijeg lica i izglađivanje lica. *„Faceswap-pytorch"* još jedan alat za kreiranje dubokih lažnih podataka koji čini čitač skupova podataka efikasnijim i može se učitati iz dva direktorijuma istovremeno [27]. Većina alata za zamenu lica koristi koncepte generativnih suparničkih mreža (*GANS*). Vizuelni dijagram enkodera (Slika 5) u *GANS-u*. Ima sposobnost da izvuče duboke informacije iz slike, u fazi kodiranja (Slika 6) izvlači izraze lica dve osobe, zatim analizira zajedničke karakteristike dva lica i pamti ih. Nakon toga, u fazi dekodiranja (Slika 6), dekodira informacije dve osobe (Slika 7) . slika prikazuje rad diskriminatora koji je doneo odluku o autentičnosti karakteristika svakog datog podatka.

Na prikazanim slikama oznake su sledeće:

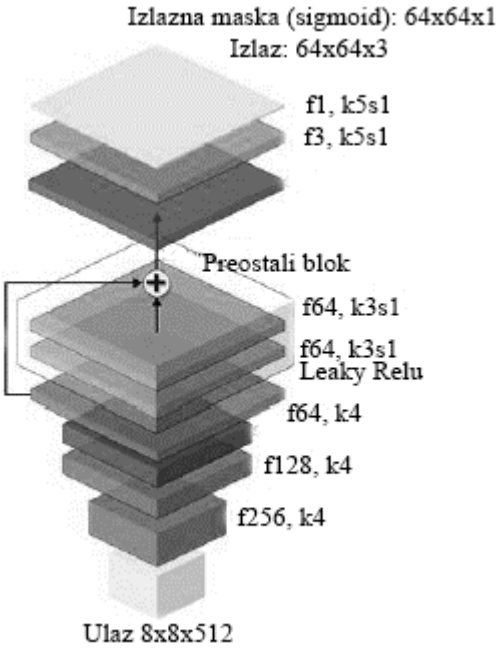
Oznaka	Značenje
f	Broj filtera
k	Veličina kernela
s	Koraci

Tabela 1 – Oznake korišćene na slikama za fazu kodiranja i dekodiranja



Slika 6 – Tehnika kodiranja alata za zamenu lica [27]

Na prethodnoj slici opisana sposobnost da izvuče duboke informacije iz okvira u fazi kodiranja je prikazana kroz različite slojeve. Predstavljeni su brojevi filtera, veličina kernela i koraci. Ulazni parametri su 64x64x3 dok su izlazni parametri 8x8x512. Na sledećoj slici prikazana je faza dekodiranja takođe slojevito. Takođe su predstavljeni brojevi filtera, veličina kernela i koraci. Ulazni parametri u ovom primeru su bili 8x8x512 a izlazna maska je bila 64x64x1.



Slika 7 – Tehnika dekodiranja alata za zamenu lica [27]

2.3. Upotreba deepfake tehnike

Kultura se stalno menja [38], a njen uticaj na društvo u uskoj je povezanosti sa načinom na koji društvo koristi moderne tehnologije. Kako bi obrazložili ovu tvrdnju dobar primer jeste sama brzina dobijanja informacija. Sredinom dvadesetog veka informacije su krenule da se šire brže nego ranije, takav trend je počeo pojavom radio uređaja, fiksnih telefona, zatim televizora a znatno se ubrzao pojavom mobilnih telefona i interneta [39]. Samim tim možemo zaključiti da je uticaj u društvu veliki i da zavisi od toga koliko je isprepleten sa načinom na koji ljudi komuniciraju i koriste moderne tehnologije. Medijatizacija obuhvata takve aspekte jer pokušava da obuhvati dugoročne procese međuodnosa između medijskih promena, s jedne strane, i društvenih i kulturnih promena s druge strane. Sve u svemu, medijatizacija je proces, ali se „može posmatrati i kao kontejner u kojem se mogu prikupljati zapažanja, i kao „dinamičan proces povećanja uticaja medija, ne može se smatrati determinističkim i linearnim razvojem. Masovni mediji su u prošlosti evoluirali kao što je rečeno od starih tradicionalnih medija do digitalnih mreža i platformi koje se doživljavaju kao zajednički prostor povećane vidljivosti i povezanosti. Masovni mediji imaju efekte koji dovode do promene ishoda jedne osobe, društvene grupe ili kompanije, koji nastaje kao posledica nakon izlaganja ili niza poruka prenetih u masovnim medijima [40]. Kao takav, sadržaj koji se plasira preko masovnih medija, a koji u sebi sadrži deepfake tehnologiju, predstavlja određenu zabrinutost. Danas je teško

regulisati upotrebu deepfake tehnike, koja je postala deo svakodnevnog društvenog života, koji mogu biti zabavnog karaktera, ali i lažnog. U novije vreme, gde interaktivnost u oblasti tehnologije postaje mnogo stvarnija, kao na primer, veštačka inteligencija u video igrama ili virtualnoj realnosti lako pobeđuje čoveka, ima sposobnost da postigne nadljudske performanse na specifičnim zadacima (npr. klasifikacija slika) [23]. Pravilna upotreba ove tehnologije može doneti mnoge pozitivne rezultate.

Iako se većinom vremena pomenuta tehnologija koristi za zlonamerni rad sa lošom namerom, ipak ima neke pozitivne upotrebe u nekoliko sektora. Deepfake kreacija više nije ograničena na stručnjake, sada je postala mnogo lakša i dostupna svima. Deepfake tehnologija ima pozitivnu upotrebu u mnogim slučajevima industrije, uključujući filmsku industriju, obrazovne medije, digitalne komunikacije, igrice i zabava, društveni mediji, zdravstvena zaštita, nauka o materijalima, kao i razne oblasti iz sveta mode i e-trgovine. U filmskoj industriji deepfake tehnologija može se koristiti pri izradi glasa za glumca koji prilikom bolesti izgubio glas, ili za ažuriranje filmskih scena umesto snimanja novih. Filmski stvaraoci će moći da rekreiraju klasične scene u filmovima, kreiraju nove filmove sa davno preminulim glumcima, koriste specijalne efekte i napredno uređivanje lica u post-produkciji i poboljšaju amaterske video snimke do profesionalnog kvaliteta. Tehnologija Deepfake takođe omogućava automatsko i realistično snimanje glasa za filmove na bilo kom jeziku, omogućavajući tako različitoj publici da bolje uživa u filmovima i obrazovnim medijima. Kao primer možemo navesti globalnu kampanju za podizanje svesti o malariji 2019. sa poznatim fudbalerom Dejvid Bekamom razbila je jezičke barijere putem obrazovnog oglasa koji je koristio vizuelnu tehnologiju i tehnologiju za menjanje glasa kako bi bio preveden na više jezika. Konkretno Dejvidov glas su uz pomoć deepfake tehnologije preveli na devet različitih jezika kako bi kampanja imala veći uticaj [41] [42]. Slično, deepfake tehnologija može da razbije jezičku barijeru u video konferencijskim pozivima tako što će prevesti govor i istovremeno menjati pokrete lica i usta kako bi se poboljšao kontakt očima i učinilo da se čini da svi govore isti jezik. Tehnologija koja stoji iza deepfake-a omogućava video igre za više igrača i virtuelne svetove u kojima je omogućeno dopisivanje sa povećanom teleprisutnošću, prirodnim zvukom i izgledom. Ova tehnologija je korišćena za stvaranje novih umetničkih dela, angažovanje publike i pružanje jedinstvenih iskustava. Dalijev muzej u Sankt Peterburgu na Floridi pružio priliku svojim posetiocima da upozna Salvadora Dalija i da se interaktivnije bavi njegovim životom kako bi upoznao ovu sjajnu ličnost putem veštačke inteligencije. Jedna od

zanimljivosti jeste da posetioci muzeja mogu napraviti sliku sa poznatim slikarom [43]. Deepfake tehnologija se sada koristi i u reklamne i poslovne svrhe. Takođe, koriste se kako bi se napravile kopije poznatih umetničkih dela (delo Monaliza). GAN metoda se može koristiti u različitim oblastima za pružanje realističnih iskustava kao što je u sektoru maloprodaje, gde bi bilo moguće videti pravi proizvod ono što vidimo u prodavnici fizički. Duboki generativni modeli su takođe pokazali velike mogućnosti razvoja u zdravstvenoj industriji. Da bi se zaštitili stvarni podaci pacijenata i istraživački rad umesto deljenja stvarnih podataka, imaginarni podaci bi se mogli generisati pomoću ove tehnologije [27].

Uz brojne navedene pozitivne strane deepfake tehnike, nažalost, u svetu se sve više koristi za zlonamerne radnje, kao što je već napomenuto. Postoji veliki broj aplikacija koje se koriste za zlonamerni rad protiv bilo kog ljudskog bića, posebno protiv javnih lica i političkih lidera. Deepfake sadržaji, koji bi mogli služiti za zabavu, ponekad se koristi za osvetu, ucenu, krađu identiteta nekoga i još mnogo toga... Deepfake tehnika predstavlja veliku pretnju društvu, političkom sistemu i biznisu, a neki od negativnih uticaja su sledeći [10]:

- vrše pritisak na novinare tako što plasiraju lažne vesti u javnost,
- ugrožavaju nacionalnu bezbednost širenjem propagande i mešanjem u izbore,
- narušavaju poverenje građana plasirajući lažne informacije od strane vlasti,
- postavljaju pitanje o sajber bezbednosti ljudi i organizacija.

Pitanja sajber bezbednosti predstavljaju jednu od pretnji koja dolazi od strane deepfake tehnologije [44]. Korporativni svet je već izrazio interes da se zaštiti od virusnih prevara, pošto bi se deepfake tehnologija mogla koristiti za manipulaciju tržištem i akcijama, na primer, tako što će se prikazati direktor koji govori rasističke ili mizoginističke uvrede, najavljuje lažno spajanje, daje lažne izjave o finansijskim gubicima ili bankrot, ili njihovo prikazivanje kao da čine zločin. Postoje hiljade video snimaka Deepfake-a i većina njih su video snimci žena čiji se likovi koriste za neprimereni sadržaj bez njihove dozvole [45]. Najčešća upotreba deepfake tehnologije je pravljenje neprimerenih snimaka poznatih glumica i ona se iz dana u dan ubrzano povećava, posebno kod holivudskih glumica [46]. Još jedna najzlonamernija upotreba deepfake-a je eksploatacija svetskih lidera i političara tako što se prave lažni video snimci o njima.

Jedna od ključnih oblasti na koje utiče veštačka inteligencija u digitalnim medijima je medijska produkcija, gde postoji dvosmerni odnos moći između proizvodnje i potrošnje. Ključno pitanje ko odlučuje šta će biti proizvedeno, često se dešavalo kroz procese koji su pokušavali da razumeju publiku i pruže joj sadržaj koji odgovara njihovim interesovanjima. Međutim, u eri digitalnih medija, takvi procesi su automatizovani i pružaju trenutni uvid sa neuporedivim detaljima. Mogućnost plasiranja sadržaja određenoj grupi ljudi, deepfake dodatno pojačava sadržaj koji je oblikovan tako da postane privlačniji za pojedince. Naročito pošto je novinarstvo u velikoj meri zavisno od pažnje javnosti, a u eri društvenih medija, to se radi u realnom vremenu, potrebno je istražiti odnose moći između medijskih platformi, društvenih platformi i drugih zainteresovanih strana. Ovo takođe podrazumeva da novinari moraju biti u stanju da provere autentičnost tvrdnji, i kao takvi, moraju da imaju neophodne alate i potencijalno stručnost da uoče deepfake video snimke i slike. Jedan od ključnih aspekta je automatizovana proizvodnja medija koju pokreće veštačka inteligencija. Ona je pokazala svoju sposobnost da kreira tekst, slike i video zapise, na osnovu onoga što uči iz dostupnih izvora na digitalnim platformama, na primer, društveni mediji kao što su Fejsbuk, Tviter, Instagram, itd [47]. Duboko učenje je omogućilo ovaj značajan napredak u napretku obrade prirodnog jezika *(NLP)*, uključujući mogućnost sumiranja tekstova, kao i reprodukcija glasova (npr. pripovedanje) na osnovu ograničenih dostupnih uzoraka (npr. uzorak govora od 5 sekundi), preko kojih se mogu kreirati deepfake, na nivou koji čak i ljudi smatraju ubedljivim. U vremenu gde su digitalne informacije lako dostupne i mogu se kopirati, moguće je kreiranje automatizovanih vesti (uključujući govorni jezik). Ovo ima značajne efekte na medijsku produkciju, jer složeni sadržaj sada može da kreira veštačka inteligencija, s obzirom na adekvatne izvore, na primer, javno dostupan uzorak glasa iz govora i fotografija sa društvenih medija, može dovesti do stvaranja realističnog deepfake videa. Novinarstvom se danas ne bave samo novinari, već i pojedini građani. Novinarstvo vođeno građanima je evidentno u digitalnim medijima pa čak i u onim vodećim, gde sve više vidimo sadržaj koji generišu korisnici (koji može uključivati deepfake). Ovo se često predstavlja i široj publici, na primer, putem televizije ili društvenih kanala velikih organizacija, međutim često se propušta provera istinitosti vesti i odgovornost se u velikoj meri prebacuje na pojedinca, na primer, jednostavnom napomenom kao što je „video ili fotografija snimljena od strane građanina“ [23]. Pojava Deepfake-a ima značajan uticaj na medije jer će njihova produkcija sigurno biti zloupotrebljena prema agendi njenih kreatora.

Zastupljenost veštačke inteligencije u celini i njene mogućnosti da bude deo digitalnih medija prikazuje koliki uticaj može imati na širu javnost. Digitalni mediji su svakako promenili način na koji ljudi komuniciraju, a sada sa veštačkom inteligencijom, čini se da će takvi odnosi mogu biti i sa negativnim posledicama. Masovni medijski efekat je promena u međuljudskim odnosima ili uopšte u društvu [23]. Era deepfake-a u kojoj se senzacionalistički, nepošteni ili čak izmišljeni sadržaji propagiraju uglavnom putem društvenih medija je uveliko prisutna [48], zato je od suštinske važnosti raditi na unapređenju algoritama za detekciju manipulisanih video materijala. Tradicionalni načini razmišljanja, obuhvaćeni popularnim izrekama kao što su „videti je verovati“, „verujem u ono što vidim“, „slika vredi hiljadu reči“, biće sve više osporavane. Uloga medija je ključna za upravljanje i demokratizaciju pošto aspekti kao što su zajednica i društveni mediji (gde se mogu koristiti deepfake video snimci i slike) utiču na ključne oblasti kao što su siromaštvo, nejednakost i društvo u celini [23]. Platforme društvenih medija su mesta gde se deepfake video snimci pretežno distribuiraju i kao takvi postaju sastavni deo njihove složene dinamike. Postojale su različite medijske tehnološke inovacije, kod kojih su društveni mediji i pametni telefoni ključni u digitalnoj eri. U ovoj eri, falsifikovanje sadržaja koji odlaze u medije nisu noviteti, ali činjenica je da svako sa niskim tehničkim veštinama lako može izraditi deepfake video ili sliku i plasirati je široj javnosti i tako uticati na medije i društvo. Tehnički upućeni građani i stručnjaci u određenim oblastima mogu procenjivati realnost video snimka.

2.4. Metode prepoznavanja deepfake tehnike

Postoje četiri načina za borbu protiv Deepfake-a [10]:

- zakonodavstvo i regulativa,
- korporativne politike i dobrovoljne akcije,
- obrazovanje i obuka i
- tehnologija protiv dubokog lažiranja koja uključuje otkrivanje deepfake-a, autentifikaciju sadržaja i prevenciju deepfake-a.

Zakonodavstvo i regulativa su očigledna sredstva protiv dubokih lažiranja. Krivični zakoni treba da pokriju zakon protiv klevete, lažiranje identiteta ili lažno predstavljanje izvedeno pomoću deepfake tehnike. Stručnjaci za pravo aktivno traže rešenja u okviru zakona za kontrolu deepfake-a od njegovog pojavljivanja. Prava izvođača su atraktivno rešenje za regulisanje ovog problema, ali deepfake dovodi u pitanje njihov obim primene. Razlog za to je

što deepfake koristi sadržaj zaštićen pravima izvođača na način koji nije bio predviđen od strane kreatora politike intelektualne svojine [49].

Korporativne politike i dobrovoljne akcije mogu da obezbede efikasnije alate protiv dubokih lažiranja. Na primer, političari se mogu obavezati da neće koristiti nedozvoljene taktike digitalne kampanje ili širiti dezinformacije kao što su deepfake u svojim izbornim kampanjama. Kako navodi institut za slobodu govora *(Institut for free speech)* postoje pravni lekovi koji bi za potencijalnu štetu od deepfake-a i oni spadaju državni zakon o kleveti. Takođe postoje i problemi koje je potrebno dodatno urediti u pravnom smislu. Zakoni se razlikuju u zavisnosti od države gde se primenjuje, ali većinski problemi su sledeći [50].

- Zakoni koji imaju za cilj regulisanje ili zabranu lažiranja u kontekstu političkog govora verovatno će biti u suprotnosti sa amandmanom koji garantuje slobodu govora.
- Standardi i definicije za ono što predstavlja deepfake i generalno manipulisani video materijali su inherentno subjektivni i nejasni
- Praktično sve političke komunikacije su na neki način uređene, regulisanje ove prakse „guši“ slobodu govora.
- Propisana odricanja od odgovornosti će označiti poruku govornika na način da je slušaoci prihvataju kao lažnu

Obrazovanje i obuku u manipulisanih video materijala je potrebno uključiti još u ranim danima školovanja. Odabir pravih tehnologija koje odgovaraju ishodima učenja učenika u današnjim učionicama koje poseduju nove tehnologije postavlja pred nastavnike višestruke izazove u pravljenju programa nastave. Kako je učenicima obezbeđen širok pristup informacijama sa raznih veb platformi, nastavnici sada imaju veći zadatak da usmere učenike i studente na autentičnu i originalnu literaturu. Shodno tome predavači trebaju da usmeravaju učenike i studente na prave alate kojima se može potvrditi autentičnost i validnost pronađenih podataka na internetu [51].

Ove veštine takođe treba promovisati među starijom populacijom koja je manje upućena u tehnologiju. Razlog za to je taj što ljudi treba da budu u stanju da kritički procene autentičnost i društveni kontekst videa koji možda žele da vide, kao i verodostojnost njegovog izvora (odnosno, ko je podelio video i šta on kaže) , kako bi se razumela prava namera video snimka. Takođe je važno zapamtiti da kvalitet nije pokazatelj autentičnosti videa. Kako

deepfake tehnologija sve više razvija, potreban je manji broj slika stvarnih lica za kreiranje deepfake slika i video snimaka. Svako ko javno objavljuje svoje fotografije i snimke rizikuje da njegovo lice bude deo deepfake tehnike. Tehnologije za detekciju deepfake videa treba da obezbede:

- otkrivanje deepfake tehnike
- autentifikaciju sadržaja
- sprečavaju korišćenje sadržaja za kreiranje deepfake videa i slika [10].

Međutim, sam razvoj tehnologije za detekciju deepfake upotrebe nekad nije sasvim dovoljan. Organizacije takođe moraju usvojiti ove tehnologije, na primer, vlada svake države može se suoči i pomogne u zaštiti svojih građana od deepfake tehnologije. Medijski forenzičari su predložili suptilne pokazatelje za otkrivanje deepfake tehnologije u videima, koje mogu biti od velike koristi bilo kojoj osobi koja nije obučena za prepoznavanje ovakvog vida manipulacije videima. Neki od predloga su: kolebanje lica, svetlucanje i izobličenje; talasanje u pokretima osobe, nedoslednosti sa govorom i pokreti usta; abnormalni pokreti fiksnih predmeti kao što je stalak za mikrofon; nedoslednost u osvetljenju, refleksije i senke; zamućene ivice; uglovi i zamućenje crta lica; nedostatak disanja; neprirodan pravac očiju; nedostaju crte lica kao npr. poznati mladež na obrazu; mekoću i težinu odeće i kose; previše glatka koža; nedostatak kose i detalji o zubima; neusklađenost u simetriji lica; nedoslednosti u nivou piksela; i čudno ponašanje pojedinca [10]. Kako je ljudima sve teže da prepoznaju pravi video zapis od lažiranog, veštačka inteligencija je idealan instrument za otkrivanje deepfake efekta u videima. Na primer, algoritmi ovih alata mogu da analiziraju obrasce neujednačenosti (PRNU) na snimku, to jest, nesavršenosti jedinstvene za svetlosni senzor određenih modela kamere, ili biometrijske podatke kao što je protok krvi naznačen suptilnim promenama koje se javljaju na licu osobe. Takođe, ovakvi alati mogu da analiziraju video zapise na osnovu okvira po kadar kako bi pratio znakove falsifikovanja, ili da pregleda ceo video odjednom da bi ispitao meke biometrijske potpise, uključujući nedoslednosti u potvrđenim odnosima između pokreta glave, obrazaca govora i izraza lica, kao što su kao osmeh, da utvrdi da li je video izmanipulisan. Detekcija deepfake video snimaka je posebno važna za medijske kampanje, kojima je, pored provere tačnosti informacija koju prenosi video, važno i otkrivanje autentičnosti video zapisa koji će se plasirati u javnost.

Često je veoma teško, a ponekad i nemoguće otkriti sadržaj deepfake-a od strane čoveka koji nije obučen za prepoznavanje deepfake-a. Potreban je viši nivo stručnosti da bi se otkrile nepravilnosti u deepfake video snimcima. Postoji nekoliko predloga za pristup uključujući detekciju, forenzičku, autentifikaciju kao i regulaciju za borbu protiv deepfake-a [27]. Stručnjaci kažu da je deepfake video kreiran algoritmom, dok originalni video pravi stvarna kamera, tako da je moguće detektovati deepfake iz postojećih tragova. Postoje i neke anomalije kao što su nedoslednosti osvetljenja, iskrivljenje slike, glatkoća u oblastima i neobične formacije piksela koje bi mogle pomoći da se otkrije deepfake [27]. Prepoznavanje deepfake-a Korshunov [52] opisao je proces koji se koristi za pronalaženje nedoslednosti usred vidljivog pokreta usta i glasa u snimku. Takođe moguće je primeniti nekoliko pristupa uključujući jednostavnu analizu glavnih komponenti *(PCA),* linearnu diskriminantnu analizu *(LDA)*, metriku kvaliteta slike *(IKM)* i mašinu za vektorsku podršku *(SVM)*. Baza podataka VidTIMIT, je javno dostupna baza video snimaka, koja se koristi za generisanje deepfake video snimaka sa kombinacijom različitih karakteristika tehnike kreiranja deepfake-a, uključujući zamenu lica, pokrete usta, treptanje očiju itd. Nakon korišćenja ovih video snimaka za verifikaciju nekoliko metoda detekcije deepfake, otkriveno je da nekoliko tehnika identifikacije zamene lica ne uspevaju da otkriju lažni sadržaj. Autori u [53] su imali za cilj da automatski i pravilno otkriju falsifikat lica koristeći najnovije tehnike dubokog učenja uz pomoć neuronske mreže. Forenzički model se suočava sa ekstremnim poteškoćama u otkrivanju falsifikata kada su izvorni podaci napravljeni *CNN* i *GAN* metodom dubokog učenja [54]. Da bi se izbegli problemi prilagodljivosti, predložen je metod koji podržava generalizaciju i koji bi mogao da locira manipulisani video na lakši način. Ekraam Sabir i dr. [55] pokušao je da otkrije zamenu lica koju generiše nekoliko dostupnih softvera kao što su Deepfake, *face2face*, *Faceswap* korišćenjem konvencionalnih mreža i rekurentne jedinice.

Kada se govori o deepfake detekciji, mogu se pronaći očigledne stvari koje nam mogu reći o „lažnosti“ videa/fotografija. Postoje neki uobičajeni indikatori koji se mogu proveriti [31]:

- Previše glatka koža, nedostatak detalja o koži – ovi pokazatelji su posledica jednog problema u deepfake algoritmima: niske rezolucije sintetizovanih lica. Ali, ponekad otkrivanje može biti veoma teško, posebno zbog šminke na jednom od dva lica. Originalni deepfake algoritam generiše lica veličine 64x64 piksela tako da obično treba da im promenimo veličinu. Sada, neki od

algoritama mogu proizvesti 128x128 ili čak 256x256 lica, ali čak ni takve veličine ne mogu biti dovoljne za dobar deepfake video.

- Neusklađenost boja između sintetizovanog lica i originalnog lica - ovaj indikator se može koristiti za prepoznavanje ljudi na deepfake videima, ali ponekad takve nepodudarnosti mogu biti veoma teško otkriti okom, ali lake u programima za detekciju deepfake-a.
- Vidljivi delovi originalnog lica ili vremensko treperenje - kada algoritam za zamenu lica dobije nepravilan izbor regiona lica možemo videti artefakte originalnog lica ili čak celo originalno lice treperi. Možda je to samo jedan kadar od celog jednosatnog videa.
- Položaj glave. Prilikom provere autentičnosti video materijala položaj glave može igrati ključnu ulogu. Maska koja se generiše uz pomoć veštačke inteligencije i koja pokriva deo lica u delovima gde se spaja sa originalnim ostatkom glave predstavlja težak zadatak. Ukoliko se položaj glave često i brzo pomera u manipulisanom video materijalu, pomenuti deo gde se spajaju maska i ostatak tela je teže generisati a da izgleda prirodno. Algoritmi za detekciju su obučeni da prepoznaju prostorne i vremenske karakteristike a pomeranje glave olakšava proces detekcije.
- Artefakti na malim pokretnim delovima – zbog ograničenja rezolucije, deepfake algoritam ne može da proizvede male pokretne delove dobrog kvaliteta. Zbog toga ponekad možemo videti artefakte na dlačicama, obrvama, trepavicama ili nekim malim defektima kože.
- Brzina treptanja – indikator koji je bio veoma koristan na samom početku popularnosti algoritama za zamenu lica. Zbog malih skupova podataka fotografija i veoma male količine slika sa zatvorenim očima, deepfake nije mogao da proizvede lice koje treperi, pa se stopa treptanja smanjuje. Sada su nove verzije algoritama rešile takav problem, tako da više nije od velike pomoći.
- Artefakti izobličenja lica – jedan od najboljih pokazatelja lažnih video zapisa, generisani algoritmima sa izlazom lica niske rezolucije (64x64 ili 128x128). Nakon što se tako mala slika sintetiše, treba je afino transformisati. Tako da se neki artefakti mogu jasno videti. Kao još jedan plus takvog indikatora je to što nam nisu potrebni deepfake skupovi podataka za obuku modela. Možemo samo koristiti algoritme za detekciju lica i napraviti neke afine transformacije u

njemu. Indikator artefakata izobličenja lica je možda najbolji izbor trenutno, ali kada se pojave novi algoritam za zamenu lica i tehnologije i kada se sintetišu slike lica visokog kvaliteta, to može postati beskorisno.

- Obrasci ponašanja ljudi – mogu biti korisni kada govorimo o tehnikama lutkarskog majstora i sinhronizacije usana u Deepfake-u. Možemo da uzmemo uobičajeno ponašanje osobe, dobijemo neke obrasce i pokušamo da identifikujemo sličnost uobičajenog i video ponašanja. To je, možda, najbolji pokazatelj lažnih video zapisa, ali je veoma teško koristiti takav indikator na fotografijama i može otkriti samo lažne sa osobom čiji su obrasci ponašanja snimljeni.

Deepfake video snimci su sve štetniji po ličnu privatnost i socijalnu sigurnost. Predložene su različite metode za otkrivanje izmanipulisanih video zapisa. Rani pokušaji su se uglavnom fokusirali na nedosledne karakteristike uzrokovane procesom sinteze lica, dok sadašnje metode detekcije uglavnom ciljaju na osnovne karakteristike [12]. Kao što je prikazano u tabeli 1 (**Error! Reference source not found.**), ove metode spadaju u pet kategorija na osnovu karakteristika koje koriste. Na početku, detekcija zasnovana na opštim neuronskim mrežama se obično koristi u literaturi, gde se zadatak detekcije dubokog lažiranja smatra redovnom klasifikacijom zadataka. Karakteristike vremenske doslednosti se takođe koriste za otkrivanje diskontinuiteta između susednih kadrova lažnog videa. Da bi se pronašle prepoznatljive karakteristike, vizuelni artefakti generisani u procesu mešanja se koriste u zadacima detekcije. Nedavno predloženi pristupi se fokusiraju na fundamentalnije karakteristike, gde šeme zasnovane na otisku prsta kamere i biološkim signalima pokazuju veliki potencijal u zadacima detekcije.

Metode	Opis
Opšte metode zasnovane na mreži	U ovoj metodi, detekcija se smatra zadatkom klasifikacije na nivou okvira koji završavaju CNN
Metode zasnovane na vremenskoj doslednosti	Utvrđeno je da u deepfake video snimcima postoje nedoslednosti između susednih kadrova zbog nedostataka algoritma za falsifikovanje. Stoga se RNN primenjuje za otkrivanje takvih nedoslednosti.
Metode zasnovane na vizuelnim artefaktima	Operacija mešanja u procesu generisanja bi izazvala suštinska odstupanja slike u granicama mešanja. Za identifikaciju ovih artefakata koriste se metode zasnovane na CNN-u.
Metode zasnovane na otiscima prstiju kamere	Zbog specifičnog procesa generisanja, uređaji ostavljaju različite tragove na snimljenim slikama.

	Istovremeno, priznaje se da lica i pozadinske slike dolaze sa različitih uređaja. Dakle, zadatak detekcije može da se završi korišćenjem ovih tragova.
Metode zasnovane na biološkim signalima	GAN je teško razumeti skrivene biološke signale lica, što otežava sintezu ljudskih lica sa razumnim ponašanjem. Na osnovu ovog zapažanja, biološki signali se izdvajaju da bi se otkrili lažni video snimci.

Tabela 2 - Klasifikacija postojećih metoda detekcije [31]

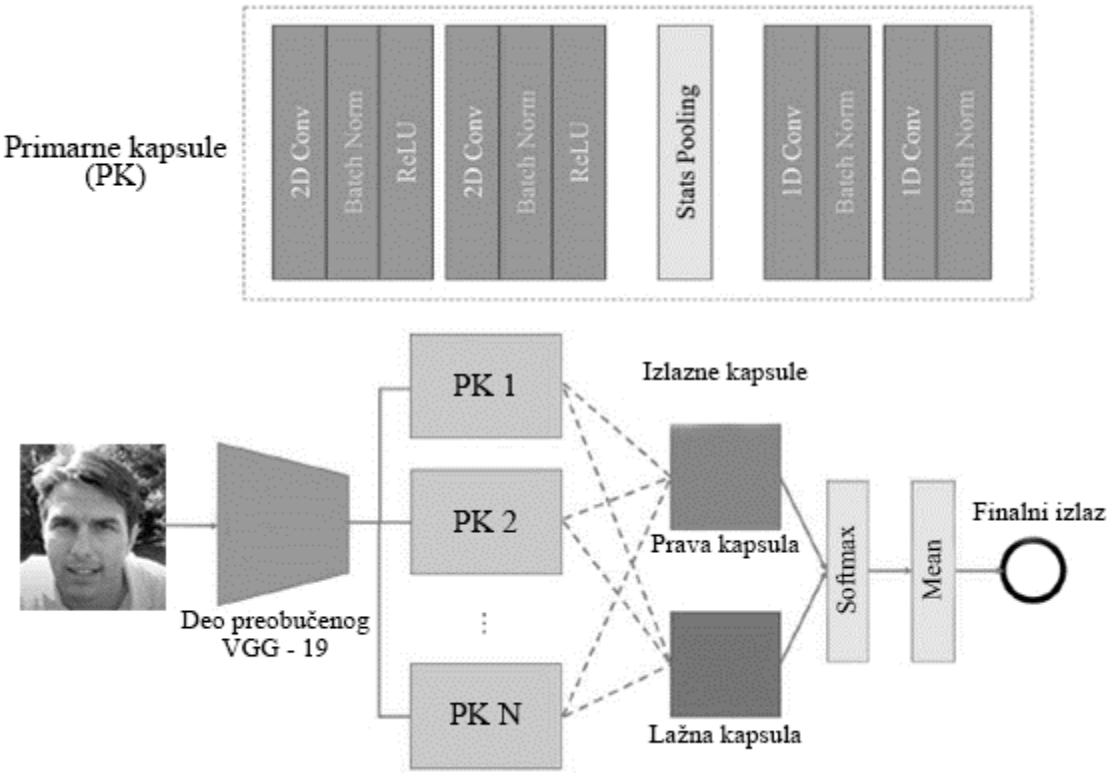
2.4.1. Opšte metode zasnovane na mrežama

Kako bi se poboljšalo otkrivanje deepfake video zapisa, primenjen je napredak u klasifikaciji slika. U ovoj metodi, slike lica ekstrahovane iz otkrivenog videa se koriste za obuku mreže za detekciju. Zatim, obučena mreža se primenjuje da bi se napravila predviđanja za sve kadrove ovog videa. Predviđanja se konačno izračunavaju usrednjavanjem ili strategijom glasanja [12]. Shodno tome, tačnost detekcije u velikoj meri zavisi od neuronskih mreža, bez potrebe da se iskoriste specifične karakteristike koje se razlikuju [12]. U narednom tekstu predstavljene su postojeće metode zasnovane na mreži u dva tipa: metode zasnovane na transferu učenja i pristupe detekcije zasnovane na posebno dizajniranim mrežama.

Metode detekcije zasnovane na mrežama trebalo bi da budu prvi i najraniji metod uveden za prepoznavanje deepfake videa. Ubrzo nakon otkrivanja prvog deepfake videa, predloženi su neki algoritmi za prepoznavanje ove vrste videa, a uglavnom su zasnovani na postojećim mrežama koje su imale dobru ulogu u klasifikaciji slika . Primeri strategija transfer učenja može se lako pronaći u ranim studijama. Kombinujući karakteristike steganalize i funkcije dubokog učenja, predložena je mreža sa dva toka za detekciju neovlašćenog pristupa. Slično, ocenjen je *XceptionNet* na skupu podataka *FaceForensic++*, kao mreža koja nadmašuje sve druge mreže u otkrivanju lažiranja. Testirana su dva postojeća modela: mali DNN (sastavljen od 6 konvolucionih slojeva i povezanog sloja) i postojeći *XceptionNet*. Rani rezultati su pokazali da najbolji metod (*XceptionNet*) pruža 93,0% preciznosti [12].

Sa pojavom velikih skupova podataka i razvojem algoritama za detekciju, više pažnje se usmerava na poboljšanje generalizacije algoritama za detekciju. Nujen [56] uveo je mrežu kapsula za poboljšanje performansi mreža za detekciju. Kao što je prikazano na slici 9, slike lica se prvo unose u prethodno obučenu VGG-19 mrežu (**Error! Reference source not found.**) . Izvučene karakteristike se zatim unose u predloženu mrežu kapsula, koja uključuje nekoliko primarnih kapsula i dve izlazne kapsule. Poklapanje karakteristika ekstrahovanih primarnim kapsulama se dinamički izračunava pomoću dinamičkog algoritma rutiranja i rezultati se

konačno rutiraju u odgovarajuću izlaznu kapsulu. Unapred obučeni VGG-19 se prvo koristi za izdvajanje karakteristika iz slika lica. Karakteristike se dalje unose u predložene kapsule, koje uključuju nekoliko primarnih kapsule i dve izlazne kapsule. Dogovor između primarnih kapsula i izlaznih kapsula je izračunat algoritmom dinamičkog rutiranja. konačno, izlaz kapsula se preslikava na verovatnoće vrednosti [12].



Slika 8 – Prikaz šeme kapsule [27]

Vizuelizacija ekstrahovanih latentnih karakteristika pokazala je da je kombinacija mreža kapsula i algoritma dinamičkog povezivanja efikasna za otkrivanje manipulacija. Međutim, mreža kapsule je imala loše rezultate kada je naišla na nepoznate deepfake video zapise, što je dokazalo da mrežama kapsula i dalje treba dodatno poboljšanje da bi se otkrili video snimci visoke vernosti. Da bi istražili mezoskopska svojstva slika, Afchar [57] je takođe predložio CNN, odnosno *MesoInception* - 4, koji se sastoji od varijante početnih modula uvedenih [58]. Predloženi pristup postigao je 98,4% tačnosti koristeći privatnu bazu podataka. Dublje mreže imaju tendenciju da postižu bolje rezultate od plitkih mreža u različitim oblastima. Razlog za dobre performanse može jednostavno biti u tome što su dizajnirane mreže dovoljno duboke [12]. U poređenju sa tradicionalnim metodama zasnovanim na učenju, Vang

[59] obraćaja više pažnje na pokrivenost neurona i interakcije, a ne na dizajn specifičnih mrežnih struktura. *FakeSpotter* koji su predložili koristi hijerarhijsko ponašanje neurona kao karakteristiku, pokazujući visoku robusnost protiv četiri uobičajena napada perturbacije. Ovo istraživanje je pružilo novi uvid u otkrivanje lažnih.

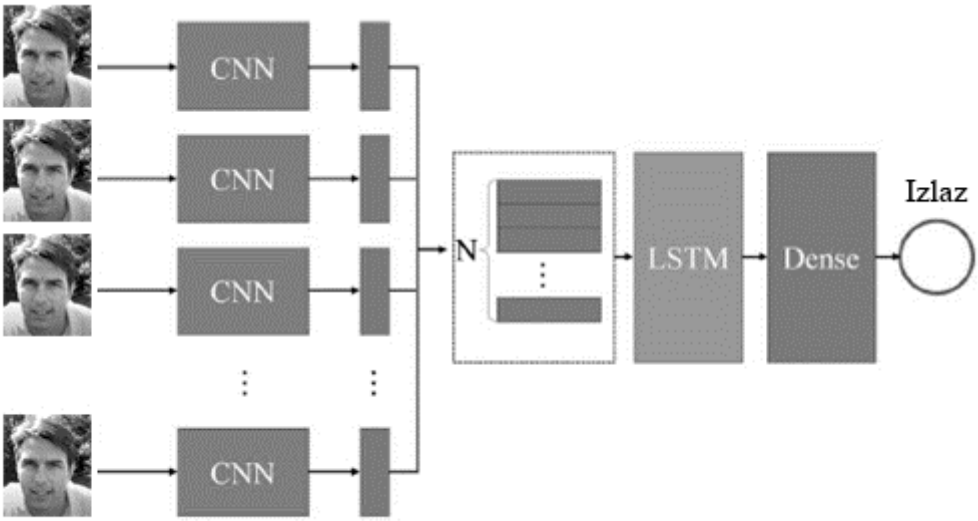
Nedostatak metoda zasnovanih na mreži je u tome što se takve metode preklapaju sa određenim skupovima podataka. U ovoj vrsti metoda, iako prilagođavanje i optimizacija strukture modela često utiču na stepen apstrakcije karakteristika, još uvek nema dovoljno relevantnosti za zadatak detekcije deepfake tehnike. Stoga se sadašnji pravac takvog rada postepeno menja. S jedne strane, dodavanjem dodatnih komponenti u model, model se može ograničiti da nauči heurističke karakteristike. U ovom slučaju je značajno smanjen značaj arhitekture modela dok dodatne komponente igraju veću ulogu. Upravo ovo je razlika između zadataka otkrivanja deepfake-a i opštih zadataka kompjuterskog vida. S druge strane, sve više metoda zasnovanih na mreži počelo je da uvodi učenje sa više zadataka, to jest, ne samo da klasifikuje stvarna i lažna lica, već i da generiše maske za neovlašćene promene na nivou piksela [12]. Vredi napomenuti da neki osnovni pravci u kompjuterskom vidu, kao što su detekcija anomalija, semantička segmentacija i metričko učenje, daju sve značajniji doprinos u ovoj oblasti.

2.4.2. Metode zasnovane na vremenskoj doslednosti

Vremenski kontinuitet je jedinstvena odlika videa. Za razliku od fotografija, video je sekvenca sastavljena od više okvira, gde susedni okviri imaju snažnu međusobnu korelaciju i kontinuitet. Kada se manipuliše video okvirima, korelacija između susednih okvira će biti uništena zbog efekta deepfake algoritma, posebno izražena u pomeranju pozicije lica i treperenja videa. Prema ovom fenomenu, istraživači su predložili više različitih metoda za detekciju primene deepfake tehnologija [27]. Prvi metod će uvesti originalnu CNN-RNN arhitekturu, a zatim demonstrirati njeno poboljšanje tokom narednih godina.

Uzevši u obzir vremenski kontinuitet u videu, Guera [60] je prvi predložio korišćenje RNN-a za detekciju deepfake videa. Tokom rada, otkriveno je da autoenkoder bio u potpunosti nesvestan prethodno generisanih lica zato što su bila generisana okvir po okvir. Ovaj nedostatak privremene nemogućnosti da detektuju pomenuta generisanja, rezultirao je u više anomalija, gde su ključni dokazi primene deepfake tehnologije izmakli detekciji [12]. Da bi se proverio kontinuitet između povezanih okvira, predložen je sistem ponavljajući *"end-to-end"*

sistem. Predloženi sistem se mahom sastojao od konvencijalne *LSTM* memorijske strukture za procesuiranje okvira sekvenci. Posebno pretrenirana incepcija V3 [58], je adaptirana da izbaci duboku reprezentaciju za svaki okvir. 2048 dimenzijini vektori ekstraktovani iz *last-pooling* sloja su primenjeni kao sekvencijalni *LSTM* input, koji karakteriše kontinuitet između sekvenci slika. Konačno, *fullyconnected* sloj i *softmax* sloj su dodati da bi se izračunala verovatnoća falsifikata testirane sekvence okvira. Eksperimenti na lično napravljenom *(selfmade)* DataSet-u pokazali su da algoritam može precizno detektuje falsifikat čak i kada je video kraći od dve sekunde. Kao što pokazuje slika 10, predloženi sistem se uglavnom sastoji od konvolucione strukture dugotrajne memorije *(LSTM)* za obradu sekvenci okvira (**Error! Reference source not found.**). Osnovni *CNN* model se prvo koristi za izdvajanje karakteristika svake slike lica u uzastopnim kadrovima. Izlazne karakteristike se zatim spajaju i koriste kao ulaz u *LSTM* mrežu, koja obrađuje karakteristike vremenske serije da bi dobila vrednost verovatnoće da li je video snimak istinit ili netačan. Iako ovo istraživanje nije pokazalo njegovu superiornost pošto nije bilo DataSetov-a velikih obima, više članaka je bilo inspirisano člankom koji je promovisao ovu tehnologiju zasnovanu na vremenskoj doslednosti.



Slika 9 - Pregled metode detekcije zasnovane na CNN-LSTM [27]

Nakon što je metoda detekcije zasnovana na vremenu pokazala svoju efikasnost, predložene su mnoge srodne studije koje su dovele do poboljšanja ove metode. Sabir [61] je koristio vremenske informacije prisutne u video prenosu da otkrije duboke lažne video snimke. Slično [60], izgrađen je model od kraja do kraja, gde je *CNN* takođe uključen u obuku koja

sledi. U međuvremenu, poravnavanje lica zasnovano na orijentirima lica i prostornoj transformatorskoj mreži se primenjuje da bi se dalje poboljšale performanse algoritma. Nakon što je metoda detekcije zasnovana na vremenu pokazala svoju efikasnost, predložene su mnoge srodne studije Iako ovakva rešenja garantuju visoku preciznost u video zapisima visokog kvaliteta, ne rade dobro na video snimcima niskog kvaliteta kada je kontinuitet između susednih kadrova poremećen operacijama video kompresije. Da bi se rešio ovaj problem, uzimajući u obzir da kvalitet lica nekih okvira nije visok, predložen je mehanizam automatskog ponderisanja kako bi se naglasili najpouzdaniji regioni prilikom predviđanja na nivou videa [62]. Eksperimenti su pokazali da se kombinovanjem *CNN-a* i *RNN-a* postiže visoka tačnost detekcije na skupu podataka *DFDC (Deepfake Detection Challenge)*. Osim robusnosti algoritama, sposobnost generalizacije je takođe neophodna za zadatke otkrivanja falsifikata. Zhao [63] je koristio optički tok da uhvati očigledne razlike u izrazima lica između susednih kadrova. Međutim, ove studije nisu pokazale jaku generalizaciju ili robusnost. Da bi se rešilo ovaj problem, predložen je novi okvir za otkrivanje manipulacije, nazvan SSTNet, koji koristi i artefakte niskog nivoa i vremenska odstupanja [64]. Druga studija koja je predložena je dobila dobru generalizaciju na više skupova podataka. U tom istraživanju, rekurentna mreža sa dve grane je primenjena da propagira originalne informacije dok potiskuje sadržaj lica. Više-pojasne frekvencije se pojačavaju korišćenjem Laplasovog Gausovog sloja kao sloja uskog grla. Nova funkcija gubitka je dizajnirana za bolje izolovanje lica kojima se manipuliše. Eksperimentalni rezultati na nekoliko skupova podataka pokazuju odlične performanse generalizacije algoritma detekcije. Ipak, šeme detekcije zasnovane na vremenu i dalje imaju mnogo prostora za poboljšanje performansi generalizacije [62]. Prebacivanje ekrana i nepoznati podaci su i dalje problemi koje treba rešiti za pristupe detekcije zasnovane na vremenu.

U poređenju sa opštim pristupima zasnovanim na mreži, metode detekcije zasnovane na vremenskoj doslednosti uzimaju u obzir kontinuitet između susednih okvira, čime se poboljšavaju performanse detekcije. Međutim, mnogi modeli teže da unište prostornu strukturu originalnih okvira prilikom izdvajanja vremenskih obeležja, dok je motivacija za projektovanje ovakvih metoda upravo izdvajanje nedoslednosti prostornih obeležja u vremenskom domenu. *CNN-RNN (Convolutional Neural Network – Recurrent neural network)* arhitekture objedinjuju karakteristike unutar okvira u vektore [12], tako da ne mogu uhvatiti prostorne karakteristike dok detektuju vremensku konzistentnost. Iako strukture kao što je 3DCNN *(3D*

Convolutional Neural Network) mogu da izbegnu uništavanje prostornih karakteristika, prekomerni parametri olakšavaju preklapanje na određeni skup podataka [12].

2.4.3. Metode zasnovane na vizuelnim artefaktima

U većini postojećih metoda deepfake-a, generisano lice mora biti uklopljeno u postojeću pozadinsku sliku, uzrokujući postojanje suštinskih neslaganja slika na granicama mešanja. Kao što je prikazano na slici 11, slike lica i pozadine potiču iz različitih izvornih slika, što dovodi do abnormalnog ponašanja sintetičke slike, kao što je granična anomalija i nedosledna osvetljenost (**Error! Reference source not found.**). Deepfake generisana slika pokazuje razliku u boji i nedoslednost rezolucije zbog nedostatka post-procesa Ovi vizuelni artefakti čine Deepfake video snimke suštinski prepoznatljivim [12]. U ovom delu biće predstavljena tri glavna vizuelna artefakta.

Slika 10 - Video frejmovi sa vizuelnim artefaktima [12]

Na osnovu zapažanja sa nedoslednošću između lica i pozadine, novi metod zasnovan na dubokom učenju su predložili Li i Liu [65]. Artefakti izobličenja lica generisani procesom mešanja korišćeni su za otkrivanje lažnih video snimaka. Kao što je prikazano na slici 2, sintetička lica su prošla afinu transformaciju da bi se poklapala sa pozama ciljanih lica. U ovom slučaju, postojala bi očigledna razlika u boji i nedoslednost rezolucije između unutrašnjeg lica i pozadine. Pošto je svrha ovde da se otkrije nedoslednost između regiona lica i pozadinske površine, negativni uzorci se generišu pojednostavljenim procesom, gde se lice podvrgava afinom iskrivljenju nazad na izvornu sliku neposredno nakon izglađivanja. Da bi se stvorili realističniji negativni primeri, koristi se konveksni poligon na osnovu obeležja lica smeđih očiju i dna usta. Takođe, informacije o bojama se takođe nasumično menjaju kako bi se povećala raznolikost treninga. Nakon toga, četiri *CNN* modela—VGG16, ResNet50, ResNet101 i ResNet152 su obučena u ovoj studiji [12]. Procenjen na osnovu nekoliko skupova podataka dostupnih deepfake video zapisa, ovaj metod je pokazao delotvornost u praksi. U

poređenju sa prethodnim metodama, ova studija se fokusira na vizuelne artefakte izazvane afinom transformacijom. Istovremeno, zbog nepostojanja dodatnih negativnih uzoraka za učešće, ovaj algoritam ne mora da se uklapa u distribuciju uzoraka deepfake video zapisa, što značajno povećava generalizaciju algoritma [66].

Uočivši da su deepfake video snimci stvoreni spajanjem sintetizovanog lica u originalnu sliku, predložena je nova metodu detekcije zasnovanu na 3D pozama glave. Tvrdili su da trenutne generativne neuronske mreže ne mogu da garantuju podudaranje orijentira, zbog čega su procenjeni 3D orijentiri na oblasti kojom se manipuliše licem bili drugačiji od 3D orijentira procenjenih iz celog područja lica [12]. U ovoj metodi, matrica rotacije procenjena korišćenjem orijentira lica sa celog lica i procenjena korišćenjem samo orijentira u centralnom regionu se izračunavaju da bi se analizirala sličnost između položaja dva vektora. Iako je eksperiment potvrdio razliku između stvarnih i lažnih položaja vektora, ova studija je izgrađena na osnovu specifičnih karakteristika koje postoje u skupu podataka koji je sami napravio koji je generisan relativno osnovnom verzijom deepfake algoritma. Dakle, ova metoda nije efikasna za otkrivanje nove verzije deepfake video zapisa kako se razvijaju deepfake algoritmi [66].

Metode zasnovane na vizuelnim artefaktima često postižu bolje performanse generalizacije jer ciljaju na opštije artefakte koji postoje u većini Deepfake sadržaja. Međutim, ovi algoritmi mogu otkriti samo određene tragove falsifikata zbog obraćanja više pažnje na određene artefakte. Sa napretkom deepfake algoritama, ovi artefakti postepeno nestaju. Ipak, pristupi zasnovani na vizuelnim artefaktima postižu bolje performanse u najnovijoj verziji deepfake skupova video podataka [12]. Takve šeme i dalje imaju veliki potencijal u zadacima otkrivanja deepfake. Istraživanja bi trebalo da budu uspostavljena kako bi se iskoristile više suštinskih karakteristika.

2.4.4. Metode zasnovane na otiscima prstiju kamere

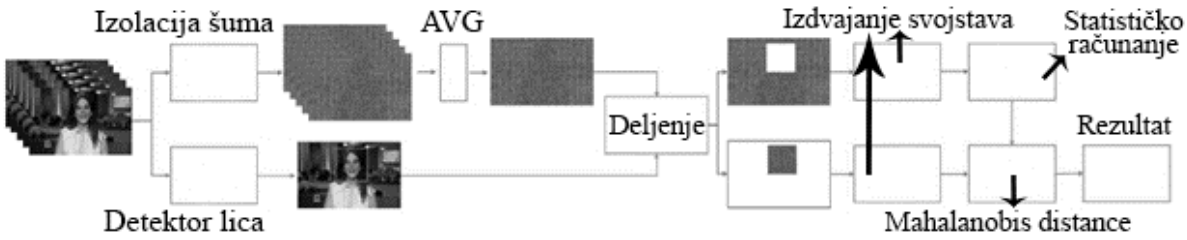
Otisci prstiju kamere su vrsta buke sa veoma slabom energijom, koja igra važnu ulogu u forenzičkim poljima, posebno u zadacima identifikacije izvora. Generalno, pristupi zasnovani na otiscima prstiju kamere prošli su kroz tri procesa: obrasci neuniformiteta foto odgovora – *PRNU (Photo Response Non-Uniformity)*, otisak šuma i nedavni obrasci video šuma.

Detekcija zasnovana na otiscima prstiju kamere potiče od forenzike slika. Uočavajući da će uređaji ostaviti različite tragove na snimljenim slikama, sledeće predloženo je *PRNU*

šum, koji se može koristiti u zadacima identifikacije kamere [27]. *PRNU* se pobuđuje zbog različite osetljivosti piksela na svetlost uzrokovanu nehomogenošću silicijumske pločice i nesavršenostima u procesu proizvodnje senzora. Zbog jedinstvenosti i stabilnosti, *PRNU* obrazac se smatra otiskom prsta uređaja, koji se može koristiti za obavljanje mnogih forenzičkih zadataka. Na osnovu ovih nalaza, Koopman je prvi predložio da se koristi *PRNU* za otkrivanje dubokih lažnih video zapisa. PRNU obrasci su verifikovani da su efikasni na malom skupu podataka. Međutim, klasifikator zasnovan na *PRNU-u* postiže mnogo nižu tačnost kada se testira u *GAN* generisanim skupovima podataka [12]. Trebalo bi sprovesti više istraživanja kako bi se verifikovala efikasnost *PRNU* obrasca u zadacima otkrivanja dubokog lažiranja.

Metode zasnovane na PRNU mogu samo da izdvoje funkcije vezane za uređaj dok potiskuju druge artefakte kamere koji postoje u procesu generisanja slike. Tragovi generisani tokom procesa akvizicije digitalne slike sastoje se od nekoliko šuma. Unutar kamere, slika se podvrgava operacijama kao što su interpolacija i gama korekcija. Izvan kamere, slika se takođe može komprimovati ili poboljšati, što će ostaviti mnogo tragova na konačnoj slici. Dakle, svaka slika ima svoje jedinstvene tragove, odnosno ostatke šuma, koji se mogu koristiti za identifikaciju izvorne kamere. Prateći ovaj pravac, uveden je otisak prsta kamere zasnovan na CNN-u pod nazivom *Noiseprint* [12]. Da bi se uklonio sadržaj scene i poboljšali artefakti vezani za model kamere, sijamska mreža je obučena korišćenjem slika koje potiču iz različitih modela kamera. U ovoj sijamskoj mreži, potpuno konvoluciona mreža, prvi put je uvedena da bi se izdvojio obrazac šuma slika. Za obuku sijamske mreže korišćeni su parovi slika sa istih ili različitih modela kamera. Na kraju procesa obuke, CNN koji se koristi u sijamskoj mreži mogao bi da se koristi za izdvajanje odgovarajućeg otiska šuma iz ulazne slike, prikazujući poboljšane artefakte modela kamere. Ovaj rad pruža nove ideje za zadatke ekstrakcije šuma otiska prsta, dodatno promovišući razvoj forenzičke oblasti slika. Nakon što je uveden koncept šumne štampe, proširen su nalazi na oblast video forenzike. Osim za identifikaciju izvora, šumna štampa je usvojena i za detekciju i lokalizaciju falsifikata. Uzimajući u obzir da je u manipulisanom video snimku manipulisani region generisan drugačije od pozadinskog regiona i stoga nosi različite šumove, utvrđeno je da se otkrivanje falsifikata može završiti korišćenjem štampanja video šuma. Kao što je prikazano na slici 12, otisci šuma izdvajaju okvir po kadru, koji se zatim usrednjavaju da bi se ukazalo na šum sadržan u video snimku (**Error! Reference source not found.**). Otisci šuma se prvo izdvajaju na dovoljnom broju video okvira. Zatim se

ekstrahovani otisci šuma usrednjavaju da predstavljaju otisak video šuma. Podeljen detektorom lica, video otisak šuma se zatim deli na region lica i region pozadine. Nakon toga, algoritam izdvaja karakteristike pozadinskog regiona i izračunava statističke informacije. Konačno, Mahalanobisovo rastojanje između karakteristika lica i pozadine se izračunava da bi se dobila konačna toplotna mapa. Regioni lica i pozadine se zatim dele da bi se izračunala sličnost. Slično, matrica prostornih ko-pojava ekstrahovanog otiska šuma se koristi za dalje izračunavanje Mahalanobisove udaljenosti između regiona lica i reference, koja se zatim koristi kao rezultat manipulacije. Algoritam je pokazao dobre performanse detekcije na skupu podataka *FaceForensic++*, iako mreža za ekstrakciju šuma nije bila obučena na njemu [12]. Međutim, pošto su otisci šuma izdvojeni iz okvira usrednjeni da predstavljaju otisak video šuma, izračunavanje otiska video šuma će biti ometano ako video ima veliko kretanje. Na ovaj način, iako je štampa sa šumom pokazala svoju efikasnost u manipulaciji slikama, njena strategija korišćenja u video forenzičkoj oblasti i dalje treba da se poboljša.



Slika 11 - Šema koja se koristi za otkrivanje deepfake tehnike u videima [12]

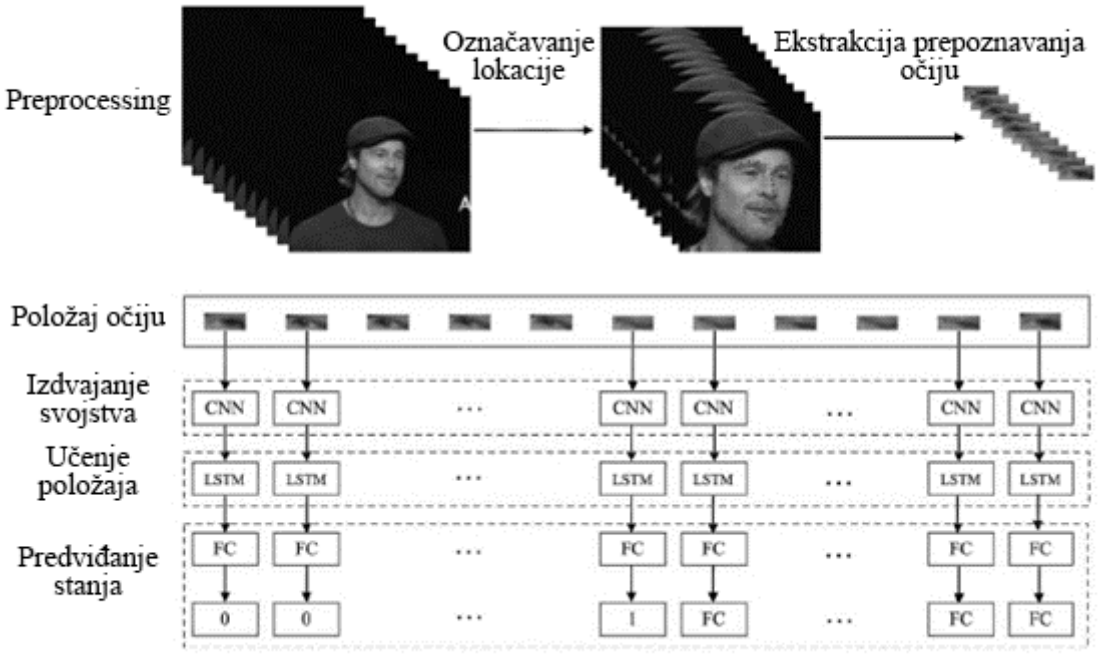
Pokazalo se da su otisci prstiju kamere efikasni u zadacima otkrivanja dubokog lažiranja. Međutim, tačna procena otisaka prstiju kamere zahteva veliki broj slika snimljenih različitim tipovima kamera. Dakle, došlo bi do smanjenja tačnosti prilikom otkrivanja slika snimljenih nepoznatim kamerama. S druge strane, metode zasnovane na otisku prsta kamere nisu robusne za jednostavnu naknadnu obradu slike kao što su kompresija, šum i zamućenje. Pošto se *GAN* slike generišu bez ikakvog procesa snimanja slike, na izlaznoj slici nema otiska prsta kamere, tako da su metode zasnovane na otisku prsta kamere veoma pogodne za otkrivanje slika koje generiše *GAN*. Međutim, nedavni radovi pokazuju da se slike mogu generisati i simulacijom otisaka prstiju kamere, čime se obmanjuju metode detekcije koje se oslanjaju na otiske prstiju kamere. Nedavna istraživanja su takođe dokazala da neuronske mreže mogu izbrisati obrazac buke. Na ovaj način, postojeće metode zasnovane na otisku prsta kamere trebalo bi da povećaju robusnost da se odupru takvim napadima [12].

2.4.5. Metode zasnovane na biološkim signalima

Detekcija zasnovana na biološkim signalima je zanimljiva šema koja se pojavila poslednjih godina. Osnovno zapažanje je da iako je *GAN* u stanju da generiše lica visokog realizma, prirodno skriveni biološki signali se i dalje ne mogu lako replicirati, što otežava sintezu ljudskih lica sa razumnim ponašanjem [67]. Koristeći prednost ovog abnormalnog ponašanja, predloženo je nekoliko studija. U ovom odeljku ćemo uvesti dva pristupa zasnovana na biološkim signalima: pristupe detekcije zasnovane na frekvenciji treptanja i detekcije zasnovane na pulsu.

Abnormalnosti sa frekvencijom treptanja su ranije identifikovane kao karakteristične karakteristike u zadacima otkrivanja dubokog lažiranja [66]. Ovo se može pripisati činjenici da deepfake algoritmi treniraju modele koristeći veliki broj slika lica dobijenih na mreži. Većina slika prikazuje ljude otvorenih očiju, zbog čega je teško generisati pogled zatvorenih očiju u manipulisanom videu. Na osnovu ovog nalaza, uveden je model duboke neuronske mreže, poznat kao dugoročni rekurentni CNN (LRCN), da bi se razlikovala stanja otvorenih i zatvorenih očiju. Da bi se izračunala frekvencija treptanja, okolni pravougaoni regioni očiju se izrezuju u novi niz ulaznih okvira nakon poravnanja lica [66]. Zatim se izrezane sekvence prosleđuju u LRCN model da bi se uhvatile vremenske zavisnosti . Kao što je prikazano na slici 13, modul za ekstrakciju karakteristika se prvo koristi za izdvajanje diskriminacionih karakteristika iz regiona ulaznog oka pomoću CNN-a zasnovanog na VGG16 okviru (**Error! Reference source not found.**). Sekvence oka se prvo izdvajaju modulom za prethodnu obradu, a zatim se unose u modul za ekstrakciju karakteristika da bi se generisale sekvence karakteristika. Modul učenja sekvenci se zatim primenjuje na analizu vremenskih sekvenci. Konačno, FC sloj se dodaje da bi se napravilo predviđanje stanja, izračunavajući brzinu treptanja. Rezultat ekstrakcije karakteristika se zatim unosi u učenje sekvence, implementirano sa RNN modelom. U fazi predviđanja konačnog stanja, dodaje se potpuno povezani sloj za izračunavanje verovatnoće otvorenih i zatvorenih stanja oka, koji se zatim koristi za izračunavanje frekvencije treptanja. Ovaj metod se procenjuje na osnovu skupova podataka koji su sami napravili, pokazujući obećavajuće performanse u otkrivanju video snimaka generisanih metodama deepfake [66]. Međutim, algoritmi za falsifikovanje mogu lako da generišu video zapise sa razumnom frekvencijom treptanja sve dok se u set za obuku dodaje dovoljno slika zatvorenih očiju. Zbog preterane pažnje na abnormalnu frekvenciju treptanja,

ova metoda više nije primenljiva za trenutne zadatke detekcije dubokog lažiranja nakon što je problem frekvencije treptanja rešen.



Slika 12 - Pregled LRCN metode [12]

Osim frekvencije treptanja, otkucaji srca su takođe pronašli razliku između stvarnih i manipulisanih video zapisa. Prethodna literatura je dokazala da se promene boje kože na video snimku mogu primeniti da bi se zaključio broj otkucaja srca. Na osnovu ovih nalaza, detektor zasnovan na biološkim signalima pod nazivom *Fake-Cachcher* je dizajniran da detektuje duboke lažne video zapise [67]. Konkretno, daljinska fotopletizmografija *(rPPG ili iPPG)* je korišćena za izdvajanje signala otkucaja srca prema suptilnim promenama boje i pokreta u RGB video zapisima. Eksperimenti su potvrdili da prostorna koherentnost i vremenska konzistentnost takvih signala nisu dobro očuvani u duboko lažnim video zapisima. Nakon statističke analize, razvijen je robustan sintetički video klasifikator zasnovan na fiziološkim promenama. Rezultati su potvrdili da *FakeCatcher* ima visoku tačnost detekcije za duboko lažne video snimke, čak i za video snimke niske rezolucije ili niskog kvaliteta. Slično, Fernandes [68] je predložio da se koriste neuralne obične diferencijalne jednačine za predviđanje brzine otkucaja srca duboko lažnih video zapisa. Velika razlika je prikazana između originalnih video snimaka i deepfake video snimaka kada se predviđanje otkucaja srca vrši odvojeno. Međutim, ovaj rad je izveo samo predviđanje otkucaja srca za deepfake video zapise, dok su nedostajali dalji eksperimenti za otkrivanje duboko lažnih. Izveden je veliki broj

radova na osnovu bioloških signala. Nedavno predloženi pristup, nazvan *DeepRhithm*, koristio je dvostruko-prostorno-vremenski mehanizam pažnje za praćenje ritmova otkucaja srca, što se pokazalo da se dobro generalizuje u različitim skupovima podataka. Slično, *DeepFakesON-Phis* predviđa brzinu otkucaja srca kroz promene u boji kože, uzimajući u obzir otkrivanje duboko lažnih video zapisa [12]. Iako su pristupi detekcije zasnovani na biološkim signalima pokazali dobre performanse na različitim skupovima podataka, prirodna mana ove vrste metoda je da se proces detekcije ne može izvoditi *end-to-end*. Takođe, na informacije koje reflektuje biološki signal ozbiljno utiče kvalitet video zapisa, tako da postoje prirodne mane i ograničen opseg primene za pristupe zasnovane na biološkim signalima.

2.4.6. Skupovi podataka i ocena učinka

Pošto sadašnje metode zasnovane na dubokom učenju u velikoj meri zavise od podataka velikih razmera, izgradnja skupova podataka visokog kvaliteta odražava važnost. Kako se duboko lažni algoritmi razvijaju, novi skupovi podataka bi trebalo da se grade za razvoj naprednih algoritama za suprotstavljanje novim metodama manipulacije. U narednom tekstu biće opisane najčešće korišćeni skupovi podataka prikazani u tabeli 3 [12]i ukratko će se predstaviti njihove karakteristike (3). Performanse detekcije na ovim skupovima podataka će takođe biti uvedene.

Skup podataka	Datum objave	Stvarno/lažno	Izvor
UADFV	11.2018.	49/49	YouTube
Deepfake-TIMIT	12.2018.	-/620	YouTube
FaceForensics++	01.2019.	1000/4000	YouTube
Google DFD	09.2019.	363/3068	Actors
DFDC-preview	10.2019.	1131/4119	Actors
DFDC	10.2019.	23.654/104.500	Actors
Celeb – DF	11.2019.	890/5639	YouTube
DeeperForensics	01.2020.	10.000/50.000	Actors

Tabela 3 - Spisak skupova podataka uključujući video manipulacije [12]

Deepfake-TIMIT, sastavljen od 620 deepfake video zapisa za 16 parova subjekata, predložen je za procenu nekoliko osnovnih algoritama za detekciju zamene lica. Takođe, skup podataka *UADFV* (Naziv DataSeta) je prikupljen da bi se otkrila stopa treptanja očiju u video snimcima [66]. Skup podataka se sastoji od 49 originalnih video snimaka sa *YouTube-a* i 49 deepfake video snimaka koje generiše FakeApp, sa tipičnom rezolucijom od 294 × 500 piksela i prosečnim vremenom od 11,14 s. Ovi sami napravljeni skupovi podataka su u velikoj meri promovisali razvoj algoritama za otkrivanje dubokog lažiranja u ranoj fazi. Takođe, skup podataka *UADFV* je prikupljen da bi se otkrila stopa treptanja očiju u video snimcima. Skup podataka se sastoji od 49 originalnih video snimaka sa *YouTube*-a i 49 deepfake video snimaka koje generiše *FakeApp*, sa tipičnom rezolucijom od 294 × 500 piksela i prosečnim vremenom od 11,14 sekundi. Ovi sami napravljeni skupovi podataka su u velikoj meri promovisali razvoj algoritama za otkrivanje dubokog lažiranja u ranoj fazi. Kao što je prikazano u tabeli 4 [12], algoritmi za detekciju dobro rade na ovim skupovima podataka (Tabeli 4).Međutim, lažni video snimci generisani u ovim skupovima podataka često su ciljani na određene algoritme detekcije i kvalitet ovih video snimaka nije dovoljan za trenutne zadatke otkrivanja. Prvi skup podataka velikog obima koji se koristi za otkrivanje dubokog lažiranja je *FaceForensic++*, predstavio je *Rossler*. Skup podataka sadrži 1000 originalnih video zapisa i 4000 izmanipulisanih video snimaka generisanih pomoću četiri različite metode falsifikovanja. Konkretno, ove metode sadrže *DeepFake, FaceSvap, Face2Face i NeuralTekture*, gde se prve dve koriste za zamenu lica, a poslednje dve se koriste za manipulaciju izrazom. Obezbeđeni su video snimci tri različita nivoa kompresije kako bi se razvile robusne metode detekcije. Nekoliko studija je potvrdilo efikasnost predloženih metoda koristeći ovaj skup podataka. Međutim, metod falsifikovanja koji se koristi za generisanje negativnih uzoraka u skupu podataka je relativno zaostao u poređenju sa trenutnim deepfake algoritmima, što uzrokuje mnoštvo vizuelnih artefakata u generisanim falsifikovanim video zapisima.

Studija	**Metoda**	**Skup podataka**	**Performamnse**	**Flop**
Zhou	Mreža sa dva toka	SwapMe and FaceSwap dataset	0.927	>5.73
Guera	CNN + LSTM	Self-made dataset	97.1%	>5.73
Yang	3D head poses	UADFV	0.974	-
Li	Eyeblink + LRCN	Self-made dataset	0.99	15.5
Ciftci	Biološki signali	Self-made deep fakes dataset	91.07%	-

Archar	MesoInception-4	Meso-data(frame-level)	91.7%	0.5
		Meso-data(video-level)	98.4%	
Nyuyen	Mreže kapsula	Meso-data(frame-level)	95.93%	>7.72
		Meso-data(video-level)	99.23%	
Li Lyu	Afekti izobličenja lica + CNN	UADFV	0.974	4.12
		Meso-data(frame-level)	0.999	
		Meso-data(video-level)	0.932	
Li	Patch jednog para CNN	Mesonet-data Deepfake-TIMIT	0.979	1.82
			1.0	

Tabela 4 - Performanse detekcije na skupovima podataka koji su napravljeni sami [12]

Kao što je prikazano u tabeli pet [12], tačnost detekcije različitih šema detekcije dostigla je čak više od 99% na skupu podataka *Face-Forensic++* (Tabela 5). Iako je *FaceForensic++* dao veliki doprinos razvoju detekcije dubokog lažiranja, on nije u skladu sa trenutnim razvojnim statusom istraživanja deepfake, pa se stoga ne može koristiti za verifikaciju performansi trenutnih algoritama za otkrivanje.

Studija	**Metoda**	**Skup podataka**	**Performanse**	**FLOP**
Bonettini	Ensemble of CNNs	FaceForensics++(c2s3)	0.9444	0.24
Nguyen	Mreža kapsula	FaceForensics++ Face2Face	93.11%	>7.72
Zhao	Optički protok	FaceForensics++ DeepFake	98.10%	0.24
Cozzolino	Noiseprint + sijamska mreža	FaceForensics++	92.14%	-
Rossler	XceptionNet	FaceForensics++(raw)	99.26%	8.42
		FaceForensics++(c23)	95.73%	
		FaceForensics++(c40)	81.00%	
Afchar	MesoInception-4	FaceForensics++(raw)	95.23%	0.5
		FaceForensics++(c23)	83.10%	

		FaceForensics++(c40)	70.47%	
Sabir	CNN + GRU + STN	FaceForensics++ DeepFake	96.9%	14.4
		FaceForensics++ Face2Face	94.35%	
		FaceForensics++ FaceSwap	96.3%	
Li	Face X-ray + multitask učenje	FaceForensics++ DeepFake	0.9912	3.99
		FaceForensics++ FaceSwap	0.9909	
		FaceForensics++ NeuralTexture	0.9927	

Tabela 5 - Performanse detekcije na FaceForensic++ skupovima podataka [12]

Da bi se dalje simulirala realistična scena, predloženi su skupovi podataka generisani novim deepfake algoritmima. Predložen je skup podataka za otkrivanje deepfake, odnosno skup podataka velikih razmera napravljen za otkrivanje deepfake. U ovom skupu podataka, 3000 deepfake video zapisa kreira 28 glumaca u različitim scenama. Nakon što su komercijalne kompanije učestvovale u istraživanju otkrivanja deepfake-a, *DFDC* je održan da promoviše razvoj otkrivanja deepfake-a [12]. Tokom izazova, uvedena su dva skupa podataka: *DFDC-preview* DataSet i *DFDC* DataSet. Skup podataka za *DFDC-preview* je izgrađen pomoću dva različita pristupa Deepdake-u, koji se sastoji od 1131 originalnog videa i 4113 odgovarajućih deepfake video zapisa. *DFDC* skup podataka je mnogo veći skup podataka koji se koristi za takmičenje u Kaggle-u, a sastoji se od preko 470 GB video zapisa (neoštećenih i manipulisanih) [12]. Vredi napomenuti da je u cilju promovisanja praktične primene algoritma za detekciju deepfake, *DFDC* više nasumičan u prikupljanju podataka, što donosi veću vizuelnu varijabilnost. Povezana istraživanja o *DFDC* skupu podataka i 3 šeme detekcije *DFDC-a* prikazana su u tabeli šest [12] (Tabela 6). Veruje se da bi *DFDC* skup podataka doneo više doprinosa razvoju zadataka detekcije dubokog lažiranja. Iako je obim trenutnog deepfake video skupa podataka bio u stanju da zadovolji potrebe algoritma za detekciju, video snimci u ovim skupovima podataka imaju očigledne vizuelne artefakte, koji nisu u skladu sa trenutnim statusom postojećih deepfake pristupa.

Studija	Metoda	Performanse	FLOP
Bonettini	Ensemble of CNNs	0.8813	>0.04
Montserrat	Mehanizam za automatsko ponderisanje	91.88%	>9.9
Tarasiou	Laka arhitektura	88.76%	-
Li	Face X-ray + multitask learning	0.892	>3.99
VertVertWM vert/vert	Ensemble of WSDAN-based networks	0.42842 (LogLoss)	18.83
NtechLab	Mixup + EfficientNet + 3D conv	0.43452 (LogLoss)	72.35 x 3

Tabela 6 - Performanse detekcije na DFDC skupovima podataka [12]

Da bi rešili ovaj problem, Li [66] je predstavio skup podataka *Celeb-DF*, generisan poboljšanim pristupom dubokog lažiranja. Problemi koji postoje u ranoj verziji duboko lažnih video snimaka, kao što su vremensko treperenje i niska rezolucija sintetizovanih lica, poboljšani su u ovom skupu podataka. Skup podataka se sastoji od 590 stvarnih video zapisa i 5639 dubokih video snimaka, koji zadovoljavaju potrebu za obukom modela. Eksperimentalni rezultati prikazani u literaturi (Tabela 7) dokazuju da je *Celeb-DF* trenutno najizazovniji skup podataka, gde je tačnost detekcije različitih metoda na *Celeb-DF* niža od one drugih skupova podataka. Napravljeno je još jedno merilo velikog obima, sastavljeno od 50.000 originalnih video zapisa i 10.000 manipulisanih video zapisa. DF-VAE, novi uslovni autoenkoder, se primenjuje za generisanje deepfake lica sa višom ocenom realizma. Studije koje koriste *DeeperForensics* pokazuju da je kvalitet generisanog video zapisa znatno bolji od kvaliteta postojećeg skupa podataka.

Studija	Metoda	Petformanse	FLOP
Li L.	Face X-ray + multitask učenje	0.8058	>3.99
Ciftci	Detekcija biološkim signalima	91.50%	-
Tarasiou	Laka arhitektura	92.62%	-
Hernandez-Ortega	DeepFakesON-Phys	91.50%	0.48
Wang	Monitoring ponašanja neurona	0.668	-

Tabela 7 - Performanse detekcije na skupovima podataka Celeb-DF [12]

3. Neuronske mreže

U kreiranju manipulisanih video materijala uz pomoć deepfake tehnike ključnu ulogu ostvaruju neuronske mreže. S obzirom na to ključno je da se posvetimo neuronskim mrežama u smislu njihove podele na jednoslojne i višeslojne, takođe u ovom poglavlju disertacije biće opisane i vrste veza između neurona kao i vrste obučavanja neuronskih mreža.

U informacionoj tehnologiji *(IT)*, veštačka neuronska mreža *(artificial neural network - ANN)* je sistem hardvera i/ili softvera napravljenog po uzoru na rad neurona u ljudskom mozgu. *ANN* - koje se jednostavno nazivaju i neuronske mreže - su razne tehnologije dubokog učenja, koje takođe spadaju pod okrilje veštačke inteligencije (*artificial intelligence – AI*). Komercijalne primene ovih tehnologija uglavnom se fokusiraju na rešavanje složene obrade signala ili problema prepoznavanja šablona. Primeri značajnih komercijalnih aplikacija od 2000. godine uključuju prepoznavanje rukopisa za obradu čekova, transkripciju govora u tekst, analizu podataka o istraživanju nafte, predviđanje vremena i prepoznavanje lica. [69]

3.1. O neuronskim mrežama

Istorija veštačkih neuronskih mreža seže u rane dane računarstva. Još 1943. godine matematičari Voren Mekalok i Volter Pits izgradili su sistem kola koji je pokretao jednostavne logaritme, a koji je funkcionisao na principu ljudskog mozga. Nakon duže pauze, istraživanje je ponovo pokrenuto 2010. godine. Trend velikih podataka, gde kompanije prikupljaju velike količine podataka, i paralelno računarstvo dali su naučnicima podatke za obuku i računarske resurse potrebne za pokretanje složenih veštačkih neuronskih mreža. U 2012. godini, neuronska mreža je uspela da nadmaši ljudske performanse u zadatku prepoznavanja slike u okviru *ImageNet* takmičenja . [69]

Veštačke neuronske mreže *(ANN)* se sastoje od slojeva čvorova, koji sadrže ulazni sloj, jedan ili više skrivenih slojeva i izlazni sloj. Svaki čvor, ili veštački neuron, povezuje se sa drugim i ima povezanu težinu i prag. Ako je izlaz bilo kog pojedinačnog čvora iznad navedene granične vrednosti, taj čvor se aktivira i šalje podatke sledećem sloju mreže. U suprotnom, podaci se ne prosleđuju na sledeći sloj mreže. *ANN* obično uključuje veliki broj procesora koji rade paralelno i raspoređenih u slojeve. Prvi nivo prima sirove ulazne informacije - analogno optičkim nervima u ljudskoj vizuelnoj obradi. Svaki sledeći nivo prima izlaz sa nivoa koji mu

prethodi, a ne sirovi ulaz - na isti način na koji neuroni dalje od optičkog nerva primaju signale od onih koji su mu bliži. Poslednji nivo proizvodi izlaz sistema.

Neuronske mreže se oslanjaju na podatke da bi na osnovu njih poboljšale svoju tačnost tokom vremena. Međutim, kada se ovi algoritmi učenja precizno podese, oni predstavljaju moćne alate u računarskoj nauci i veštačkoj inteligenciji, omogućavajući nam da klasifikujemo i grupišemo podatke velikom brzinom. Zadaci u prepoznavanju govora ili slike mogu trajati nekoliko minuta u odnosu na sate u poređenju sa ručnom identifikacijom od strane ljudi. Jedna od najpoznatijih neuronskih mreža je *Google*-ov algoritam za pretragu.

3.2. Jednoslojne neuronske mreže

Perceptron je samo moderniji naziv za jednostavni model neurona s aktivacijom funkcijom praga. Bio je to jedan od prvih službenih modela proračuna neurona, a s obzirom na njegovu ključnu ulogu u istoriji neuronskih mreža naziva se i „majkom svih veštačkih neuronskih mreža”. Perceptron se može upotrebljavati kao jednostavni klasifikator u zadacima binarne klasifikacije. Metode za izučavanje težine perceptrona iz podataka, koja se naziva algoritam perceptrona, uveo je psiholog *Frank Rosenblatt* 1957. godine. Algoritam perceptrona je gotovo jednako jednostavan kao metoda klasifikatora najbližih suseda. Taj se algoritam temelji na načelu unosa jednog po jednog podatka za učenje u mrežu. Sa svakom pogrešnom klasifikacijom težina se prilagođava novoj situaciji. [70]

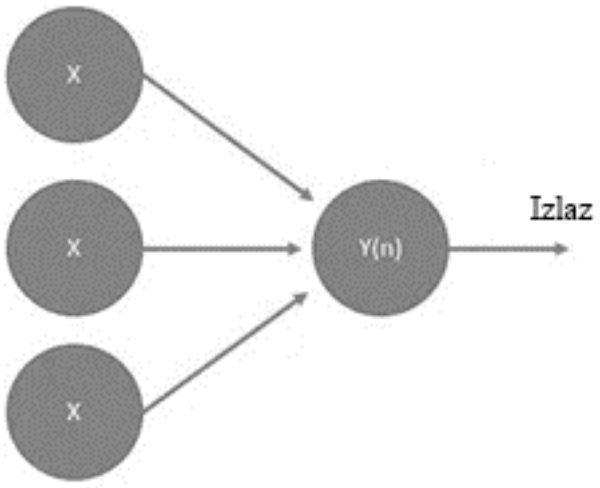
Prethodno navedeno znači da najjednostavnija vrsta neuronske mreže je jednoslojna mreža perceptrona, koja se sastoji samo od jednog sloja izlaznih jedinica. Ulazni podaci se smeštaju direktno u izlaz preko niza težina. U svakoj jedinici se izračunava suma proizvoda težina i ulaza i ukoliko je vrednost iznad nekog praga (obično 0), neuron emituje signal i uzima aktiviranu vrednost (najčešće 1). U suprotnom uzima deaktiviranu vrednost (npr. -1). Neuroni sa ovom vrstom funkcije aktivacije nazivaju se i veštački neuroni ili linearne jedinice praga *(Linear threshold units).*

Perceptron se najčešće napravi tako što koristi bilo koje vrednosti za stanja aktivacije i deaktivacije, ali se vrednost praga mora nalaziti između te dve vrednosti. Veliki broj perceptrona ima izlaze 1 ili -1 sa pragom 0. Smatra se da takve mreže mogu brže da se treniraju

u odnosu na mreže sa drugim vrednostima. Perceptroni mogu da se treniraju uz jednostavan algoritam učenja, koji se obično naziva delta pravilo. Zadatak algoritma je da računa grešku između sračunatog izlaza i primeraka izlaznih podataka. Upotrebom tih vrednosti mogu se ispraviti vrednosti težina, tako implementirajući formu gradijentnog spusta.

Zaključaj prethodno navedenog je da jednoslojna neuronska mreža predstavlja najjednostavniji oblik neuronske mreže, u kojoj postoji samo jedan sloj ulaznih čvorova koji šalju ponderisane ulaze sledećem sloju prijemnih čvorova, ili u nekim slučajevima, jednom prijemnom čvoru. Ovaj jednoslojni dizajn bio je deo temelja za sisteme koji su sada postali mnogo složeniji. Perceptron bi vratio funkciju zasnovanu na ulazima, opet, zasnovanu na pojedinačnim neuronima u fiziologiji ljudskog mozga. U izvesnom smislu, modeli perceptrona su slični „logičkim kapijama“ koje ispunjavaju pojedinačne funkcije: perceptron će ili poslati signal, ili ne, na osnovu ponderisanih ulaza. Drugi tip jednoslojne neuronske mreže je jednoslojni binarni linearni klasifikator, koji može da izoluje ulaze u jednu od dve kategorije. [71]

Jednoslojne neuronske mreže se takođe mogu smatrati delom klase neuronskih mreža unapred, gde informacije putuju samo u jednom pravcu, preko ulaza, do izlaza. Opet , ovo definiše ove jednostavne mreže za razliku od mnogo komplikovanijih sistema, kao što su oni koji za funkcionisanje koriste propagaciju unazad ili spuštanje gradijenta.



Slika 13 - Jednoslojna neuronska mreža [72]

Jednoslojna neuronska mreža će imati neprekidni izlaz, dok je standardna alternativa da navodno snabdevanje funkcioniše.

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

Sa ovom alternativom, jednoslojna mreža je mrtvo zvono za model regresije ponude, koji se široko koristi u primenjenom matematičkom modelovanju. Operacija snabdevanja se dodatno naziva i sigmoidna operacija. To je neprekidni nusproizvod koji mu omogućava da se koristi u povratnom razmnožavanju. Ova operacija je dodatno najomiljenija jer je njen nusproizvod definitivno proračunat.

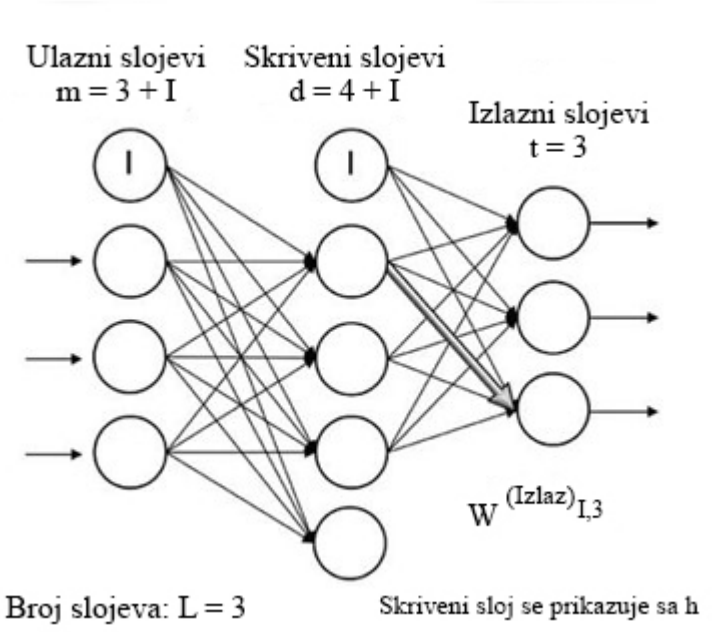
Ako aktivacija jednoslojne neuronske mreže radi u Mod1, onda će ova mreža rešiti nedostatak sa tačno jednom somatskom ćelijom.

$$f(x)=x mod 1;\ f'(x)=1$$

3.3. Višeslojne neuronske mreže

Duboko učenje se bavi obukom višeslojnih veštačkih neuronskih mreža, koje se nazivaju i duboke neuronske mreže. Nakon što je Rozenblatov perceptron razvijen 1950-ih, postojao je nedostatak interesovanja za neuronske mreže sve do 1986. godine , kada su dr Hinton i njegove kolege razvili algoritam za propagaciju unazad za obuku višeslojne neuronske mreže. Danas je to glavna tema kod mnogih vodećih firmi poput Gugla, Fejsbuka i Majkrosofta koje ulažu u aplikacije koje koriste duboke neuronske mreže. O dubokom učenju detaljnije informacije se nalaze kasnije u tekstu u posebnom poglavlju.

Potpuno povezana višeslojna neuronska mreža naziva se višeslojni perceptron (MLP).



Slika 14 - Višeslojne neurnoske mreže [73]

Ima 3 sloja uključujući jedan skriveni sloj. Ako ima više od jednog skrivenog sloja, naziva se duboka *ANN*. *MLP* je tipičan primer veštačke neuronske mreže unapred. Na ovoj slici, i-ta aktivaciona jedinica u l-tom sloju je označena kao ai(l).

Broj slojeva i broj neurona se nazivaju hiperparametrima neuronske mreže, i njima je potrebno podešavanje. Tehnike unakrsne provere moraju se koristiti da bi se pronašle idealne vrednosti za njih.

Trening prilagođavanja težine se vrši putem propagacije unazad. Dublje neuronske mreže bolje obrađuju podatke. Međutim , dublji slojevi mogu dovesti do problema sa nestajanjem gradijenta. Za rešavanje ovog problema potrebni su posebni algoritmi.

Procedura je predstavljena kao pojednostavljen prikaz višeslojnosti. Prikazano je potpuno povezana troslojna neuronska mreža sa 3 ulazna neurona i 3 izlazna neurona. Ulaznom vektoru se dodaje termin pristrasnosti.

Procedura učenja MLP-a je sledeća: [70]

- Počevši od ulaznog sloja, širiti podatke napred do izlaznog sloja. Ovaj korak je širenje unapred.

- Na osnovu rezultata izračunati grešku (razliku između predviđanog i poznatog ishoda). Greška treba da se minimizira.
- Proširiti grešku unazad. Pronaći njegov izvod u odnosu na svaku težinu u mreži i ažurirati model.

U prvom koraku potrebno je izračunati jedinicu aktivacije skrivenog sloja.

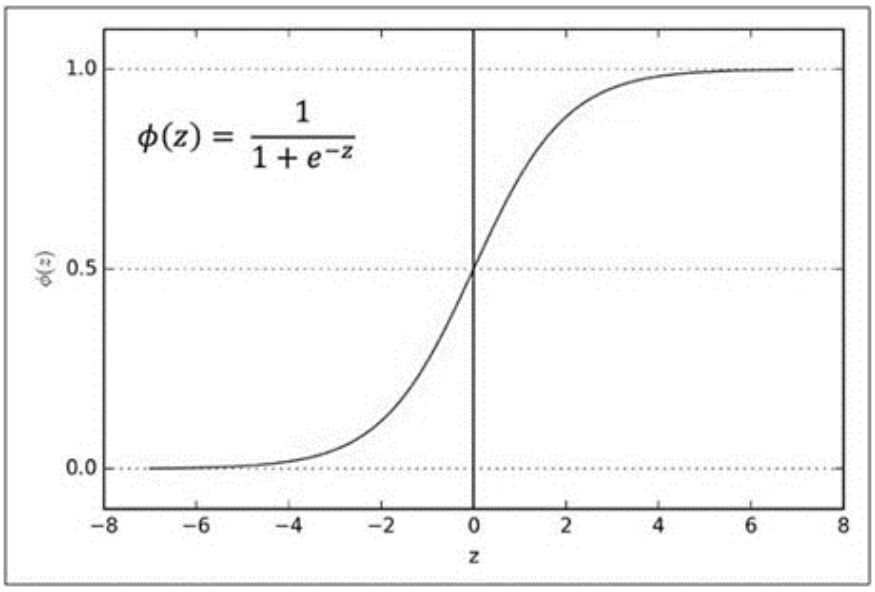
$$z_1^{(h)} = a_0^{(in)} w_{0,1}^{(h)} + a_1^{(in)} w_{1,1}^{(h)} + \cdots + a_m^{(in)} w_{m,1}^{(h)}$$

$$a_1^{(h)} = \phi\left(z_1^{(h)} \right)$$

Jedinica za aktiviranje je rezultat primene aktivacione funkcije φ na vrednost z. Mora se razlikovati da bi mogle da se nauče težine koristeći gradijentni pad. Aktivaciona funkcija φ je često sigmoidna (logistička) funkcija.

$$\phi(z) = \frac{1}{1+e^{-z}}$$

Omogućava nelinearnost potrebnu za rešavanje složenih problema kao što je obrada slike. Sigmoidna kriva je kriva u obliku slova S.



Slika 15 - Sigmoidna kriva

3.4. Vrste veza između neurona

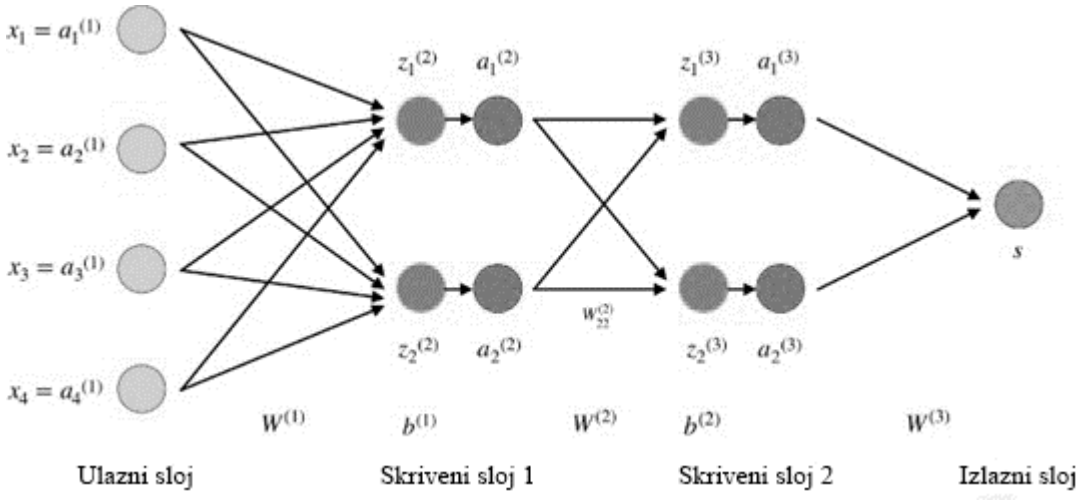
Neuronske mreže se mogu podeliti prema vrstama veza na: [74]

- slojevite – Algoritam propagacije unazad,
- potpuno povezane - Hopfildova neuronska mreža,
- *CNN – Cellular Neural Network.*

Algoritam propagacije unazad je verovatno najosnovniji građevinski blok u neuronskoj mreži. Prvi put je predstavljen 1960-ih, a skoro 30 godina kasnije (1989.) popularizirali su ga Rumelhart, Hinton i Viliams u radu pod nazivom „Učenje reprezentacija greškama koje se šire unazad“.

Algoritam se koristi za efikasnu obuku neuronske mreže kroz metod koji se zove lančano pravilo. Jednostavnim rečima, nakon svakog prolaska unapred kroz mrežu, propagacija unazad vrši prolaz unazad dok prilagođava parametre modela (težine i pristrasnosti).

4 -slojna neuronska mreža sastoji se od 4 neurona za ulazni sloj, 4 neurona za skrivene slojeve i 1 neurona za izlazni sloj.



Slika 16 - Jednostavna četvoroslojna ilustracija neuronske mreže

Neuroni u ulaznom sloju predstavljaju ulazne podatke. Oni mogu biti jednostavni kao skalari ili složeniji poput vektora ili višedimenzionalnih matrica.

Hopfildova neuronska mreža spada u tip rekurentnih mreža. Ova mreža se može iskoristiti za realizaciju nelinearne asocijativne memorije tj. memorije adresirane sadržajem. Uz sposobnost adresiranja sadržajem ova neuronska mreža ima karakterističnu mogućnost ispravljanja grešaka, i pored prisustava pogrešnih bitova na ulazu mreže na izlazu možemo dobiti tačnu informaciju. Hopfildova neuronska mreža nema fazu obučavanja, već se sinaptički koeficijenti mogu odrediti računskim putem [75]. Bitna karakteristika ove mreže je kapacitet. Kapacitet predstavlja koliki broj stabilnih stanja mreža poseduje, a da mreža zadrži prihvatljivu tačnost.

Kod **celularnih neuronskih mreža** međusobno su povezani samo susedni neuroni. Bez obzira na lokalnu povezanost, signali se prostiru i na neurone i van susedstva zbog indirektnog prostiranja informacija.

Takođe bitno je napomenuti da postoje tri pristupa obučavanju neuronskih mreža [76].

1. Nadzirna obuka
2. Delimično nadgledano obučavanje
3. Obuka bez nadzora

3.5. Smerovi prostiranja informacija

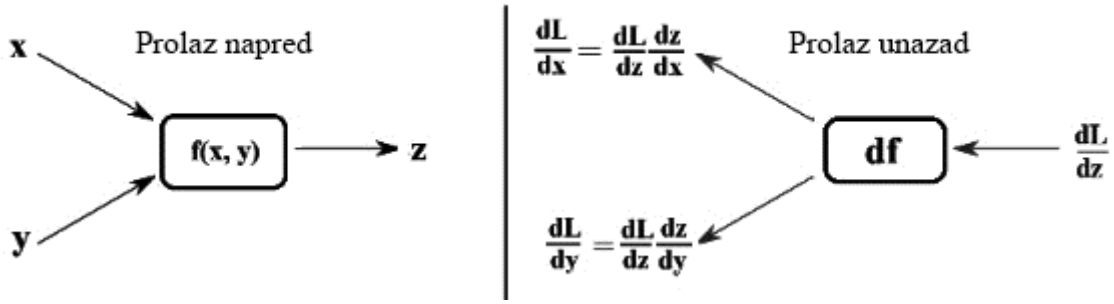
Neuronske mreže se prema smeru prostiranja informacija kroz mrežu dele na [77]:

- *Feedforward* (neprepoznati, nepredvidivi ili bespovratni),
- *Feedback* (rekurzivne ili rekurentne ili povratne).

3.5.1. Nepovratno prostiranje informacija

Backpropagation predstavlja jednu od najkorišćenijih metoda učenja koja je prvi put predstavljena 70-tih godina, iako su temelji počeli da se grade i ranije. Sam naziv navodi da se vrši nekakva povratna sprega. Proces učenja počinje jednim primerom koji prolazi kroz definisanu mrežnu arhitekturu, tzv. prolaz napred (eng. *Forward-pass*) sa početnim nasumično inicijalnim vrednostima težinskih faktora. [78]

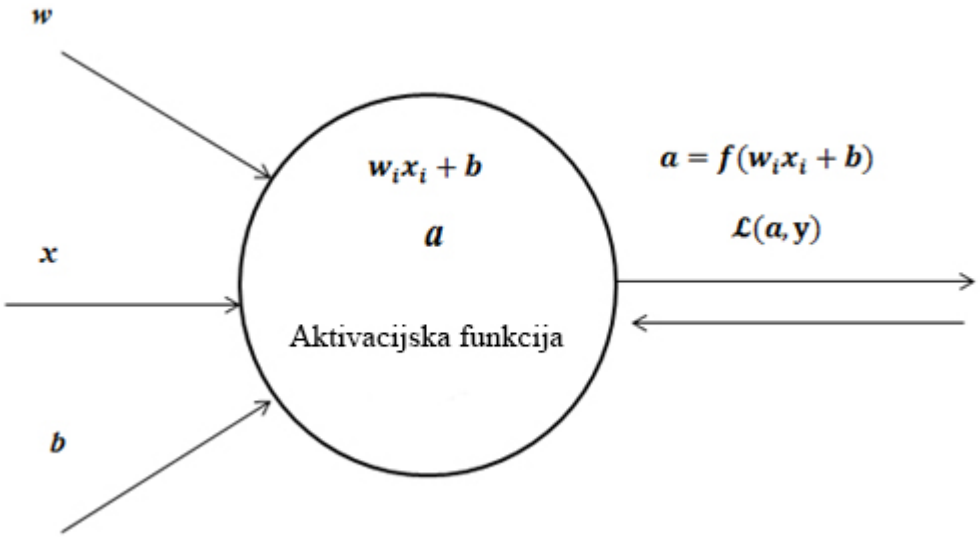
Na izlazu se vrši izračunavanje definisane *Loss funkcije* koja za rezultat daje vrednost greške između predviđene i stvarne vrednosti. Kako bi se ta greška svela na što manju vrednost ili idealnu stvarnu, potrebno je da se podešavaju težinski koeficijenti u mreži. Tu nastupa *backpropagation* algoritam gde se vrši izračunavanje prvog izvoda poslednjeg sloja po osnovu izlazne greške. Dalje se nastavlja izračunavanje na isti način svakog narednog sloja u istom smeru sve do ulaznog sloja gde se vrši ažuriranje inicijalnih težinskih faktora pomoću optimizacionog algoritma *gradient descent*.



Slika 17 - Šematski prikaz prolaza napred (Forward pass) i unazad (Backward pass)

U radnim platformama, potrebno je implementirati slojeve za *forward pass*, a dalje se automatski izvršava prolaz unazad *(Backward-pass)*. Razlika predviđene vrednosti izlaznog

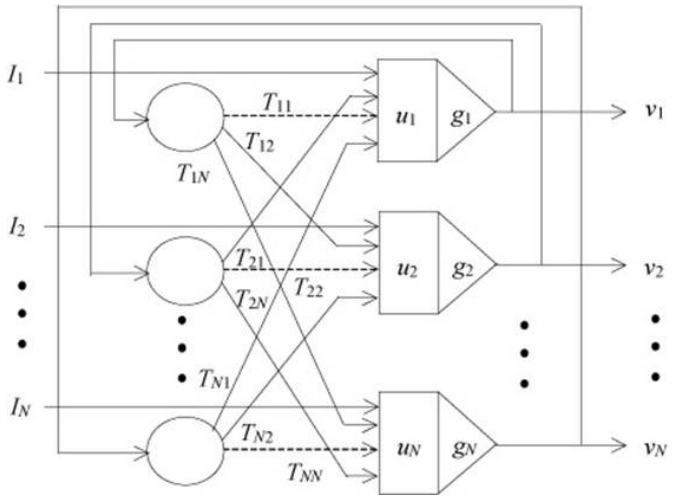
sloja, definisana *cost* funkcijom, u literaturi se još nalazi pod nazivom gubitak ili greška (uglavnom se označava sa L - *loss* ili E - *Error*). Dakle Sigmoid funkcija glasi ovako.



Slika 18 - Šematski prikaz prolaz napred za klasičnu neuronsku mrežu

3.5.2. Povratno prostiranje informacija

Specifični tip neuronskih mreža predstavljaju rekurentne neruonske mreže. Ovakve mreže poseduju posebnu karakteristiku a to je povratna veza izlaznog sloja neurona ka ulaznom sloju, pri čemu se nad signalom može primeniti određena modifikacija. Jedna od najpoznatijih mreža ovog tipa jeste Hopfildova neuronska mreža [79]. Ovakva mreža se u najvećem broju slučajeva koristi kod problema sa optimizacijom.



Slika 19 – Struktura Hopfildove neuralne mreže [75]

Hopfildova neuralna mreža je dizajnirana tako da prati model klasičnih rekurentnih mreža. Svaki neuron zasebno realizovan je kao operacioni pojačavač sa sigmoidalno rastućom funkcijom koja povezuje izlaz [80] Vi i ulaz Ui i-tog neurona. Na ovaj način mreža dobija karakteristiku nelinearnosti. Izlazne vrednosti su skalirane na opseg od 0 do 1. Funkcija prenosa (aktivaciona funkcija) za svaki od neurona data je kao [81].

$$V_{xi} = g_{xi}(U_{xi}) = \frac{1}{1+e^{-a_{xi} \cdot U_{xi}}} \quad (4)$$

gde je a konstanta koja određuje nagib karakteristike.

Shodno pravilu rekurzivnih mreža, izlazni signal i-tog neurona vodi se na sve ulaze drugih neurona sem na sopstveni ulaz, preko rezistivnih veza. Ova povezanost definisana je matricom povezanosti T =[Tij]. Pored signala koji prima od izlaznih neurona, na svaki od ulaznih neurona deluje dodatni strujni signal (*bias current*) Ii. Njime se podešava polarizacija neurona.

Izlazi neurona, *Vj*, se preko otpornika *Rij* stiču na kondezator *Ci*. Promena napona na kondenzatoru data je jednačinom stanja.

$$c_i \frac{du_i}{dt} = -\frac{u_i}{R_i} + \sum_{j=1}^{N} T_{ij} V_j + I_i$$

Napon kondezatora Ci deluje na ulazu nelinearnog diferencijalnog pojačavača, na čijem izlazu se dobijaju signali Vi i -Vi ove ćelije, prema relaciji [21], [26].

Hopfild je pokazao da će u slučaju da je pojačanje pojačavača relativno veliko (teoretski a →∞, kada je prenosna funkcija odskočna) energijska funkcija biti. [75]

$$E = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} \cdot V_j V_i - \sum_{i=1}^{N} V_i I_i$$

Za veliko pojačanje operacionog pojačavača, minimum energije u datom N dimenzionalnom prostoru se raspoređuje u 2 N rogljeva.

Relacija definiše promenu ulaznog signala i promenu energije, u svakoj od iteracija. Može se pokazati da ovako definisana mreža obezbeđuje konvergenciju ka stabilnim stanjima. Ovako postavljena mreža služi kao osnovna struktura za rešavanje optimizacionih problema. [77]

4. Duboko učenje

Same tehnike dubokog učenja koje implementiraju duboke neuronske mreže postale su jako popularne zbog povećanja kvaliteta računarskih uređaja. Duboko učenje postiže znatno veću snagu i fleksibilnost, zbog svoje sposobnosti da obradi veliki broj karakteristika određenih ne struktuiranih podataka. Algoritam dubokog učenja koristi podatke sa više slojeva. Svaki sloj je sposoban da izvlači karakteristike progresivno i prenosi ih na sledeći sloj. Početni slojevi izdvajaju karakteristike niskog nivoa i prenose u naredne slojeve, kombinuju karakteristike i formiraju potpunu informaciju što će biti detaljnije objašnjeno u daljem tekstu.

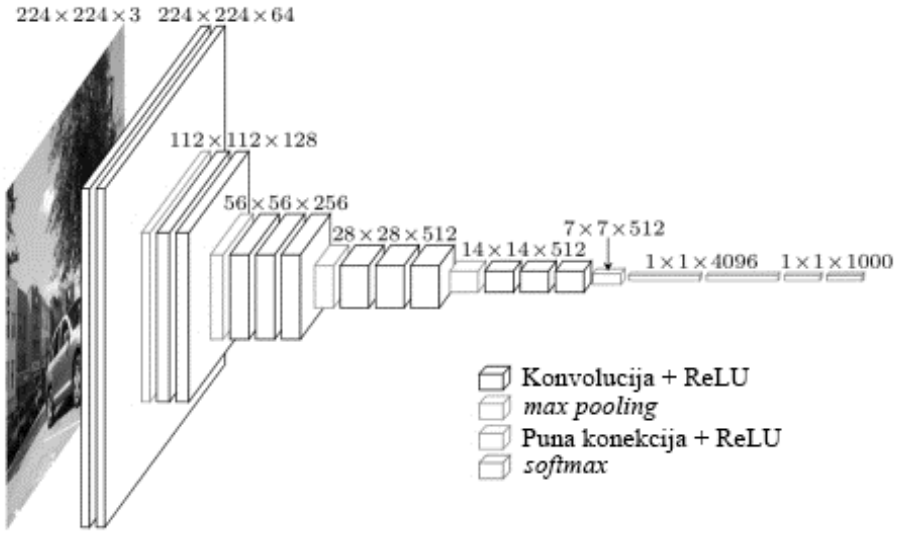
4.1. Postojeće metode zasnovane na vremenskim odlikama

U ovom poglavlju se opisuju postojeće metode zasnovane na vremenskim odlikama. Ključne su tri metode a to su VGG Net, GoogleNet i ResNet. Svaka od ovih metoda je posebno opisana i detaljno prikazana.

4.1.1. VGG Net

VGG se koristi kao skraćenica za *Visual Geometri Group*. VGG predstavlja standardnu duboku konvolucionu neuronsku mrežu (CNN) sa više različitih slojeva. Dubokom konvolucionom mrežom se naziva jer se odnosi na dubinu tj. na broj slojeva sa VGG-16 ili VGG-19. Pomenute mreže se sastoje iz 16 i 19 slojeva.

VGG arhitektura se smatra osnovom revolucionarnih modela za prepoznavanje objekata. Razvijen kao duboka neuronska mreža, VGG Net takođe svojim karakteristikama prevazilazi same osnove za mnoge zadatke i skupove podataka izvan ImageNet-a. Takođe kao takva i dalje je jedna od najčešće korišćenih i najpopularnijih arhitektura za prepoznavanje slika.



Slika 20 – VGG arhitektura neuronske mreže [82]

VGG model ili konkretno VGGNet koji koristi 16 slojeva duboku mrežu se takođe naziva VGG16 što je model konvolucione neuronske mreže koji je predložen od strane A. Zisserman i K. Simonyan sa Univerziteta u Oksfordu. Pomenuti istraživači su objavili svoj model u radu pod nazivom „Deep Inside Convolutional Networks: Cisualising Image Classification Models and Saliency Maps“ [83]

Model VGG16 funkcioniše s preciznošću od 92,7% i spada u jedan od popularnijih model. Takođe spada u jedan od prvih 5 modela po tačnosti u ImageNet-u. ImageNet je skup podataka koji se sastoji od više od 14 miliona slika koje se razvrstavaju u 1000 različitih klasa. Sa pomenutim rezultatima VGG16 je postao kao jedan od najpopularnijih modela priloženih na ILSVRC-2014. On zamenjuje filtere velikih veličina jezgra sa nekoliko filtera veličine jezgra 3x3 jedan za drugim, čineći tako značajna poboljšanja u odnosu na prethodno korišćeni AlekNet. Model VGG16 je treniran koristeći Nvidia Titan Black GPU u trajanju od više nedelja. [84]

VGGNet-16 Podržava 16 slojeva i može da klasifikuje slike u 1000 kategorija različitih objekata koji podrazumevaju kako predmete tako i životinje, ljudske oblike itd. Pored navedenog, model ima ulaznu veličinu slike 224x224. Koncept modela VGG19 je isti kao i VGG16 a jedina razlika u broju podržavanih slojeva. Brojevi 16 i 19 označavaju broj težinskih slojeva u modelu a to su konvolucijski slojevi.

VGGNet arhitektura je zasnovana na najvažnijim karakteristikama konvolucionih neuronskih mreža *(CNN)*. U nastavku slika pokazuje osnovni koncept kako *CNN* funkcioniše. VGG mreža je konstruisana sa veoma malim konvolucionim okvirima. VGG-16 se sastoji od 13 konvolucionih slojeva i tri potpuno povezana sloja**.**

Sistematičnost VGG arhitekture se zasniva na sledećim segmentima: [85]

- Ulaz
- Konvolucijski slojevi
- Skriveni slojevi
- Potpuno povezani slojevi

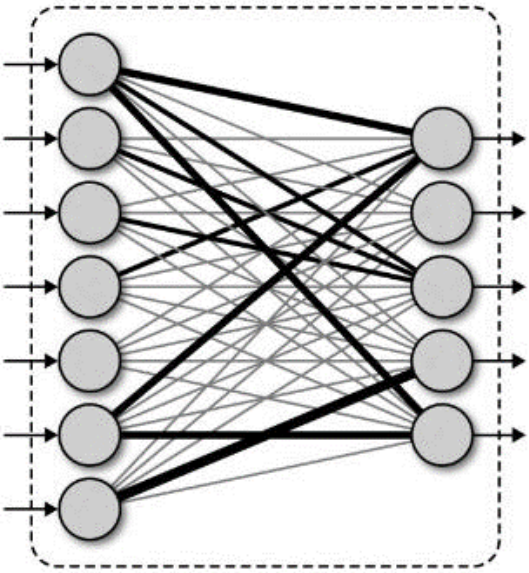
Ulaz: VGGNet uzima veličinu ulazne slike od 224 × 224. Za takmičenje ImageNet, kreatori modela su izrezali centralnu zakrpu od 224 × 224 na svakoj slici kako bi održali ulaznu veličinu slike doslednom.

Konvolucijski slojevi: VGG-ovi konvolucijski slojevi koriste minimalno prijemčivo polje, tj. 3 × 3, najmanju moguću veličinu koja i dalje hvata gore/dole i levo/desno. Takođe, postoje i 1 × 1 konvolucijski filteri koji deluju kao linearna transformacija ulaza.

Zatim sledi ReLU jedinica, što je ogromna inovacija iz AlekNet-a koja smanjuje vreme obuke. ReLU je skraćenica za funkciju aktivacije ispravljene linearne jedinice; to je linearna funkcija koja će dati ulaz ako je pozitivna; inače, izlaz je nula. Korak konvolucije je fiksiran na 1 piksel da bi se prostorna rezolucija sačuvala nakon konvolucije (korak je broj pomeranja piksela preko ulazne matrice).

Skriveni slojevi: Svi skriveni slojevi u VGG mreži koriste ReLU. VGG obično ne koristi normalizaciju lokalnog odgovora (LRN) jer povećava potrošnju memorije i vreme treninga. Takođe, ne poboljšava ukupnu preciznost.

Potpuno povezani slojevi: VGGNet ima tri potpuno povezana sloja. Od tri sloja, prva dva imaju po 4096 kanala, a treći ima 1000 kanala, po jedan za svaku klasu.



Slika 21 - Potpuno povezani slojevi

Broj 16 u nazivu VGG odnosi se na činjenicu da je to 16 slojeva duboke neuronske mreže *(VGGnet)*. To znači da je VGG16 prilično obimna mreža i da ima ukupno oko 138 miliona parametara. Čak i po savremenim standardima, to je ogromna mreža. Međutim , jednostavnost arhitekture VGGNet16 je ono što čini mrežu privlačnijom. Već posmatrajući njegovu arhitekturu, može se reći da je prilično ujednačena.

Postoji nekoliko konvolucionih slojeva nakon kojih sledi sloj za spajanje koji smanjuje visinu i širinu. Ako pogledamo broj filtera koje možemo da koristimo, dostupno je oko 64 filtera koje možemo udvostručiti na oko 128, a zatim na 256 filtera. U poslednjim slojevima možemo koristiti 512 filtera. Broj filtera koje možemo da koristimo udvostručuje se na svakom koraku ili kroz svaki stog sloja konvolucije. Ovo je glavni princip koji se koristi za dizajniranje arhitekture VGG16 mreže. Jedan od ključnih nedostataka VGG16 mreže je to što je to glomazna mreža, što znači da je potrebno više vremena za obuku njenih parametara. Zbog svoje dubine i broja potpuno povezanih slojeva, VGG16 model ima više od 533MB. Ovo čini implementaciju VGG mreže dugotrajnim zadatkom. Model VGG16 se koristi u nekoliko problema klasifikacije slika dubokog učenja, ali manje mrežne arhitekture kao što su GoogLeNet i SkueezeNet su često poželjnije. U svakom slučaju, VGGNet je odličan građevinski blok za učenje jer je jednostavan za implementaciju. VGG16 u velikoj meri nadmašuje prethodne verzije modela na takmičenjima ILSVRC-2012 i ILSVRC -2013. Rezultat VGG16 se takmiči za pobednika zadatka za klasifikaciju (GoogleNet sa greškom od

6,7%) i znatno nadmašuje pobednički podnesak ILSVRC-2013 Clarifay. Dobio je 11,2% sa eksternim podacima o obuci i oko 11,7% bez njih. U pogledu performansi jedne mreže, VGGNet-16 model postiže najbolji rezultat sa oko 7,0% greške testa, čime je nadmašio jedan GoogleNet za oko 0,9%. [86]

VGG je skraćenica za Visual Geometry Group i sastoji se od blokova, gde je svaki blok sastavljen od slojeva 2 D Convolution i Mak Pooling. Dolazi u dva modela — VGG16 i VGG19 — sa 16 i 19 slojeva. [87] Kako se broj slojeva povećava u *CNN-u*, povećava se i sposobnost modela da uklopi složenije funkcije. Dakle , više slojeva obećava bolje performanse. Ovo ne treba mešati sa veštačkom neuronskom mrežom *(ANN)*, gde povećanje broja slojeva ne mora nužno dovesti do boljih performansi.

Postavlja se pitanje, zašto se ne koristi VGGNet sa više slojeva, kao što su VGG20, ili VGG50, ili VGG100? Tu nastaje problem. Težine neuronske mreže se ažuriraju kroz algoritam preparacije unazad, što čini manju promenu svake težine tako da se gubitak modela smanjuje. On ažurira svaku težinu tako što pravi korak u pravcu u kojem se gubitak smanjuje. Ovo nije ništa drugo nego preliv ove težine koji se može naći pomoću pravila lanca. Međutim , kako gradijent nastavlja da teče unazad do početnih slojeva, vrednost se povećava za svaki lokalni gradijent. Ovo dovodi do toga da gradijent postaje sve manji i manji, čime se promene u početnim slojevima čine veoma malim. Ovo , zauzvrat, značajno povećava vreme obuke. Problem se može rešiti ako lokalni gradijent postane jedan. Ovde se pojavljuje ResNet pošto to postiže kroz funkciju identiteta. Dakle , kako se gradijent širi unazad, on se ne smanjuje u vrednosti jer je lokalni gradijent jedan.

Duboke rezidualne mreže *(ResNets)*, kao što je popularni model ResNet-50, su još jedan tip arhitekture konvolucionih neuronskih mreža *(CNN)* koja je duboka 50 slojeva. Preostala neuronska mreža koristi umetanje prečica za pretvaranje obične mreže u njen pandan rezidualne mreže. U poređenju sa VGGNet-ovima, ResNets su manje složeni jer imaju manje filtera [88]. ResNet , takođe poznat kao Residual Network, ne dozvoljava da se pojavi problem nestajanja gradijenta. Preskočene veze deluju kao nagibni superautoputevi, koji omogućavaju neometano odvijanje nagiba. Ovo je takođe jedan od najvažnijih razloga zašto ResNet dolazi u verzijama kao što su ResNet50, ResNet101 i ResNet152

4.1.2. GoogleNet

GoogleNet je 22-slojna duboka konvoluciona neuronska mreža koja je varijanta Inception Network-a, duboke konvolucione neuronske mreže koju su razvili istraživači u Google-u. GoogleNet arhitektura predstavljena u *ImageNet Large-Scale Visual Recognition Chalenge 2014 (ILSVRC14)* rešavala je zadatke kompjuterskog vida kao što su klasifikacija slika i detekcija objekata. Danas se GoogleNet koristi i za zadatke koje predstavljaju detekciju i prepoznavanje lica [89]. Zadaci u kojima se koristi GoogleNet su prikazani u tabeli ispod [90].

Zadatak	Broj radova	Udeo u procentima
Klasifikacija slika	21	17.95%
Kvantizacija	11	9.4%
Detekcija objekata	11	9.4%
Prepoznavanje objekata	7	5.98%
Adaptacija domena	5	4.27%
Klasifikacija objekta	3	2.56%
Kompresija modela	3	2.56%
Prepoznavanje lica	3	2.56%
Kontradiktorna obuka	2	1.71%

Tabela 8 - Zadaci u kojima se koristi GoogleNet

GoogleNet arhitektura je rešila većinu problema sa kojima su se velike mreže suočavale, uglavnom kroz korišćenje Inception modula.Inception modul je arhitektura neuronske mreže koja koristi detekciju karakteristika na različitim skalama kroz konvolucije sa različitim filterima i smanjuje troškove računara za obuku ekstenzivne mreže kroz smanjenje dimenzija.GoogleNet arhitektura se sastoji od 22 sloja (27 slojeva uključujući slojeve za udruživanje), a deo ovih slojeva je ukupno 9 početnih modula.

Tip	Veličina zakrpe	Veličina izlaza	Dubina	#1x1	#3x3 smanjiti	#3x3	#5x5 smanjiti	#5x5	Broj filtera	Parametri u hiljadama	MOP u milionima
Konvolucija	7 x 7 / 2	112 x 112 x 64	1							2.7	34
Max pool	3 x 3 / 2	56 x 56 x 64	0								
Konvolucija	3 x 3 / 1	56 x 56 x 192	2		64	192				112	360
Max pool	3 x 3 / 2	28 x 28 x 192	0								
Inception (3a)		28 x 28 x 256	2	64	96	128	16	32	32	159	128
Inception (3b)		28 x 28 x 480	2	128	128	192	32	96	64	380	304
Max pool	3 x 3 / 2	14 x 14 x 480	0								
Inception (4a)		14 x 14 x 512	2	192	96	208	16	48	64	364	73
Inception (4b)		14 x 14 x 512	2	160	112	224	24	64	64	437	88
Inception (4c)		14 x 14 x 512	2	128	128	256	24	64	64	463	100
nception (4d)		14 x 14 x 528	2	112	144	288	32	64	64	580	119
Inception (4e)		14 x 14 x 832	2	256	160	320	32	128	128	840	170
Max pool	3 x 3 / 2	7 x 7 x 832	0								
Inception (5a)		7 x 7 x 832	2	256	160	320	32	128	128	1072	54
Inception (5b)		7 x 7 x 1024	2	384	192	384	48	128	128	1388	71
prosečan pool	7 x 7 / 1	1 x 1 x 1024	0								
Dropuout (40%)		1 x 1 x 1024	0								
Linerar		1 x 1 x 1000	1							1000	1
Softmax		1 x 1 x 1000	0								

Tabela 9 – Konvencionalna GoogleNet arhitektura

Karakteristike GoogleNet konfiguracione tabele:

- Ulazni sloj arhitekture GoogleNet dobija sliku dimenzija 224 x 224.
- Tip : Ovo se odnosi na ime trenutnog sloja komponente unutar arhitekture.
- Veličina zakrpe: Odnosi se na veličinu prostora za čišćenje koji se koristi u slojevima konv i objedinjavanja. Prostori za čišćenje imaju jednaku visinu i širinu.
- Korak : Definiše količinu pomeranja filtera/kliznog prozora koji preuzima ulaznu sliku.
- Veličina izlaza: rezultirajuće izlazne dimenzije (visina, širina, broj mapa karakteristika) trenutne komponente arhitekture nakon što se ulaz prođe kroz sloj.
- Dubina : Odnosi se na broj nivoa/slojeva unutar komponente arhitekture.
- #1x1 #3x3 #5x5: Odnosi se na različite konvolucione filtere koji se koriste u okviru početnog modula.
- #3x3 smanjiti #5x5 smanjiti: Odnosi se na brojeve 1x1 filtera koji su korišćeni pre konvolucija.
- Broj filtera odnosi se na broj 1x1 filtera koji se koriste nakon objedinjavanja unutar početnog modula.
- Kolona parametri se na broj parametara unutar trenutne komponente arhitekture.
- Kolona MOP odnosi se na broj matematičkih operacija izvedenih unutar komponente.

Na svom početku, arhitektura GoogleNet –a je dizajnirana da bude moćna arhitektura sa povećanom računarskom efikasnošću u poređenju sa nekim od svojih prethodnika ili sličnih mreža stvorenih u to vreme [89]. Jedna metoda kojom GoogleNet postiže efikasnost je smanjenjem ulazne slike, dok istovremeno zadržava važne prostorne informacije.

Prvi na tabeli 9 koristi filter (zakrpe) veličine 7x7, što je relativno veliko u poređenju sa drugim veličinama zakrpe unutar mreže. Primarna svrha ovog sloja je da odmah smanji ulaznu sliku, ali da ne izgubi prostorne informacije korišćenjem velikih filtera. Veličina ulazne

slike (visina i širina) se smanjuje za faktor četiri na drugom sloju i za faktor osam pre nego što se stigne do prvog početnog modula, ali se generiše veći broj mapa karakteristika.

Drugi sloj ima dubinu od dva i koristi 1x1 blok, što je efekat smanjenja dimenzionalnosti. Smanjenje dimenzionalnosti kroz 1x1 blok omogućava smanjenje računarskog opterećenja smanjenjem broja operacija slojeva. Svrha ovih slojeva maksimalnog objedinjavanja je da smanje uzorkovanje ulaza dok se on šalje unapred kroz mrežu. Ovo se postiže smanjenjem visine i širine ulaznih podataka. Smanjenje veličine ulaza između početnog modula je još jedan efikasan metod za smanjenje računarskog opterećenja mreže. [83]

Prosečan sloj objedinjavanja uzima srednju vrednost za sve mape obeležja koje je proizveo poslednji početni modul i smanjuje ulaznu visinu i širinu na 1x1.

Sloj za ispadanje (40%) se koristi neposredno pre linearnog sloja. Sloj ispadanja je tehnika regularizacije koja se koristi tokom treninga da bi se sprečilo preopterećenje mreže. Tehnika ispadanja funkcioniše tako što nasumično smanjuje broj neurona koji se međusobno povezuju unutar neuronske mreže. Na svakom koraku treninga, svaki neuron ima šansu da bude izostavljen, ili bolje rečeno, da ispadne iz uporednog doprinosa povezanih neurona.

Linearni sloj se sastoji od 1000 skrivenih jedinica, što odgovara 1000 klasa prisutnih u okviru ImageNet skupa podataka. Poslednji sloj je *Softmax* sloj. Ovaj sloj koristi *softmak* funkciju, aktivacionu funkciju koja se koristi za izvođenje distribucije verovatnoće skupa brojeva unutar ulaznog vektora. Izlaz *softmak* aktivacione funkcije je vektor u kome njen skup vrednosti predstavlja verovatnoću pojave klase ili događaja. Sve vrednosti unutar vektora imaju zbir do 1. [83]

Pre nego što se istraživanje arhitekture GoogleNet privede kraju, postoji još jedna komponenta koju su implementirali kreatori mreže da bi se regulisalo i sprečilo prekomerno prilagođavanje. Ova dodatna komponenta je poznata kao pomoćni klasifikator. Jedan od glavnih problema ekstenzivne mreže je to što one imaju problem sa nestajanjem gradijenta. Spuštanje gradijenta nestajanja se dešava kada je ažuriranje težina koje proizilazi iz povratnog širenja zanemarljiva unutar donjih slojeva kao rezultat relativno male vrednosti gradijenta. Ukoliko se mreža ne održava ona prestaje da uči tokom treninga. Pomoćni klasifikatori se koriste samo tokom treninga i uklanjaju se tokom zaključivanja. Svrha pomoćnog klasifikatora je da izvrši klasifikaciju zasnovanu na ulazima unutar srednjeg preseka mreže i doda gubitke

izračunate tokom obuke nazad ukupnom gubitku mreže. Svaki od uključenih pomoćnih klasifikatora prima kao ulaz aktivacije iz prethodnih početnih modula.

Razumevanje GoogleNet arhitekture je ključno za svakog praktičara dubokog učenja koji želi da razume razvoj dubokih konvolucionih mreža u polju dubokog učenja. Postoji ogromno znanje koje se može steći ponovnim razmatranjem istraživačkih napora učinjenih nekoliko godina unazad.

4.1.3. ResNet

Duboke rezidualne mreže poput popularnog modela ResNet-50 su konvoluciona neuronska mreža *(CNN)* koja je duboka 50 slojeva. Preostala neuronska mreža (ResNet) je veštačka neuronska mreža *(ANN)* vrste koja slaže preostale blokove jedan na drugi da bi se formirala mreža.

Poslednjih godina, oblast kompjuterskog vida je pretrpela dalekosežne transformacije usled uvođenja novih tehnologija. Kao direktan rezultat ovog napretka, postalo je moguće da modeli kompjuterskog vida nadmaše ljude u efikasnom rešavanju različitih problema u vezi sa prepoznavanjem slike, detekcijom objekata, prepoznavanjem lica, klasifikacijom slika itd. S tim u vezi, uvođenje dubokih konvolucionih neuronskih mreža ili *CNN-a* zaslužuje posebnu pažnju. Ove mreže su se intenzivno koristile za analizu vizuelnih slika sa izuzetnom tačnošću [74]. Bez obzira što nam daje mogućnost dodavanja više slojeva *CNN-u* za rešavanje komplikovanijih zadataka u kompjuterskom vidu, dolazi sa sopstvenim skupom problema. Primećeno je da obučavanje neuronskih mreža postaje teže sa povećanjem broja dodatih slojeva, au nekim slučajevima i tačnost opada. Iz navedenih razloga upotreba ResNet dobija na značaju. Dublje neuronske mreže je teže obučiti. Sa ResNet-om postaje moguće prevazići poteškoće u obuci veoma dubokih neuronskih mreža.

ResNet je skraćenica od Residual Network. To je inovativna neuronska mreža koju su prvi predstavili Kaiming He, Ksiangiu Zhang, Ksaoking Ren i Jian Sun u svom istraživačkom radu kompjuterskog vida iz 2015. pod nazivom „Duboko rezidualno učenje za prepoznavanje slike“ [91]. Ovaj model je bio izuzetno uspešan, što se može zaključiti iz činjenice da je njegov ansambl osvojio prvo mesto na ILSVRC 2015. godine sa greškom od samo 3,57%. Pored toga, takođe je bio prvi u ImageNet detekciji, ImageNet lokalizaciji, *COCO (Common Object in Context)* detekciji i *COCO* segmentaciji na *ILSVRC (ImageNet*

Large Scale Visual Recognition Challenge) & *COCO* takmičenjima 2015. ResNet ima mnogo varijanti koje rade na istom konceptu, ali imaju različit broj slojeva. ResNet50 se koristi za označavanje varijante koja može da radi sa 50 slojeva neuronske mreže.

Kada rade sa dubokim konvolucionim neuronskim mrežama na rešavanju problema u vezi sa kompjuterskim vidom, stručnjaci za mašinsko učenje se angažuju u slaganju više slojeva. Ovi dodatni slojevi pomažu u efikasnijem rešavanju složenih problema jer se različiti slojevi mogu obučiti za različite zadatke kako bi se dobili veoma precizni rezultati.

Dok broj naslaganih slojeva može obogatiti karakteristike modela, dublja mreža može pokazati problem degradacije. Drugim rečima, kako se broj slojeva neuronske mreže povećava, nivoi tačnosti mogu postati zasićeni i polako se degradirati nakon jedne tačke. Kao rezultat toga, performanse modela se pogoršavaju i na podacima o obuci i testiranju [92]. Ova degradacija nije rezultat preterivanja. Umesto toga, to može biti rezultat inicijalizacije mreže, funkcije optimizacije ili, što je još važnije, problema nestajanja ili eksplodiranja gradijenata.

ResNet je kreiran sa ciljem da se uhvati u koštac sa ovim problemom. Duboke preostale mreže koriste preostale blokove kako bi poboljšale tačnost modela. Koncept „preskakanja veza“, koji leži u srži preostalih blokova, je snaga ove vrste neuronske mreže. Ove veze za preskakanje rade na dva načina. Prvo , oni ublažavaju problem nestajanja gradijenta tako što postavljaju alternativnu prečicu kroz koju gradijent prolazi. Pored toga, omogućavaju modelu da nauči funkciju identiteta. Ovo osigurava da viši slojevi modela ne rade ništa lošije od nižih slojeva. Sažeto rečeno , rezidualni blokovi znatno olakšavaju slojevima da nauče funkcije identiteta. Kao rezultat, ResNet poboljšava efikasnost dubokih neuronskih mreža sa više neuronskih slojeva dok minimizira procenat grešaka. Drugim rečima, veze za preskakanje dodaju izlaze iz prethodnih slojeva na izlaze naslaganih slojeva, što omogućava obuku mnogo dubljih mreža nego što je to ranije bilo moguće.

Prva ResNet arhitektura bila je ResNet-34 koja je uključivala umetanje prečica za pretvaranje obične mreže u njenu rezidualnu mrežu. U ovom slučaju, obična mreža je inspirisana VGG neuronskim mrežama (VGG-16, VGG-19), sa konvolucionim mrežama koje imaju 3 × 3 filtere. Međutim, u poređenju sa VGGNet-ima, ResNet imaju manje filtera i manju složenost. 34-slojni ResNet postiže učinak od 3,6 milijardi *FLOP-a (Floating Point Ops per Second)*, u poređenju sa 1,8 milijardi FLOP-a manjih 18-slojnih ResNet -ova. Takođe je sledio dva jednostavna pravila dizajna – slojevi su imali isti broj filtera za istu veličinu

izlazne karte obeležja, a broj filtera se udvostručio u slučaju da je veličina karte obeležja prepolovljena da bi se sačuvala vremenska složenost po sloju. Sastojao se od 34 ponderisana sloja. Prečice su dodate ovoj običnoj mreži. Dok su ulazne i izlazne dimenzije bile iste, prečice do identiteta su direktno korišćene. Sa povećanjem dimenzija, trebalo je razmotriti dve opcije. Prvi je bio da bi prečica i dalje vršila mapiranje identiteta dok bi dodatni nulti unosi bili dopunjeni za povećanje dimenzija. Druga opcija je bila da se koristi prečica za projekciju da bi se uskladile dimenzije. Iako je arhitektura ResNet 50 zasnovana na gore navedenom modelu, postoji jedna velika razlika. U ovom slučaju, građevinski blok je modifikovan u dizajn uskog grla zbog zabrinutosti oko vremena potrebnog za obuku slojeva. Ova metoda je koristila snop od tri sloja umesto ranija dva [74].

Stoga je svaki od dvoslojnih blokova u ResNet 34 zamenjen troslojnim blokom uskog grla, formirajući ResNet 50 arhitekturu. Ovo ima mnogo veću preciznost od 34-slojnog ResNet modela. 50-slojni ResNet postiže učinak od 3,8 milijardi *FLOPS-a*, gde *FLOPS* predstavlja skraćenicu od *„Floating-point operations per second“*. Velike rezidualne mreže kao što su 101-slojni ResNet 101 ili ResNet 152 se konstruišu korišćenjem više troslojnih blokova. Čak i sa povećanom dubinom mreže, ResNet od 152 sloja ima mnogo manju složenost (na 11,3 milijardi FLOPS) od VGG-16 ili VGG-19 mreža (15,3/19,6 milijardi FLOPS).

5. Novi metod učenja prethodno obučenih modela

Novom modelu učenja prethodio je odabir DataSeta koji će biti korišćen za testiranje i analizu. Kompanija Google je 24.09.2019. godine objavila bazu od 3.000 video snimaka generisanih veštačkom inteligencijom koji su napravljeni korišćenjem različitih javno dostupnih algoritama [93]. DataSet je objavljen sa željom kompanije da pomogne istraživačima u borbi protiv zloupotrebe malipulisanih video materiijala, ali i fotografija [94]. Isti je postao javno dostupan besplatno i svaki istraživač je mogao i može preuzeti ovu bazu. Baza se može pronaći na više različitih lokacija jer su je različite istraživačke ustanove objavljivale. Baza korišćenja u daljem istraživanju preuzeta je sa github.com sajta i može se pronaći na linku: https://github.com/ondyari/FaceForensics/

Komanda za preuzimanje dataseta je:

```
python faceforensics_download_v1.py -d compressed files
```

Preuzeta baza video materijala je ukupno imala više od šest hiljada fajlova. Fajlovi su podeljeni u video materijale za testiranje, treniranje i validaciju. Svaka od ove tri grupe ima tri podgrupe a to su originalni video snimci, korišćene maske i video snimci sa zamenjenim licem.

Prethodno pomenuti DataSet je prilagođen istraživanju na način da je iz svakog video materijala uzet okvir sa tačno treće sekunde i na taj način dobijen je jednak broj slika koliko je preuzeto video materijala. Kreirana je skripta za uzimanje određenog okvira.

```
for AVI in `find . -name "*.avi" -type f`; do
  ffmpeg -hide_banner -loglevel error -y -i $AVI -ss 00:00:03.000 \
  -vframes 1 -f image2 $( echo $AVI | sed 's/.avi/.jpg/' )
done
```

Gore navedenim postupcima došli smo do finalnog DataSet-a koji je korišćen u istraživanju. U daljem tekstu će biti opisana tri prethodno obučena modela koji su uz odgovarajuće konfiguracije dali rezultate prikazane u tabli u nastavku ovog poglavlja. Tri modela koja su korišćena su:

- XceptionNet
- EfficientNetB4
- EfficientNetV2

Pomenuti modeli su već obučeni i formirani obučavanjem na osnovu miliona slika velikog broja različitih objekata (ImageNet).

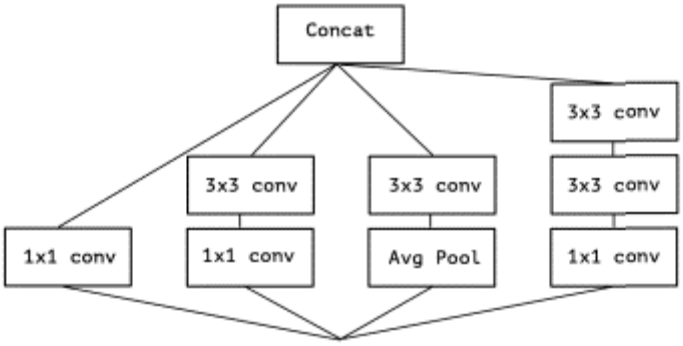
5.1. Prethodno obučeni modeli

U ovom delu disertacije su detaljno objašnjeni prethodno obučeni modeli. Sva tri modela sa posebnim konfiguracijama su korišćena u istraživanju i u predloženoj metodi za detekciju manipulisanih video snimaka.

5.1.1. XceptionNet

Konvolucione neuronske mreže su se pojavile kao glavni algoritam u kompjuterskom svetu, a razvoj projektovanja neuronskih mreža za njihovo bio je predmet značajne pažnje. Istorija dizajna konvolucionih neuronskih mreža započela je modelima u LeNet stilu [95], koji su bili jednostavni skupovi konvolucija za ekstrakciju karakteristika i operacije maksimalnog skupljanja za prostorno poduzorkovanje. Godine 2012. ove ideje su prerađene u AlekNet arhitekturu [96], gde su se operacije konvolucije ponavljale više puta između operacija maksimalnog skupljanja, omogućavajući mreži da nauči bogatije karakteristike na svakoj prostornoj skali. Ono što je usledilo bio je trend da se ovaj stil mreže učini sve dubljim, uglavnom vođen godišnjim ILSVRC takmičenjem; prvo sa Zeilerom i Fergusom 2013. [97], a zatim sa VGG arhitekturom 2014. [98]. Iste godine pojavio se novi stil mreže Inception arhitektura, koju su uveli Szegedi et al. 2014. Godine [99] kao GoogLeNet (Inception V1), kasnije rafiniran kao Inception V2 [100], Inception V3 [99], a najskorije Inception-ResNet [19]. Sam početak bio je inspirisan ranijom NetvorkIn-Netvork arhitekturom [11]. Od svog prvog uvođenja, Inception je bio jedan od porodice modela sa najboljim učinkom u skupu podataka ImageNet [101], kao i internim skupovima podataka koji se koriste u Google-u, posebno JFT.

Osnovni gradivni blok modela Inception stila je Inception modul, od kojih postoji nekoliko različitih verzija. Na slici 22 prikazan je kanonski oblik Inception modula, koji se nalazi u Inception V3 arhitekturi. Početni model se može shvatiti kao skup takvih modula. Ovo je odstupanje od ranijih mreža u VGG stilu koje su bile skup jednostavnih konvolucionih slojeva.



Slika 22 – Kanonski oblik Inception modula [102]

5.1.2. Inception hipoteza

Konvolucijski sloj pokušava ima cilj da pretvori filtere u 3D prostoru, sa dve prostorne dimenzije (širina i visina) i dimenzijom kanala; stoga je jedno konvoluciono jezgro zaduženo da istovremeno mapira međukanalne korelacije i prostorne korelacije. Ideja modula Inception je ta da ovaj proces učini lakšim i efikasnijim faktoringom u niz operacija koje bi nezavisno posmatrale međukanalne korelacije i prostorne korelacije. Tačnije, tipični Inception modul prvo posmatra međukanalne korelacije preko skupa 1x1 konvolucija, mapirajući ulazne podatke u 3 ili 4 odvojena prostora koji su manji od originalnog ulaznog prostora, a zatim mapira sve korelacije u ovim manjim 3D prostorima, preko obične 3x3 ili 5x5 konvolucije, prikazano na slici 22 [102]. Fundamentalna hipoteza koja stoji iza Inceptiona je da međukanalne korelacije i prostorne korelacije budu dovoljno odvojene da ih je poželjno ne mapirati ih zajedno.

„Ekstremna“ verzija Inception modula, zasnovana na ovoj jačoj hipotezi i prvo bi koristila konvoluciju 1x1 za mapiranje međukanalnih korelacija, a zatim bi posebno mapirala prostorne korelacije svakog izlaznog kanala.

Dve razlike između i „ekstremne“ verzije početnog modula i dubinski odvojene konvolucije su:

- Redosled operacija: dubinski odvojene konvolucije kao što se obično primenjuje (npr. u TensorFlov) prvo izvode prostornu konvoluciju u kanalima, a zatim izvode konvoluciju 1x1, dok Inception prvo izvodi konvoluciju 1x1.

- Prisustvo ili odsustvo nelinearnosti nakon prve operacije. U Inception, obe operacije su praćene ReLU nelinearnošću, međutim dubinski odvojive konvolucije se obično implementiraju bez nelinearnosti [102].

5.1.3. Arhitektura Xception

Arhitektura konvolucione neuronske mreže u potpunosti je zasnovanu na dubinski odvojivim slojevima konvolucije. Može se postaviti sledeću hipoteza: da se mapiranje međukanalnih korelacija i prostornih korelacija u mapama karakteristika konvolucionih neuronskih mreža može u potpunosti odvojiti. Pošto je ova hipoteza jača verzija hipoteze koja leži u osnovi Inception arhitekture, navedena arhitektura se naziva Xception, što je skraćenica za „Ektreme Inception“ [102].

Podaci prvo prolaze kroz ulazni tok, zatim kroz srednji tok koji se ponavlja osam puta, i na kraju kroz izlazni tok. Arhitektura Xception ima 36 konvolucionih slojeva koji čine osnovu za ekstrakciju karakteristika mreže.

U predstavljenoj studiji istražena je klasifikacija slika i stoga konvolucionu bazu prati sloj logističke regresije. Arhitektura Xception je linearni skup slojeva konvolucije koji se mogu odvojiti po dubini sa zaostalim vezama. Ovo čini arhitekturu veoma lakom za definisanje i modifikovanje; potrebno je samo 30 do 40 linija koda koristeći biblioteku visokog nivoa kao što je Keras ili TensorFlov-Slim [99], za razliku od arhitektura kao što su Inception V2 ili V3 koje je daleko složenije definisati. Implementacija Xception otvorenog koda koristeći Keras i TensorFlov je obezbeđena kao deo modula Keras Applications, pod licencom MIT [102].

U eksperimentu arhitekture Xception [102] izvršena je uporedna analiza Xception sa arhitekturom Inception V3, zbog njihove sličnosti, odnosno, Xception i Inception V3 imaju skoro isti broj parametara (Tabela 10) i stoga se bilo kakav jaz u performansama ne može pripisati razlici u kapacitetu mreže. Poređenje je provedeno na dva zadatka klasifikacije slika: jedan je zadatak klasifikacije jedne oznake od 1000 klase na skupu podataka ImageNet [14], a drugi je zadatak klasifikacije sa više oznaka od 17 000 klasa na velikom JFT-u. skup podataka.

	Broj parametara	**Koraka po sekundi**
Inception V3	**23.626.728**	**31**
Xception	**22.855.952**	**28**

Tabela 10 Broj parametara za Xception i InceptionV3

JFT je interni Google skup podataka za skup podataka o klasifikaciji slika velikih razmera, koji obuhvata preko 350 miliona slika visoke rezolucije obeleženih oznakama iz skupa od 17.000 klasa. Da bi se procenio performans modela obučenog na JFT, koristi se pomoćni skup podataka, FastEval14k. FastEval14k je skup podataka od 14.000 slika sa gustim napomenama iz oko 6.000 klasa (u proseku 36,5 oznaka po slici). Na ovom skupu podataka procenjuju se performanse koristeći srednju prosečnu preciznost za 100 najboljih predviđanja (MAP@100), i teži se doprinosu svake klase MAP@100 sa rezultatom koji procenjuje koliko je uobičajena (i stoga važna) klasa među slikama društvenih medija [102]. Ova procedura evaluacije ima za cilj da uhvati učinak na etiketama koje se često pojavljuju sa društvenih medija, što je ključno za proizvodne modele u Google-u

Korišćena je drugačija konfiguracija optimizacije za ImageNet i JFT:

• Na ImageNet-u:

– Optimizator: SGD

– Zamah: 0,9

– Inicijalna stopa učenja: 0,045

– Opadanje brzine učenja: opadanje brzine 0,94 svake dve epohe

• Na JFT:

– Optimizator: RMSprop [22]

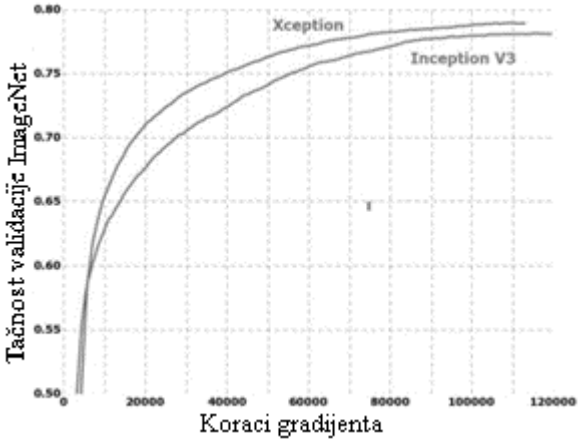
– Zamah: 0,9

– Početna stopa učenja: 0,001

– Opadanje brzine učenja: opadanje brzine 0,9 na svakih 3.000.000 uzoraka

Za oba skupa podataka, potpuno ista konfiguracija optimizacije je korišćena i za Xception i za Inception V3. Ova konfiguracija podešena je za najbolje performanse sa Inception V3; dok nije podešena optimizacija za hiperparametre Xception. Pošto mreže imaju različite profile obuke (Grafikon 1), ovo može biti ne optimalno, posebno na skupu podataka

ImageNet, na kojem je korišćena optimizacijska konfiguracija pažljivo podešena za početak V3 [102].



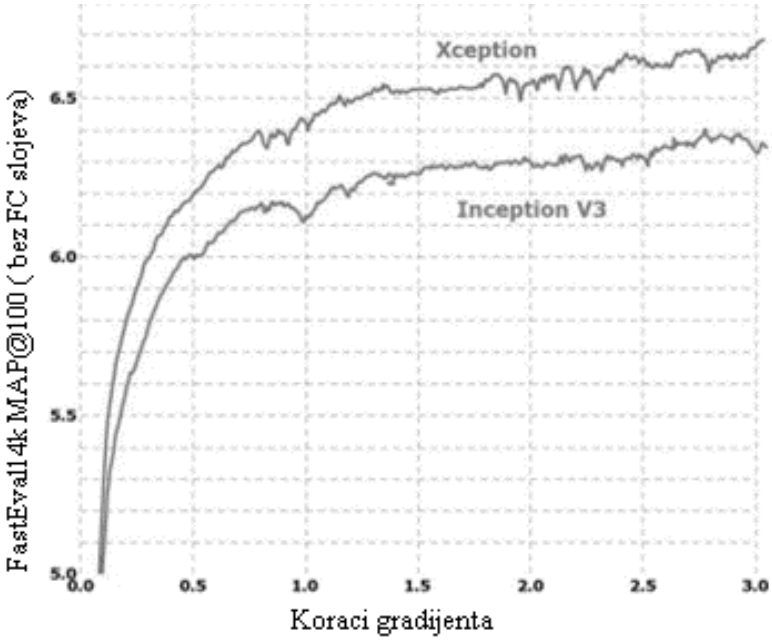
Grafikon 1 – Performanse obuke na skupu podataka ImageNet

Sve mreže su implementirane koristeći TensorFlov framevork [102] i obučene na po 60 NVIDIA K80 GPU-ova. Za ImageNet eksperimente, korišćeni su paralelni podaci sa sinhronim gradijentom spuštanja da bi se postigle najbolje performanse klasifikacije, dok je za JFT korišćen asinhroni gradijentni spuštanja kako bi se ubrzala obuka. ImageNet eksperimenti su trajali 3 dana, dok su JFT eksperimenti trajali više od mesec dana. JFT modeli nisu bili obučeni za punu konvergenciju, što bi trajalo više od tri meseca po eksperimentu.

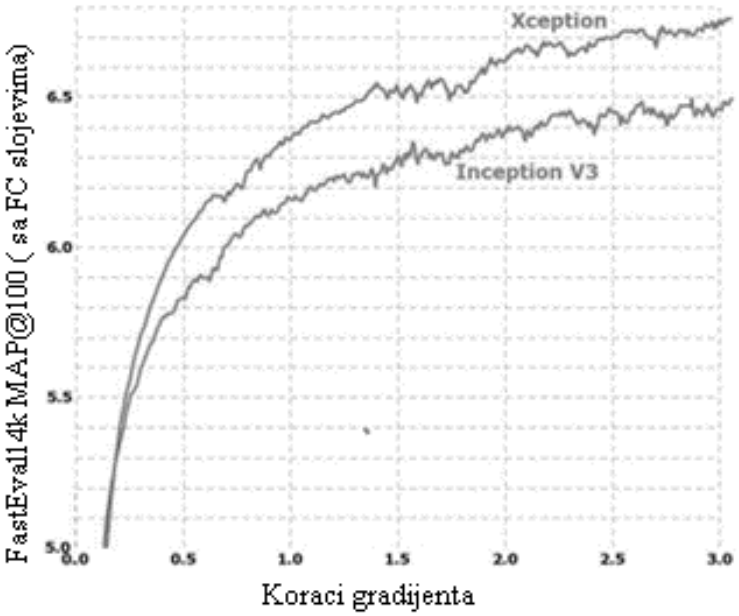
Sve evaluacije su sprovedene sa jednim isečenim ulaznim slikama i jednim modelom. ImageNet rezultati se izveštavaju o setu za validaciju, a ne o skupu za testiranje (tj. na slikama koje nisu na crnoj listi iz skupa za validaciju *ILSVRC 2012*). Rezultati JFT-a se prijavljuju nakon 30 miliona iteracija (jedan mesec obuke), a ne nakon potpune konvergencije. Rezultati su dati u tabeli 11, kao i na grafikonima dva i tri. Na JFT-u smo testirali obe verzije naših mreža koje nisu uključivale nijedan potpuno povezan sloj, kao i verzije koje su uključivale dva potpuno povezana sloja. slojeva od po 4096 jedinica pre sloja logističke regresije [102].

	FastEval14k MAP@100
Inception V3 bez FC slojeva	6.36
Xception bez FC slojeva	6.70
Inception V3 sa FC slojevima	6.50
Xception sa FC slojevima	**6.78**

Tabela 11 – Pikaz sprovedene evaluacije sa jednom isečenom slikom i jednim modelom ImageNet-a



Grafikon 2 – Rezultati za FastEval14k MAP@100 (Bez FC slojeva)

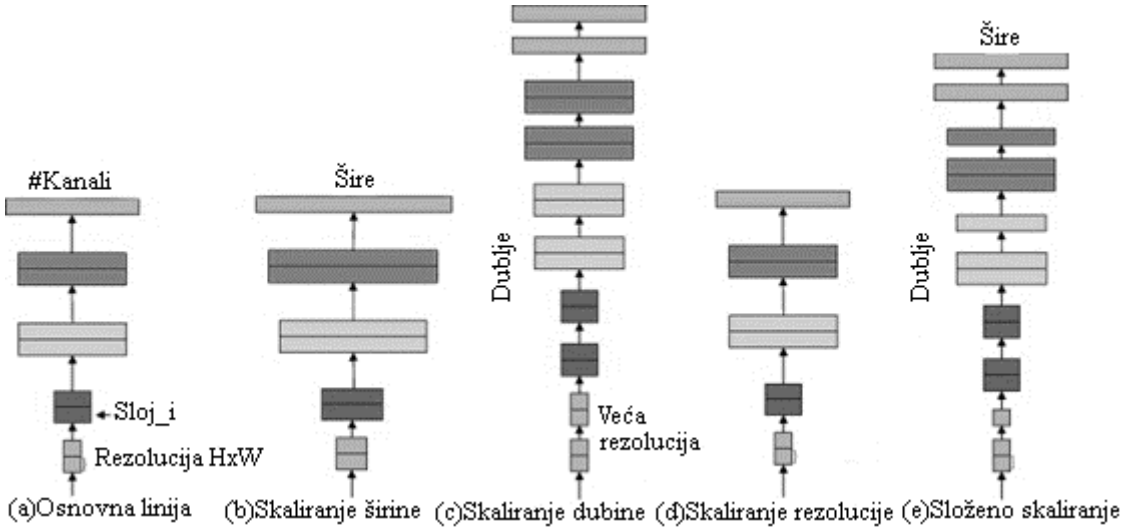


Grafikon 3 - Rezultati za FastEval14k MAP@100 (sa FC slojevima)

Na ImageNet-u, Xception pokazuje neznatno bolje rezultate od Inception V3. Arhitektura Ksception pokazuje mnogo veće poboljšanje performansi na JFT skupu podataka u poređenju sa skupom podataka ImageNet što možemo zaključiti iz prethodnih dijagrama.

5.1.4. EfficientNet-B

U procesu obuke modela dubokog učenja, najčešće korišćene metode za poboljšanje tačnosti modela su povećanje širine mreže, produbljivanje dubine mreže i poboljšanje rezolucije ulazne slike. Iako su prethodne studije kao što su ResNet, VideResNet, potvrdile gore navedene metode, balansiranje svih dimenzija širine/dubine/rezolucije mreže je kritično, i iznenađujuće, ova ravnoteža može biti postignuto jednostavnim skaliranjem svake dimenzije u konstantnom odnosu (Slika 23). Model EfficientNet je predložio Tan [103], koji može postići odgovarajući efekat na proširenje dubine, širine i rezolucije mreže, a zatim dobiti dobre performanse modela [103].



Slika 23 – Skaliranje svih dimenzija u konstantnom odnosu [103]

Prvo, možemo definisati CNN kao funkciju: $Y_i\ F_i(X_i)$, gde je F_i *operator (op),* Y_i je tenzor izlaza X_i je ulazni tenzor od oblik $<H_i\ W_I\ C_i>$ gde su $H_i\ W_I\ C$ su visina, širina i broj kanala ulazne slike. CNN Net se može opisati nizom sastavljenih slojeva: $Net = Fk$

$\odot\ldots\ \odot F_2\ \odot F_1\ (X_1) = \odot_{j=\ 1\ldots k}\ F_j(X_1)$. U stvarnom procesu primene, CNN sloj se obično primenjuje u više faza, a svaka faza koristi istu mrežnu arhitekturu. Dakle, možemo ga definisati kao:

$$Net = \bigodot_{i=1\ldots s} F_i^{L_i}\left(X_{\langle H_i, W_i, C_i\rangle}\right)$$

Gde je F_i^{Li} predstavlja sloj F_i koji se ponavlja L_i puta u etapi i $<H_i\ W_I\ C_i>$ predstavlja visinu, širinu i broj kanala ulaznog tenzora X sloja i.

Drugo, redovni CNN dizajni se uglavnom fokusiraju na pronalaženje optimalnog arhitektura slojeva F_i .Međutim, na osnovu unapred definisanog F_i osnovna mrežna struktura, skaliranje modela uglavnom pokušava da proširi dužinu mreže (L_i) širina (C_i) i rezolucija (H_i ,W_i) [103]. U međuvremenu, skaliranje modela popravlja problem dizajna za nova ograničenja resursa F_i i možemo istraživati različite L_i , C_i , H_i ,W_i , za svaki sloj zbog velikog prostora za dizajn. EfficientNet naglašava da svi slojevi moraju biti ravnomerno skalirani sa konstantnim odnosom da bi se smanjio prostor za dizajn. Cilj je da se značajno poboljša tačnost modela pod bilo kojim ograničenjem resursa, a što se može smatrati sledećim problemom optimizacije:

$$Max\ accuracy\ \big(Net(d,w,r)\big)$$

Gde *Max acuucracy* označava maksimalnu tačnost

$$Net(d,w,r) = \odot_{i=1\ldots s}\ \hat{F}_i^{\,d\cdot\hat{L}_i}\left(X_{\langle r\cdot\hat{H}_{i,r}\cdot\hat{W}_{i,w}\cdot\hat{C}_i\rangle}\right)$$

$$Memory(Net) \leq target\_memory$$

$$FLOPS\ (Net) \leq target\_flops$$

Gde je w, d i r predstavlja koeficijente koji se mogu koristiti za skaliranje širine mreže, dubine, a rezolucija predstavlja unapred definisane parametre u osnovnoj mreži. Treće, nova složena metoda skaliranja sa složenim koeficijentom ϕ može se koristiti za ravnomerno proširenje dubine, širine i rezolucije mreže na određeni način:

$$dubina: d = a^{\phi}$$

$$širina: w = \beta^{\phi}$$

$$Rezolucija: r = \gamma^{\phi}$$

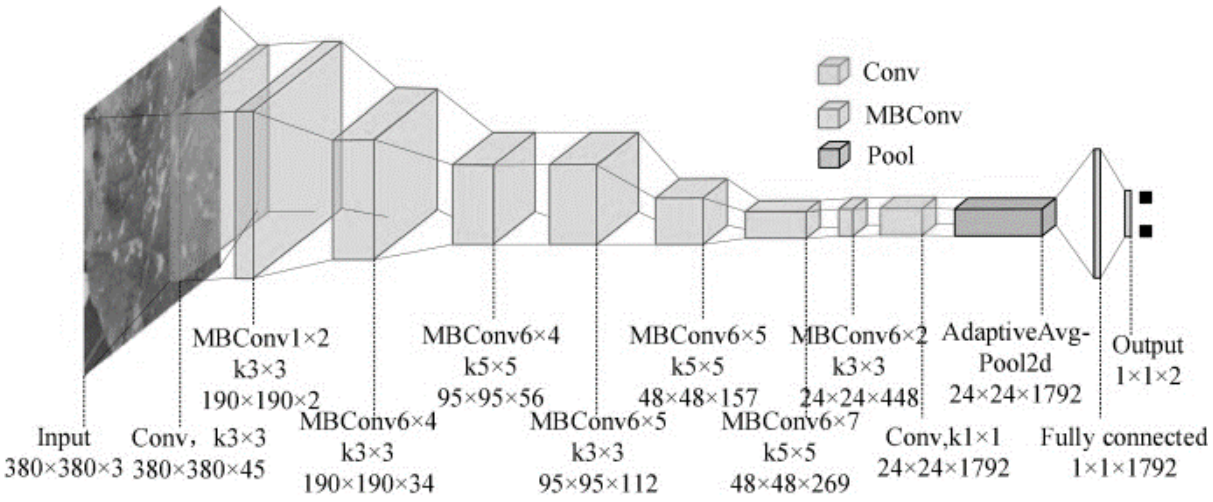
$$s.t.\ \alpha\cdot\beta^2\cdot\gamma^2\ \approx 2$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

gde su α, β, γ pre dstavljene konstante. Među njima, ph je određena vrednost koja određuje koliko još resursi važe za proširenje modela, dok α, β, γ određuju kako distribuirati ove dodatne resurse na širinu, dubinu i rezoluciju mreže redom. Osim toga, postoji specifičan odnos između FLOPS-a pravilne konvolucije op i d, v^2, r^2. Kada se dubina mreže udvostruči, FLOPS se takođe udvostručuje. Ali kada širina mreže ili rezolucija se udvostručuje, FLOPS četvorostruko. Jer konvolucija ops često kontrolišu troškove izračunavanja u CNN-ima, proširujući CNN korišćenjem jednačine (3) će približno povećati ukupni FLOPS za $(\alpha^2, \beta^2, \gamma^2)^{\phi}$. Konačno, zato što skaliranje modela ne menja operatore sloja Fi in unapred definisanu osnovnu mrežu, tako da je takođe važno imati dobru osnovna mreža. EfficientNet, nova osnovna mreža mobilne veličine je bila razvijen korišćenjem višeciljne pretrage neuronske arhitekture koja optimizuje i tačnost FLOPS. EfficientNet-B0 je proizveden pretragom [103], a u kojoj je arhitektura prikazana u tabeli ispod (Tabela 12). Njegov glavni građevinski blok uključuje mobilno obrnuto usko grlo.

i	**Stage**	**Operator** $\widehat{\int \iota}$	**Rezolucija** $\widehat{H_\iota} \times \widehat{W_\iota}$	**#Kanali** $\widehat{C_\iota}$	**#Slojevi** $\widehat{L_\iota}$
	1	Conv3 x 3	224 x 224	32	1
	2	MBConv1, k3 x 3	112 x 112	16	1
	3	MBConv6, k3 x 3	112 x 112	24	2
	4	MBConv6, k5 x 5	56 x 56	40	2
	5	MBConv6, k3 x 3	28 x 28	80	3
	6	MBConv6, k5 x 5	14 x 14	112	3
	7	MBConv6, k5 x 5	14 x 14	192	4
	8	MBConv6, k3 x 3	7 x 7	320	1
	9	Conv1 x 1&Pooling&FC	7 x 7	1280	1

Tabela 12 – Arhitektura pretrage za EfficientNet-B0



Slika 24 – Mrežna struktura EfficientNet-B4

5.1.5. EfficientNet - V

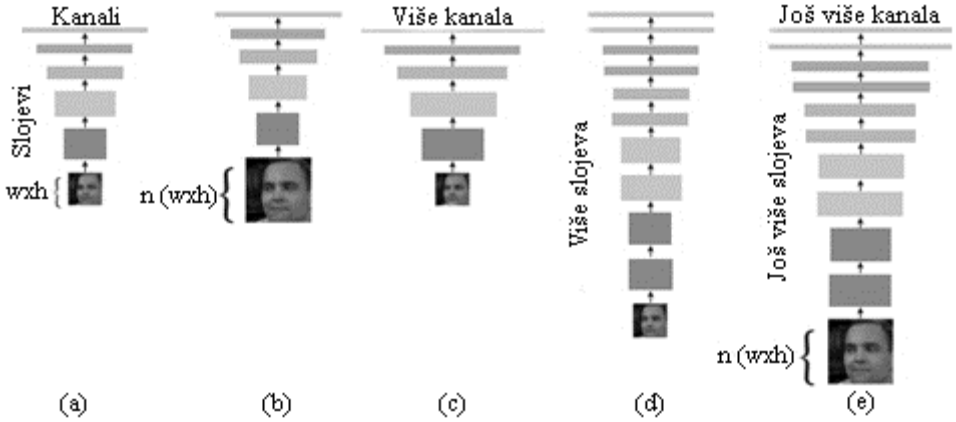
EfficientNet koristi jednostavan i efikasan složeni faktor za skaliranje mreže od tri dimenzije, dubine mreže, širine mreže i rezolucije ulazne slike, umesto tradicionalne metode proizvoljnog skaliranja dimenzija mreže, zasnovane na tehnologiji pretraživanja neuronske arhitekture [104] do dve Računarske inteligencije i neuronauka dobijaju optimalan skup parametara (kompozitnih koeficijenata).

U narednim istraživanjima i istraživanju, naučnici su otkrili da mreža serije EfficientNet nije mogla odlično da završi zadatke obuke i učenja zbog ograničenja i ograničenja eksperimentalne opreme nakon kontinuiranih testova praktične obuke. (zbog toga, 2021. godine, Mingking Tan su dodatno poboljšali EfficientNet mrežu, kreirali novu EfficientNet-V2 mrežu i podelili je na tri podmreže S, M i L. Nakon eksperimentalne verifikacije, u poređenju sa starom EfficientNet-V1 , nova mreža je modernizovanija, koristi manje resursa i ima veću stvarnu tačnost testiranja [104].

Metoda skaliranja modela (mreža serije EfficientNet) koju je predložio Google 2019. privukla je veliku pažnju u akademskoj zajednici. Da bi se istražio metod skaliranja modela koji uzima u obzir i brzinu i tačnost, prvo je predložena mreža serije EfficientNet koja istovremeno skalira tri dimenzije dubine mreže, širine mreže i rezolucije slike [104].

Kao što je prikazano na slici (Slika 25) predstavlja osnovni model mreže zasnovan na konvolucionoj neuronskoj mreži. Ulaz je trokanalna slika u boji sa širinom W i visinom H. Nakon konvolucije sloj po sloj, mreža bi mogla da nauči odgovarajuće karakteristike na slici;

slici 25(b)–25(d) predstavljaju tri uobičajene metode jednostrano skaliranje mreže, odnosno: Slika 25(b) je poboljšanje mreže iz perspektive rezolucije ulazne slike i poboljšanje efikasnosti učenja mreže proporcionalnim povećanjem ili smanjenjem veličine ulazne slike; Slika (25) je poboljšanje mreže i poboljšanje performansi mreže promenom broja kanala u svakom sloju mreže; Slika 25(d) je da povećati ili smanjiti broj mrežnih slojeva tako da mreža može naučiti više specifičnih informacija o karakteristikama i poboljšati efikasnost mreže. Slika 25(e) pokazuje da se u mreži EfficientNet, kompozitni parametri koriste za istovremeno skaliranje gornje tri dimenzije, čime se poboljšavaju ukupne performanse mreže.



Slika 25 – Poređenje sklairanja modela

Trodimenzionalni sveobuhvatni problem skaliranja u EfficientNet-u će biti izražen u formuli. Celu konvolucionu mrežu nazivamo J, a njen n konvolucioni sloj se može smatrati sledećim mapiranjem funkcije:

$$Y_i = F_i(X_i)$$

Među njima, $Y_i$je izlazni tenzor, X_i je ulazni tenzor, a njegova dimenzija je postavljena na $\langle H_i, V_i, C_i \rangle$. Radi pogodnosti izražavanja, dimenzija Batch u tenzoru je izostavljena; tada se cela konvoluciona mreža J, sastavljena od k konvolucionih slojeva, može izraziti

$$J = F_k \odot \cdots \odot F_2 \odot F_1(X_1)$$

5.2. Pregled predložene metode

Da bi se realizovalo predloženo rešenje bilo je neophodno sagledati sve prethodno obučene modele koji su opisani u prethodnim poglavljima. Nakon sagledavanja prethodno obučenih modela razvijena je tehnika za preobučavanje *XceptionNet, EfficientNetB* i *EfficentNetV* modela. Modeli su preobučeni uz upotrebu pretprocesiranja i tehnikama regulacije. Korišćen je *Single DLCNN (Deep Learning Convolutian Network)*. *DCLNN* je korišćena za prepoznavanje okvira iz manipulisanih video materijala. Ova neuronska mreža sa konvolucijom dubokog učenja podešavana kroz različite parametre davala je drugačija rešenja u pogledu odvajanja originalnih okvira od manipulisanih okvira. Parametri koje podrazumeva ovakva mreža su broj filtera, veličina jezgra filtera i parametri za proračun gradijenta.

Ovakvu tehniku smo primenili na već pomenutom DataSetu i za izdvajanje rezultata je korišćen *GridSearch*. *GridSearch* tehnika je obezbeđena u klasi *scikit-learn* (besplatna biblioteka programa za mašinsko učenje). Kada se konstruiše ova klasa mora se obezbediti rečnik hiperparametara za procenu. Rečnik se formuliše u argumentu ***param_grid.*** Ovaj rečnik predstavlja mapu imena parametra modela i niz vrednosti koje treba isprobati. Podrazumevano, tačnost jeste rezultat koji je optimizovan, ali možemo navesti i druge rezultate u argumentu konstruktora ***GridSearchCV***. Na ovaj način dobijamo i ostale rezultate koje možemo uporediti sa najboljim dobijenim. Takođe, pretraga mreže će po osnovnim podešavanjima koristiti samo jednu nit, ali to menjamo postavljanjem argumenta ***n_jobs*** u konstroktoru *GridSearchCV* na minus jedan. U ovom slučaju proces će koristiti sva jezgra na dostupnoj mašini. U nastavku rada *GridSearchCV* proces će konstruisati i proceniti jedan model za svaku kombinaciju parametara. Unakrsna validacija *Cross-Validation* se koristi za procenu svakog pojedinačnog modela. Podrazumevana vrednost je trostruka unakrsna provera, ali i ovaj parametar je menjan prilikom istraživanja. Zamena ovog parametra se vrši navođenjem argumenta ***cv***. U daljem tekstu prikazujemo primer jednog testiranja u toku eksperimenta.

```
param_grid = dict(epochs=[10,20,30])
grid = GridSearchCV(estimator=model, param_grid=param_grid,\
     n_jobs=-1, cv=5)
grid_result = grid.fit(X,Y)
```

Kada se testiranje sa podešenim parametrima završi možemo pristupiti rezultatu pretrage *GridSearch-a* u objektu rezultata koji se vaća iz grid.fit(). Kada pristupimo ovim

rezultatima član ***best_score_*** pruža pristup najboljem rezultatu uočenom tokom procedure optimizacije. Uvid u najbolju kombinaciju parametara koji su ostvarili ovaj rezultat dobijamo kroz ***best_params_***.U poglavlju „Rezultati eksperimenta“ predstavljeni su parametri koji su dali najbolje rezultate.

Takođe bitno je naglasiti da je menjana podrazumevana vrednost unakrsne validacije *(CV).* U unakrsnoj validaciji podaci su podeljeni na pet delova, zatim je model uklopljen 5 puta *(folds),* uz izostavljanje jedne petine podatka. Preostala petina podataka se koristi za merenje performansi. Za jednu kombinaciju vrednosti hiperparametara, prosek od pet grešaka predstavlja grešku unakrsne validacije. Ovakvo podešavanje omogućila nam je jedna od tehnika koje se koriste za *CV.* Izabrana tehnika je *Hold-out* a alternativa ovoj tehnici su bile i *k-folds, leave-one-out, leave-p-out, stratified K-folds, repeated K-folds, nested K-folds* i *time series CV.* Tehniha *Hold-out cross-validation* je izabrana jer se ona najčešće koristi na velikim DataSet-ovima. S obzirom da DataSet koji je korišćen ima preko 6.000 datoteka izbor je bio logičan. Pomenuta tehnika podrazumeva da se oko 80% datoteka odvoji za obuku a oko 20% za testiranje.

Obavljen je veći broj ponavljanja sa različito podešenim parametrima, izdvojene rezultate koji su vredni prikazivanja, zajedno sa vrednostima parametra prikazujemo u poglavlju „Rezultati eksperimenta“.

5.3. Konfiguracija DataSeta-a i prethodno obučenih modela

U ovom poglavlju je definisana tačna konfuguracija DataSet-a koji je korišćen u eksperimentu. Radi se podela podataka na tri grupe i obrazlaže se razlog podele što je ključni faktor za ispravnost i validnost rezultata. Takođe definišu se konfiguracije neuronske mreže sa konvolucijom dubokog učenja kao i modeli koji su korišćeni.

5.3.1. Konfiguracia DataSeta

Na početku petog poglavlja opisano je koji DataSet je korišćen, odakle je preuzet i kako su izvedene slike iz video materijala koje su korišćene u eksperimentu. U ovom poglavlju se bavimo DataSet-om opširnije. Važno je napomenuti da je neophodno da DataSet bude podeljen u više grupa kako bi rezultati bili što tačniji. Podaci su podeljeni u tri grupe, prva grupa je za obuku *(train)*, druga grupa je za test *(test)* i treća grupa je za validaciju *(validation)*. Razlog

zbog čega pravimo ovakvu podelu jeste jer je sama suština mašinskog učenja da računari obavljaju zadatak tako što mu se nešto pokaže na velikom broju primera. Na taj način učimo računar koristeći podatke koje imamo na raspolaganju. U idealnom slučaju algoritam će raditi jednako dobro sa novim podacima koje mu damo kao i sa podacima koji su već obrađeni. Razlog zašto delimo podatke na obuku, test i validaciju je da se neki od podataka zapravo žrtvuju u svrhe obuke a da kasnije procenimo algoritam sa novim podacima koje računar do tada nije koristio. Ovakva podela je bitna jer želimo da budemo sigurni da algoritam zna da primeni naučeno na novim podacima koji do tada nisu korišćeni. Broj podataka koji je korišćen je prikazan u tabelama ispod (Tabele 13,14,15 i 16). Sam odnos podataka koji se koristi za ovakav vid podele nije striktno definisan. Veliki broj istraživača koristi oko 80% podataka za treniranje. Ostali podaci se podele na testiranje i validaciju.

Grupa	**Broj datoteka**
Grupa za treniranje	4.224
Grupa za testiranje	900
Grupa za validaciju	900

Tabela 13 – Grupe datoteka preuzetog DataSeta

Datoteke koje su korišćene za trening podeljene su u tri osnovne grupe. Prva predstavlja okvire sa treće sekunde originalnih videa, druga grupa predstavlja korišćene maske koje su generisane i koje su takođe preuzete iz okvira sa treće sekunde. Poslednja grupa je treća grupa koja predstavlja okvir iz treće sekunde videa koji je izmenjen. Svaka od navedenih grupa ima po 1408 okvira koji su analizirani u ovom eksperimentu.

Grupa	**Broj datoteka**
Original	1.408
Maska	1.408
Izmenjeni	1.408

Tabela 14 – Grupe datoteka korišćenih za trening

Datoteke koje su korišćene za testiranje su takođe podeljene tri osnovne grupe. Grupe su iste kao i za trening datoteke. U ovim grupama je korišćen manji broj okvira, shodno pravilu koje je ranije pomenuto. Primenjivajući pravilo da veći broj datoteka koristimo za trening a manji broj za testiranje i validaciju u test datotekama svaka grupa je imala po 300 okvira iz treće sekunde originalnih video materija preuzetih sa Google-ovog DataSeta.

Grupa	**Broj datoteka**

Original	300
Maska	300
Izmenjeni	300

Tabela 15 - Grupe datoteka korišćenih za testiranje

Kao poslednji DataSet korišćen je set za validaciju. Kod njega je primenjeno isto pravilo kao i za DataSet testiranja. To znači da je i tu koiršćen znatno manji broj datoteka u odnosu na datoteke korišćene u trening grupama. Kod validacije je podela takođe na originalne okvire, maske i izmenjene okvire sa treće sekunde.

Grupa	**Broj datoteka**
Original	300
Maska	300
Izmenjeni	300

Tabela 16 – Grupe datoteka korišćenih za validaciju

5.3.2. Konfiguracija neuronske mreže sa konvolucijom i korišćeni modeli

Pomenuti i opisani prethodno obučeni modeli su konfigurisani uz pomoć *Single DLCNN* (*Deep Learning Convolutional Neural Network*). Neuronska mreža sa konvolucijom dubokog učenja koja je korišćena je dizajnirana da razlikuje okvire preuzete iz video materijala kao prava lica ili lažna u sistemu računarski potpomognute detekcije CAD (*Computer-aided detection*). Za obuku i testiranje DLCNN korišćen je DataSet koji je definisan u prethodnom poglavlju. Za prave i ispravne strukture lica korišćene su slike kreirane od okvira iz treće sekunde sa originalnih video materijala iz DataSeta. Takođe za obuku i trening DLCNN bili su neophodni i okviri iz treće sekunde iz manipulisanih video materijala koji su takođe deo ovog DataSet-a. DLCNN arhitektura je izabrana variranjem broja filtera, veličine jezgra filtera i parametara proračuna gradijenta u slojevima konvolucije.

Za pretragu najbolje varijacije korišćen je *GridSearch.* Pomenuti alat se koristi za optimizaciju prilikom podešavanja hiperparametara. Definiše se mreža parametara koju želimo da obradimo i odatle biramo najbolju kombinaciju parametara. Kada dobijemo takve rezultate potrebno je da se uzme u obzir i greška unakrsne validacije CV *(Cross-Validation).* Prilikom testiranja performansi modela sa ovakvim kombinacijama hiperparametara, verovatno da postoji rizik od preterivanja. Rizik od preterivanja znači da se može desiti da je samo slučajno

skup podataka za obuku dobro odgovarao ovoj specifičnoj kombinaciji hiperparametara. Učinak na ovim podacima u stvarnom i realnom okruženju bi mogao biti lošiji. Da bismo dobili pouzdaniju procenu performansi kombinacije hiperparametara moramo uzeti u obzir grešku unakrsne validacije. U unakrsnoj validaciji podaci su podeljeni na pet delova, zatim je model uklopljen 5 puta *(folds),* uz izostavljanje jedne petine podatka. Preostala petina podataka se koristi za merenje performansi. Za jednu kombinaciju vrednosti hiperparametara, prosek od pet grešaka predstavlja grešku unakrsne validacije. Uz pomoć ove metode dobijamo konačni izbor kombinacije parametara pouzdanijim.

Takođe, neophodno je odrediti pravu tehniku kojom se koristi CV. Postoji više različitih tehnika, neke od njih se obično koriste, dok druge rade samo u teoriji [105]. Izdvojeno je 8 tehnika kao najčešće koriste a posebno je opisana i tehnika zadržavanja. Izdvojene tehnike su:

- *Hold-out*
- *K-folds*
- *Leave-one-out*
- *Leave-p-out*
- *Stratified K-folds*
- *Repeated K-folds*
- *Nested K-folds*
- *Time series CV*

Prilikom ovog eksperimenta korišćena je tehnika *Hold-out cross-validationt.* Ovakva tehnika unakrsne provere podrazumeva da se oko 80% datoteka odvoji za obuku a oko 20% za testiranje, o čemu smo već pričali u prethodnom poglavlju gde je obrazloženo kako je konfigurisan ceo DataSet koji je korišćen prilikom eksperimenta. Ovakava tehnika se najčešće koristi za velike DataSet-ove.

Takođe prilikom ovog eksperimenta učeni su ansambali dubokih neuronskih mreža tehnikom težinskog usrednjavanja *(Weighting Average)*. Težinsko usrednjavanje ili ponderisani prosek je proračun koji uzima u obzir različite stepene važnosti brojeva u skupu podataka. Prilikom izračunavanja težinskog usrednjavanja svaki broj u skupu podataka se

množi sa unapred određenom težinom pre nego što se napravi konačni proračun. Težinski usrednjavanji rezultat može biti tačniji od jednostavnog proseka u kome je svim brojevima u skupu podataka dodeljena ista težina. Formula kojom se izračunava ova vrednost je predstavljena u formuli ispod.

$$W = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

Oznaka	Značenje
$\boldsymbol{W}$	Težinsko usrednjavanje
$\boldsymbol{n}$	Broj pojmova za koji se radi usrednjavanje
$\boldsymbol{W_i}$	Težine primenjene na X vrednostima
$\boldsymbol{X_i}$	Vrednosti podataka koji se usrednjuju

Tabela 17 – Parametri korišćeni za izračunavanje težinskog usrednjavanja

5.3.3. Korišćena metrika

U ovom poglavlju opisuje se metrika uz pomoć koje smo došli do rezultata. Parametri koje smo izračunavali su tačnost *(accuracy)*, preciznost *(precision)*, FAR *(fake image classified as real image).*

Za svaki od navedenih rezultata izdvojeni su opisi i formule uz pomoć kojih su dobijeni rezultati.

Tačnost se definiše kao procenat tačnih predviđanja za podatke testa. Izračunava se deljenjem tačnih predviđanja sa brojem ukupnih predviđanja

$$Tačnost = \frac{Tačna\ predviđanja}{Ukupna\ predviđanja}$$

Preciznost se definiše kao deo relativnih primera, istinski pozitivnih među svim primerima za koje je predviđeno da pripadaju određenoj klasi.

$$Preciznost = \frac{Istinski\ pozitivni}{Istinski\ pozitivn + Lažno\ popzitivni}$$

FAR se definiše kao deo primera za koje je predviđeno da pripadaju klasi u odnosu na sve primere koji zaista pripadaju klasi. U mnogim radovima se FAR naziva *recall* pa je bitno naglasiti da su te vrednosti iste.

$$FAR = \frac{Istinski\ pozitivni}{Istinski\ pozitivni + Lažno\ negativni}$$

5.4. Rezultati eksperimenta

Najbolji rezultati prepoznavanja veštački modifikovanih video materijala za ovaj dataset ostvareni su metodima učenja dubokih neuronskih mreža (Deep Learning), koji kao osnovu koriste već obučene modele, formirane obučavanjem na osnovu miliona slika velikog broja različitih objekata (ImageNet). Usvojeni metod vrši preobučavanje ovakvih modela (Transfer Learning), uz upotrebu pretprocesiranja (Facial Extraction), metoda augmentacije slika i više tehnika regularizacije u obučavanju pojedinačnih mreža. Tačnost je povećana učenjem ansambala dubokih neuronskih mreža tehnikom težinskog usrednjavanja (Weighting Average) s dodatnom nelinearnom optimizacijom. U sledećoj tabeli prikazani su rezultati testa.

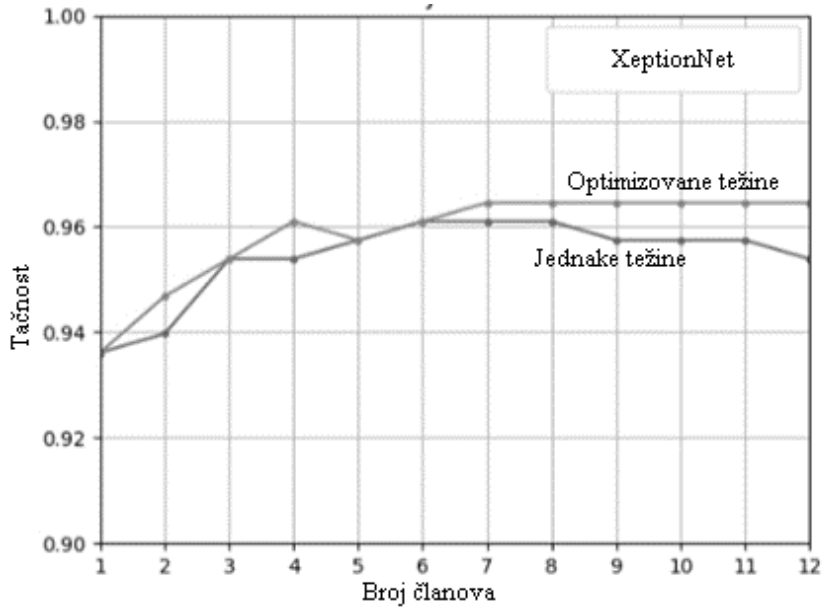
Prethodno trenirani modeli	Konfiguracija	Metod evaluacije	Tačnost	Preciznost	FAR
XceptionNet	Single DLCNN	5-fold CV	92.9% +- 0.6	0.9297	0.0810
	Ensemble1, n=7	Hold-out 0.8	96.5%	0.9622	0.0597
EfficientNetB4	Single DLCNN	5-fold CV	92.9% +- 2.0	0.9314	0.0611
	Ensemble2, n=7	Hold-out 0.8	**96.8%**	**0.9484**	**0.0597**
EfficientNetV2	Single DLCNN	5-fold CV	94.4% +-1.0	0.9469	0.0643
	Ensemble3, n=5	Hold-out 0.8	96.5%	0.9423	0.0672

Tabela 18 – Rezultai eksperimenta preučenih modela

U daljem tekstu se prikazuju pojedinačni rezultati za korišćene prethodno obučene modele. Prikazana su tri grafika na kojima se vidi srednja vrednost za WA ansamble trenirane prethodno opisanim metodama i konfiguracijama.

Prikaz rezultata učenja WA ansambla za osnovni model *XeptionNet,* broj članova *(n)* od 1 do 12. Na grafikonu ispod može se primetiti razlika rezultata kada je izvršena težinska optimizacija tj. usrednjavanje i rezultata pre izvršavanja optimizacije. Za skup podataka sa

istom težinom važnosti rezultati su niži nego za skup podataka koji je koristio usrednjavanje *(Weighting Average)*. Tačnost ovog eksperimenta je 96.45% a do ovog proračuna smo došli koristeći metrike koje su objašnjene u prethodnom poglavlju poglavlju. FAR vrednost je 5.97%.



Grafikon 4 - Prikaz rezultata za XceptionNet

		Predviđeno	
		Manipulisani	***Pravi***
Istinito	***Manipulisani***	126	**8**
	Pravi	1	146

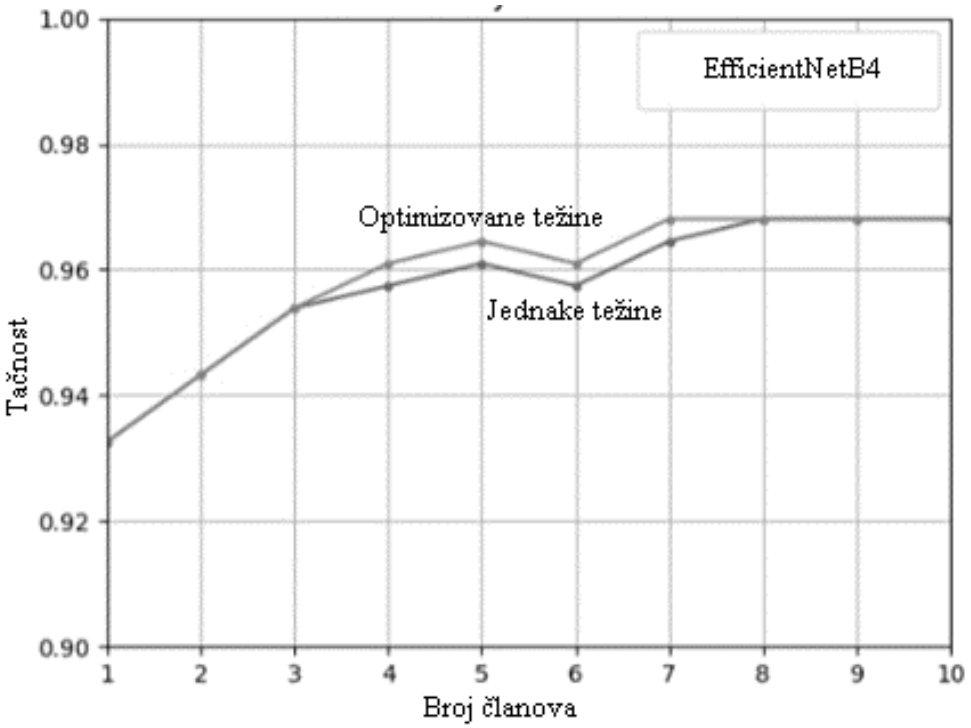
Tabela 19 – Tabelarni prikaz za XceptionNet

Accuracy (Xceptionnet) = 96.45%

False Acceptance Rate (FAR) = 5.97% (fake image classified as real image)

Prikaz rezultata učenja WA ansambla za osnovni model Efficient*NetB4,* broj članova *(n)* od 1 do 10. Na grafikonu ispod može se primetiti razlika rezultata kada je izvršena težinska optimizacija tj. usrednjavanje i rezultata pre izvršavanja optimizacije za pomenuti model. Za skup podataka sa istom težinom važnosti rezultati su niži nego za skup podataka koji je koristio usrednjavanje *(Weighting Average)*. Tačnost ovog eksperimenta je 96.81% a do ovog

proračuna smo došli koristeći metrike koje su objašnjene u prethodnom poglavlju poglavlju. FAR vrednost je 5.97% i ujedno ovo je najbolji rezultat koji smo ostvarili ovom metodom.



Grafikon 5 - Prikaz rezultata za EfficientNetB4

		Predviđeno	
		Manipulisani	***Pravi***
Istinito	***Manipulisani***	126	**8**
	Pravi	1	147

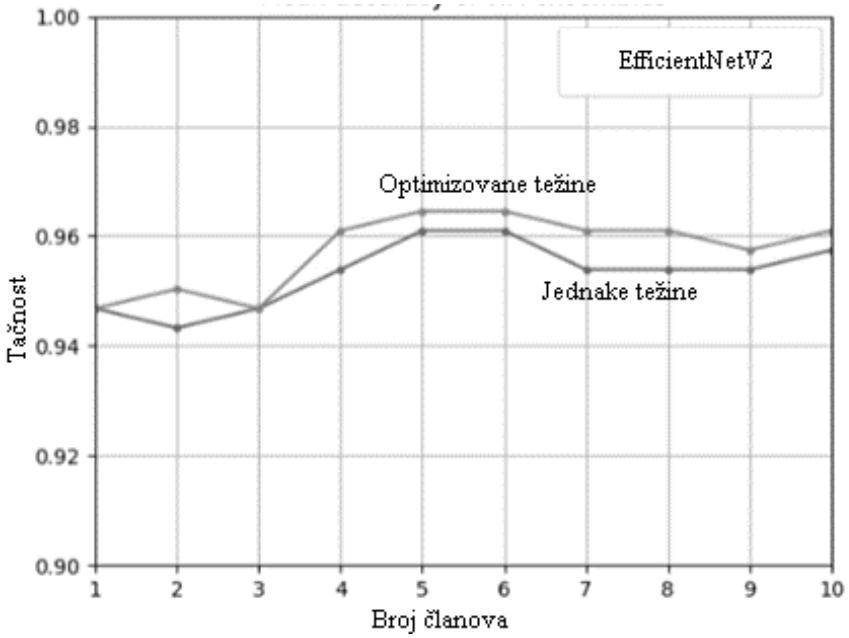
Tabela 20 - Tabelarni prikaz za EfficientNetB4

Accuracy (EfficeintB4) = 96.81%

False Acceptance Rate (FAR) = 5.97% (fake image classified as real image)

Prikaz rezultata učenja WA ansambla za osnovni model *EfficientNetV2,* broj članova *(n)* od 1 do 10. Na grafikonu ispod može se primetiti razlika rezultata kada je izvršena težinska optimizacija tj. usrednjavanje i rezultata pre izvršavanja optimizacije. Za skup podataka sa istom težinom važnosti rezultati su niži nego za skup podataka koji je koristio usrednjavanje *(Weighting Average)*. Tačnost ovog eksperimenta je 96.45% a do ovog proračuna smo došli

koristeći metrike koje su objašnjene u prethodnom poglavlju poglavlju. FAR vrednost je 6.72%.



Grafikon 6 - Prikaz rezultata za EfficientNetV2

		Predviđeno	
		Manipulisani	***Pravi***
Istinito	***Manipulisani***	125	**9**
	Pravi	1	147

Tabela 21 - Tabelarni prikaz rezultata za EfficientNetV2

Accuracy (EfficeintV2M) = 96.45%

False Acceptance Rate (FAR) = 6.72% (fake image classified as real image)

Iz prethodnog istraživanja i eksperimenta može se zaključiti da smo ostvarili preciznost od 96.8%. Ova preciznost je ostvarena kroz korišćenje tri prethodno obučena modela. Preobučavanjem ovih modela uz upotrebu pretprocesiranja i tehnikama regulacije u experimentu su dobijeni različiti rezultati. Najbolji rezultat je pokazao preobučeni model *EfficientNetB4* u kome je primenjena tehnika zadržavanja *(Hold*-out) koji je uklopljen pet puta *(5 folds).* Kako bi preciznost i tačnost eksperimenta bila pouzdana koristila se tehnika CV *(cross-validation).* Ostvareni rezultati se smatraju zadovoljavajućim, ali postoje ideje i za dalje

usavršavanje. Neke od ideja su prikazane u razmatranju diskusije rezultata a neke su ostavljen za dalju razradu i moguće komercijalno rešenje.

5.5. Poređenje dobijenih rezultata sa rezultatima drugih relevantnih metoda

U ovom poglavlju porede se druge različite metode i tehnike za prepoznavanje manipulisanih video materijala sa tehnikom i metodom koja je korišćena u ovom istraživanju. U poglavljima 5.4.1, 5.4.2 i 5.4.3 su opisana tri naučno istraživačka rada različitih autora sa osvrtom na njihove rezultate.

5.5.1. Poređenje predložene metode sa sistemom za detekciju koji uključuje LSTM za obradu okvira

Pravimo osvrt na naučno istraživački rad pod nazivom *„Deepfake Video Detection Using Recurrent Neural Networks“* [106].

U ovom radu je predložen sistem detekcije koji uključuje *LSTM* za obradu okvira izdvojenih iz videa. U konvolucionom *LSTM-u* postoje dve bitne komponente koje su koristili da bi uočili nedoslednosti između različitih okvira zbog procesa zamene. Jedna komponenta je *CNN* za izdvajanje karakteristika a druga je *LSTM* koji je korišćen za analizu vremenskih sekvenci. Za evaluaciju metode koristili su DataSet od 600 video zapisa. Koristeći ovaj metod ostvarili su 94% tačnosti.

5.5.2. Poređenje predložene metode sa metodom koja za detekciju koristi karakteristike na mezoskopskom nivou

Pravimo osvrt na naučno istraživački rad pod nazivom *„MesoNet: A Compact Facial Video Forgery Detection Network“* [57].

U ovom radu je predložen metod koji ima funkcionalnost koja bez ljudskog napora prati manipulaciju video snimaka i uglavnom se fokusira na tehnike koje su korišćene za stvaranje realističnih i spojenih video zapisa kao što su *Face2Face* i Deepfake. Obe tehnike imaju slične prirode manipulisanja i one imaju sledeće mane:

- Ulazni podaci su komprimovani na manjem prostoru kodiranja što čini izlaz zamućenim
- U nekim okvirima nema rekonstrukcije lica

Za detekciju tih mana autori su koristili dve arhitekture koje koriste karakteristike na mezoskopskom nivou. Rezultati su prikazali da postoji smanjenje tačnosti sa porastom nivoa kompresije u video zapisima. Takođe dokazali su da usta i oči igraju važnu ulogu u otkrivanju manipulisanih video materijala koji su napravljeni Deepfake tehnikom. Tvrdnja autora je da su na ovaj način dobili stopu detekcije od 98% za Deepfake materijale i 95% za *Face2Face* materijale.

5.5.3. Poređenje predložene metode sa metodom koja se fokusira na fiziološke signale ljudskog ponašanja

Pravimo osvrt na naučno istraživački rad pod nazivom *„In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking“* [107].

Autori u ovom radu su se fokusirali na fiziološke signale ljudskog ponašanja, takvi signali nisu dobro predstavljeni u manipulisanim video snimcima. Oni su iskoristili činjenicu da je prepoznatljivo odsustvo ovakvih znakova. Znakovi mogu upućivati na prirodne i neprirodne pojave kod pojedinca. U obzir su uzeti znakovi kao što su treptanje oka (brzo ili sporo), izraz lica i slično. Deepfake generisani video materijali nemaju realnu funkciju treptaja oka. Ovo predstavlja značajnu manu koju su autori iskoristili i predložili su algoritam koji obuhvata dva dela. Jedan je prethodna obrada koja uključuje ekstrakciju lica i poravnjanje, a zatim se šalje na drugi korak dugoročne rekurentne konvolucione mreže *(LRCN).*

Da bi testirali i obučili ovaj metod, koristili su sopstveni generisani skup podataka i autori su zadovolji preciznošću. Preciznost neće biti procentualno prikazana jer su autori koristili svoj skup podataka, dok u ovom radu izdvajamo zanimljivost same metode.

5.5.4. Poređenje predložene metode sa metodom mašinskog učenja koja koristi Support Vector Machine

Pravimo osvrt na naučno istraživački rad pod nazivom *„Image feature detectors for deepfake video detection“* [108].

U ovom radu Autori predlažu jedinstvenu metodu za praćenje Deepfake manipulisanih video materijala uz pomoć mašinskog učenja koristeći algoritam *SVM (Support Vector Machine)*. Ovakva tehnika se pokazala kao jedna od najkorisnijih. Ima četiri različita tipa jezgra. Tipovi jezgra su linearni, sigmoidni, radijalna bazna funkcija i polinom funkcija. Jezgra

se koriste za rešavanje različitih vrsta problema u mnogim oblastima. Oni koriste činjenicu da postoje neke nedoslednosti u generisanim slikama kao što su razlike u uslovima i okruženju. Iskoristili su ovaj nedostatak tako što su posmatrali ivice između osobe i okruženja. Koristili su različite metode matrica konfuzije a najbolji učinak koji su dobili je 94.5%.

5.6. Oblast primene predloženog rešenja

Predloženo rešenje se može primeniti u mnogim oblastima. U prvom i drugom poglavlju disertacije spominjane su negativne strane deepfake tehnologije. Rečeno je da se masovno koristi u video materijalima koji se danas lako mogu deliti preko interneta. Takođe iste tehnike se mogu primenjivati i na fotografijama koje mogu izazvati negativne događaje i neprijatne situacije. Predloženim rešenjem možemo detektovati manipulisane video materijale kojima imamo pristup u relativno kratkom vremenu. Ukoliko se radi o fotografijama, ceo proces detektovanja lažiranog materijala je još lakši i brži. Shodno ovim činjenicama bez obzira da li se na globalnoj mreži pojavio zlonamerni manipulisani video političara, glumca, glumice, sportiste ili bilo koje javne ličnosti, u mogućnosti smo da adekvatno i sa velikom tačnošću dokažemo da je video manipulisan. Time bi dokazali da osoba sa fotografije ili video materijala nije stvarna i samim tim diskreditovali bi bilo kakve informacije prenete na toj fotografiji ili u video materijalu.

U prethodnom pasusu za primer su uzete javne ličnosti, bitno je znati da se napretkom tehnologije omogućilo brzo i lako kreiranje manipulisanih materijala tako da sada svako ko poseduje računar za komercijalnu upotrebu ili kvalitetniji mobilni telefon može da kreira takav sadržaj. Uzimajući u obzir ove činjenice sasvim je prirodno da su se i fizička lica koja nisu javne ličnosti niti funkcioneri našli na ovakvim foto ili video materijalima. Ovakvi materijali su uglavnom šaljivog karaktera i brzo se dele po društvenim mrežama bez opasnosti da prouzrokuju velike svetske probleme, ali svakako mogu dovesti do neprijatnih situacija za pojedinca.

Razvijanjem metode za detektovanje manipulisanih materijala smatramo da smo pomogli u nameri da se smanji broj zlonamernog sadržaja na globalnoj mreži. U budućem radu planiramo unapređenje ovog modela, ali i eventualnu komercijalizaciju istog.

6. Zaključak

6.1. Ostvareni rezultati i doprinosi

Istraživanje je započeto analizom prethodnih modela predviđenih za detekciju manipulacije video materijala kroz Deepfake tehniku. Pažljivo su posmatrani prethodno obučeni modli i njihovi parametri. Posmatrani prethodno obučeni modeli su *XceptionNet, EfficientNetB* i *EfficientNetV*. Parametri ovih modela koji su menjani u procesu preobučavanja su konfiguracija mreže *SingleDLCNN*, broj *fold-ova* kao i vrednost za *Hold-out* tehniku. Korišćen je DataSet sa preko 6000 datoteka od kojih je veći broj datoteka korišćen za treniranje neuronske mreže a ostale datoteke su korišćene za testiranje i validaciju. Za izdvajanje najboljih rezultata korišćen je *CV (Cross-Validation)* a tačnost istih je uvećana tehnikom težinskog usrednjavanja tj. optimizacijom težine. Prikazani su rezultati za sva tri prethodno obučena modela a najbolji rezultat je ostvaren uz pomoć *EfficientNetbB4.*

Iz prethodno opisanog istraživanja i eksperimenta može se zaključiti da je najbolji ostvareni rezultat 96.8% *(FAR = 5.97%)*koji je dobijen uz pomoću *EffecientNetB4* preobučenog modela. Takođe predstavljeni su i rezultati druga dva posmatrana preobučena modela. Model *XceptionNet* ostvario je rezultat od 96.45% dok je model model *EfficientNetV2* ostvario isti taj rezultat odnosno 96.45%. Iako je tačnost ova dva modela na prvi pogled ista vrednost FAR *(fake images classified as real image)* se razilikuje. U prvom slučaju za *Xception net je iznosila* 5.97% a u drugom slučaju za *EfficientNetV2* iznosila je 6.72%. U poređenju sa ostalim rezultatima smatramo da je ovaj rezultat dobar i da se dodatnim unapređenjem može dodatno poboljšati. Ideja za dodatno unapređivanje je opisana u sledećem poglavlju.

Predloženo rešenje se može primeniti u mnogim oblastima. Razvijanje Deepfake tehnologije pokazalo je da se može koristiti u dobre, ali da može koristiti i u zlonamerne svrhe. Ovakvim rešenjem možemo sa velikom tačnosti i preciznosti odrediti koji video materijali su modifikovani tj. sa kojim video materijalima je vršena manipulacija i na takav način možemo smanjiti učinak zloupotrebe. Npr. ukoliko je neko kreirao zlonamerni manipulisani video uz pomoć Deepfake-a u nameri da diskredituje određenu osobu, na ovaj način možemo dokazati manipulaciju i demantovati diskreditaciju.

Ukoliko se radi o manipulaciji foto materijala, ovakav proces bi bio još lakši i brži i shodno tome mogli bi u relativno kratkom vremenskom periodu da reagujemo i upozorimo na diskriminišuću fotografiju.

6.2. Predlog daljeg rada

Za dalje usavršavanje ovakve tehnike predlaže se višestepena provera. U ovoj disertaciji DataSet je sastavljen od jednog okvira sa tačno treće sekunde originalnih i manipulisanih video materijala. Ostavljamo mogućnost da se u daljem razvijanju tehnike ista tehnologija iskoristi više puta, ali na više različitih okvira. Npr. ukoliko video traje dva minuta tj. 120 sekundi, može se uzeti okvir na svakih deset sekundi i na taj način izdvojiti 12 različitih okvira nad kojima treba ponoviti isti proces. Ovakvim postupkom verujemo da bi sa još više preciznosti i tačnosti mogli kroz mašinsko učenje ustanoviti sa kojim video materijalima je manipulisano a koji su originalni.

Takođe, kao predlog daljeg rada predlaže se analiza primene opisanih modela u slučajevima različitih formata. Napretkom tehnologije raste zastupljenost video materijala drugačijih formata. Prvenstveno mislimo na formate *Augmented Reality (AR), Virtual Reality (VR) i Reality* 360°.

7. Literatura

[1] M. Ghor, A. Sankaranarayanan i R. Picard, „Human Detection of Political Deepfakes across Transcripts, Audio, and Video,“ *arXiv preprint arXiv:2202.12883,* 2022.

[2] S. Chen, Y. Taiping, C. Yang, D. Shouhong, L. Jilin i J. Rongrong, „Local Relation Learning for Face Forgery Detection,“ u *Thirty-Fifth AAAI Conference on Artificial Intelligence*, 2021.

[3] W. Xinyao, Y. Taiping, D. Shouhong i M. Lizhuang, „Face Manipulation Detection via Auxiliary Supervision,“ u *Neural Information Processing - 27th International Conference (ICONIP 2020)*, Bangkok, Thiland, November 2020, 2020.

[4] Q. Yuyang, Y. Guojun, S. Lu, C. Zixuan i S. Jing, „Thinking in Frequency: Face Forgery Detection by Mining Frequency-aware Clues,“ 2020. [Na mreži]. Available: https://arxiv.org/abs/2007.09355. [Poslednji pristup 25 01 2022].

[5] D. Hao, L. Feng, S. Joel, L. Xiaoming i J. Anil, „On the Detection of Digital Face Manipulation,“ u *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[6] L. ingzhi, B. Jianmin, Z. Ting, Y. Hao, C. Dong, W. Fang i G. Baining, „Face X-Ray for More General Face Forgery Detection,“ u *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[7] T. Du, B. Lubomir, F. Rob, T. Lorenzo i P. Manohar, „C3D: Generic Features for Video Analysis,“ 2014. [Na mreži]. Available: https://arxiv.org/abs/1412.0767v1. [Poslednji pristup 11 2022].

[8] C. João i Z. Andrew, „A New Model and the Kinetics Dataset,“ u *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[9] H. Sepp i S. Jürgen, „Long Short-Term Memory,“ u *Neutral Computation 9*, 1997.

[10] M. Westerlund, „The Emergence of Deepfake Technology,“ t. 9, br. 11, 2019.

[11] P. Korshunov and S. Marcel, DeepFakes: a New Threat to Face Recognition?Assessment and Detection, 2018.

[12] Peipeng Yu, Zhihua Xia, Jianwei Fei, Yujiang Lu, „A Survey on Deepfake Video Detection," China, 2021.

[13] Borges, L., Martins, B., & Calado, P., „Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News," *Journal of Data and Information Quality,* t. 11, br. 14, 2019.

[14] Aldwairi, M., & Alwahedi, A., „Detecting Fake News," *Procedia Computer,* pp. 215-222, 2018.

[15] Figueira, A., & Oliveira, L., „The current state of fake news: challenges and opportunities.," *Procedia Computer Science,* pp. 121: 817-825, 2017.

[16] K. E. Anderson, „Getting acquainted with social networks and apps: combating fake news on social media," *Library HiTech News,* t. 35, pp. 1-6, 2018.

[17] R. Chawla, „Deepfakes: How a pervert shook the world," *International Journal of Advance Research and Development,* t. 4, br. 6, pp. 4-8, 2019.

[18] Maras, M. H., & Alexandrou, A., „Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos," *International Journal of Evidence & Proof,* t. 23, br. 3, pp. 255-262, 2019.

[19] F. J., „Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance," *Theatre Journal,* pp. 455-471, 2018.

[20] A. Grigoryan, „CGI in Film," [Na mreži]. Available: https://www.scribd.com/document/483271029/CGI-in-film. [Poslednji pristup 2022].

[21] M. staff, „MasterClass Articles," MasterClass, 25 02 2022. [Na mreži]. Available: https://www.masterclass.com/articles/what-is-cgi#what-is-computergenerated-imagery. [Poslednji pristup 03 2022].

[22] D. Grabham, „25 best movie CGI effects ever," Stuff, [Na mreži]. Available: https://www.stuff.tv/news/25-best-movie-cgi-effects-ever/.

[23] S. Karnouskos, „Artificial Intelligence in Digital Media:," *IEEE TRANSACTIONS ON TECHNOLOGY AND SOCIETY,* 2020.

[24] A. Siarohin, S. Lathuili`ere, S. Tulyakov, E. Ricci, and N. Sebe, „First order motion model for image animation,“ *Advances in Neural Information Processing Systems,* pp. 7137-7147, 2019.

[25] A. Smidi and S. Shahin, „Social Media and Social Mobilisation in the Middle East: A Survey of Research on the Arab Spring,“ *India Quarterly: A Journal of International Affairs,* t. 72, br. 2, pp. 196-209, 2017.

[26] Chesney, R. and Citron, D., „Deepfakes And The New Disinformation War.,“ Foreign Affairs, 11 12 2018. [Na mreži]. Available: https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war. [Poslednji pristup 15 05 2022].

[27] Bahar Uddin Mahmud and Afsana Sharmin, „Deep Insights of Deepfake Technology : A Review,“ t. 5, br. 1, pp. 13-23, 2020.

[28] A. Zucconi, „A Practical Tutorial for FakeApp,“ 14 03 2018. [Na mreži]. Available: https://www.alanzucconi.com/2018/03/14/a-practical-tutorial- for-fakeapp. [Poslednji pristup 14 05 2022].

[29] BuzzFeedVideo, „youtube.com,“ YouTube, 17 4 2018. [Na mreži]. Available: https://www.youtube.com/watch?v=cQ54GDm1eL0&t=1s. [Poslednji pristup 2022].

[30] S. Karnouskos, „Artificial Intelligence in Digital Media: The Era of Deepfakes,“ *Transactrions on technology and society,* 2020.

[31] Artem A. Maksutov, Viacheslav O. Morozov, Aleksander A. Lavrenov, Alexander S. Smirnov, „Methods of Deepfake Detection Based on Machine,“ Moscow, Department of Computer Systems and Technology, pp. 408-411.

[32] I. Korshunova, „Fast face-swap using convolutional neural networks,“ u *IEEE International Conference on Computer Vision*, 2017.

[33] K. Olszewski, „Realistic dynamic facial textures from a single image using gans,“ u *IEEE International Conference on Computer Vision*, 2017.

[34] Eyerys, „A Reddit User Starts 'Deepfake',“ Eyerys, 27 10 2017. [Na mreži]. Available: https://www.eyerys.com/articles/timeline/reddit-user-starts-deepfake#event-a-href-articles-timeline-metaverse-standards-forum-created-solve-interoperabilitythe-039-metaverse-standards-forum039-created-to-solve-interoperability-of-the-emerging-metaverses-a. [Poslednji pristup 05 2021].

[35] A. Harvey, „exposing.ai,“ 01 01 2021. [Na mreži]. Available: https://exposing.ai/vgg_face/. [Poslednji pristup 2022].

[36] D. Vlasic, „Face transfer with multilinear models,“ *In ACM SIGGRAPH,* p. 24, 2006.

[37] Chorowski, J., Weiss, R. J., Bengio, S., and Oord, A. V. D., „Un- supervised speech representation learning using wavenet autoencoders,“ *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* t. 2053, br. 12, p. 2041, 2019.

[38] M. Marana, Culture and Development, UNESCO Etxea, 2010.

[39] Study.com, „How Technology Affects Global & Local Cultures,“ 10 4 2015.

[40] M. Vučurnović, „Uticaj društvenih mreža Interneta na društvo,“ *ResearchGate,* 2010.

[41] O. Oakes, „PRWeek,“ 2022. [Na mreži]. Available: https://www.prweek.com/article/1581457/deepfake-voice-tech-used-good-david-beckham-malaria-campaign. [Poslednji pristup 2022].

[42] „malariamustdie.com,“ Malaria No More UK., [Na mreži]. Available: https://malariamustdie.com/david-beckham-launches-malaria-must-die-campaign.

[43] D. Lee, „Deepfake Salvador Dalí takes selfies with museum visitors,“ The Verge, 10 5 2019. [Na mreži]. Available: https://www.theverge.com/2019/5/10/18540953/salvador-dali-lives-deepfake-museum. [Poslednji pristup 07 2021].

[44] T. Hwang, „Deepfakes A Gruonded Threat Assessment,“ CSET - Center for Security and Emerging Technology, 07 2020. [Na mreži]. Available: https://cset.georgetown.edu/wp-content/uploads/CSET-Deepfakes-Report.pdf. [Poslednji pristup 03 2021].

[45] A. Khalid, „Deepfake Videos Are A Far, Far Bigger Problem For Women,“ QUARTZ, 09 10 2019. [Na mreži]. Available: https://qz.com/1723476/ deepfake-videos-feature-mostly-porn-according-to-new-study-from-deeptrace-labs/. [Poslednji pristup 16 05 2022].

[46] E. J. Dickson, „Deepfake Porn Is Still a Threat, Particularly for K-Pop Stars,“ RollingStone, 07 10 2019. [Na mreži]. Available: https://www.rollingstone.com/culture/culture-news/deepfakes-nonconsensual-porn-study-kpop-895605/. [Poslednji pristup 16 05 2022].

[47] P. Janaszkiewicz, J. Krysinska, M. Prys, M. Kierzuel, T. Lipcyznski i P. Rozewski, „Text Summarization For Storytelling: Formal Document Case,“ *Procedia Computer Science,* t. 126, pp. 1154-1161, 2018.

[48] Z. Xia, T. Qiao, M. Xu, X. Wu, L. Han i Y. Chen, „Deepfake Video Detection Based on MesoNet with,“ *Symmetry,* t. 14, 2022.

[49] M. Pavis, „Rebalancing our regulatory response to Deepfakes with performers' rights,“ *Sage journals,* t. 17, p. 08, 2021.

[50] A. Baiocco, „Political "Deepfake" Laws Threaten Freedom of Expression,“ Institute for Free Speech, 5 01 2022. [Na mreži]. Available: https://www.ifs.org/research/political-deepfake-laws-threaten-freedom-of-expression/. [Poslednji pristup 3 2022].

[51] R. J. Blankenship, „Educational Responsibility in the Deepfake Era: A Primer for TPACK Reform,“ IGI Global, 2021. [Na mreži]. Available: https://www.igi-global.com/chapter/educational-responsibility-in-the-deepfake-era/285053. [Poslednji pristup 2022].

[52] P. K. a. S. Marcel, „Vulnerability assessment and detection of Deepfake videos,“ u *International Conference on Biometrics*, Greec, 2019.

[53] Nguyen, T.T., Nguyen, C.M., Nguyen, D.T., Nguyen, D.T., & Nahavandi, S., „Deep Learning for Deepfakes Creation and Detection.,“ u *ArXiv, abs/1909.11573.*, 2019.

[54] Nguyen, H.H., Fang, F., Yamagishi, J., & Echizen, I., „Multi-task Learning For Detecting and Segmenting Manipulated Facial Images and Videos,“ [Na mreži]. Available: ArXiv, abs/1906.06876.

[55] Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan,, „Recurrent Convolutional Strategies for Face Manipulation Detection in Video,“ 2019.

[56] Nguyen, H.H., Yamagishi, J., Echizen, I., „Capsule-forensics: Using capsule networks to detect forged images and videos,“ u *In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Japan, 2019, pp. 2307-2311.

[57] D. Afchar, V. Nozick, J. Yamagishi i I. Echizen, „Mesonet: A compact facial video forgery detection network,“ *2018 IEEE International Workshop on Information Forensics and Security (WIFS),* pp. 1-7, 2018.

[58] C. Szegedy, „Going deeper with convolutions,“ *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 1-9, 2015.

[59] R. Wang, „Fakespotter: a simple yet robust baseline for spotting aisynthesized fake faces,“ u *International Joint Conference on Artificial Intelligence (IJCAI)*, 2020.

[60] D. E. J. Güera, „Delp: Deepfake video detection using recurrent neural networkS,“ u *15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, 2018, pp. 1-6.

[61] C. Szegedy, „Recurrent-convolution approach to deepFake detectionstate-of-art results on FaceForensics++,“ arXiv preprint arXiv:1905.00582, 2019.

[62] D. Montserrat, „Deepfakes Detection with Automatic Face Weighting,“ u *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, IEEE, 2020, pp. 2851-2859.

[63] Y. Zhao, „Capturing the persistence of facial expression features for deepfake video detection,“ u *International Conference on Information and Communications Security*, Springer, 2019, pp. 630-645.

[64] Xi Wu; Zhen Xie; YuTao Gao; Yu Xiao, „SSTNet: Detecting Manipulated Faces Through Spatial, Steganalysis and Temporal Features,“ u *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, IEEE, 2020, pp. 2952-2956.

[65] Li, Y., Lyu, S, „Exposing deepfake videos by detecting face warping artifacts,“ u *Computer Vision and Pattern Recognition*, arXiv:1811.00656, 2018.

[66] Y. Li, „Celeb-DF: a new dataset for deepfake forensics,“ pp. 1-10, 2016.

[67] Ciftci, U.A., Demir, I., Fakecatcher, L.Y., „Detection of synthetic portrait videos using biological signals,“ u *IEEE Trans. Pattern Anal. Mach. Intell. 1*, IEEE, 2020.

[68] S. Fernandes, „Predicting heart rate variations of deepfake videos using neural ODE,“ u *Proceedings of the IEEE International Conference on Computer Vision Workshops*, IEEE, 2019.

[69] S. Russell i P. Norvig, Veštačka inteligencija, savremeni pristup, t. 3, 2011.

[70] Plata, D. Rueda, R. Ramos-Pollan i F. A. Gonzalez, „Effective training of convolutional neural networks with small, specialized datasets,“ *Journal of Intelligent \& Fuzzy Systems,* t. 32, pp. 1333-1342, 2017.

[71] K. Alex, I. Sutskever i G. E. Hinton, „Imagenet classification with deep convolutional neural networks,“ *Advances in neural information processing systems,* t. 25, 2012.

[72] „www.educba.com,“ [Na mreži]. Available: www.educba.com. [Poslednji pristup 14 04 2022].

[73] NN, „www.simplilearn.com,“ 21 02 2022. [Na mreži]. Available: https://www.simplilearn.com/tutorials/deep-learning-tutorial/multilayer-perceptron. [Poslednji pristup 10 03 2022].

[74] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, H. Zhiheng, K. Andrej, A. Khosla, Berg, A. C. Berg i L. Fei-Fei, „ImageNet Large Scale Visual Recognition Challenge“.

[75] N. Kojić, I. Reljin i B. Reljin, „Dinamičko multicast rutiranje primenom Hopfildove neuralne mreže,“ *Telekomunikacije,* t. 1.

[76] L. Y. Boureau, J. Ponce i Y. Lecun, „A theoretical analysis of feature pooling in visual recognition,“ u *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, Haifa, 2010.

[77] N. Kojić, I. Reljin i B. Reljin, „Neural network for optimization of routing in communication networks,“ *Facta universitatis - series: Electronics and Energetics,* t. 19, pp. 317-329, 02 06 2007.

[78] J. Schmidhuber, „Deep learning in neural networks: An overview,“ *Neutral Networks,* pp. 85-117, 2015.

[79] J. Hopfield, „“Neural” computation of decisions in optimization problems,“ *Biol. Cybern. 52,* pp. 141-152, 1985.

[80] T. F. Gonzalez, Handbook of Approximation Algorithms and Metaheuristics, Chapman & Hall/CRC.

[81] L. Jzau-Sheng, N.-F. Huang i L. Mingshou, „The Shortest-Path Computation in MOSPF Protocol through an Annealed Chaotic Neural Network," *Proc. Natl. Sci. Counc. ROC(A),* t. 24, pp. 463-471, 2000.

[82] A. Anwar, „towardsdatascience.com," 07 06 2019. [Na mreži]. Available: https://towardsdatascience.com/the-w3h-of-alexnet-vggnet-resnet-and-inception-7baaaecccc96. [Poslednji pristup 11 2021].

[83] K. Simonyan, „Visualising Image Classification Models and Saliency Maps," *Deep Inside Convolutional Networks,* 2013.

[84] G. Rohini, „https://medium.com," 23 9 2021. [Na mreži]. Available: https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918. [Poslednji pristup 1 2022].

[85] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke i A. Rabinovich, „Going Deeper with Convolutions," *Computer Vision and Pattern Recognition.*

[86] R. B. Girshick, J. Donahue, T. Darrell i J. Malik, „Rich feature hierarchies for accurate object detection and semantic segmentation," *arXiv,* 2013.

[87] S. Mascerenhas i M. Agarwal, „A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," u *2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON)*, Bengaluru, India, 2021.

[88] B. Mandal, A. Okeukwu i Y. Theis, „Masked Face Recognition using ResNet-50," *arXiv,* 2021.

[89] A. Krizhevsky, I. Sutskever i G. E. Hinton, „ImageNet Classification with Deep Convolutional Neural Networks," u *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012.

[90] M. D. Zeiler i R. Fergus, „Visualizing and Understanding Convolutional Networks".

[91] K. He, X. Zhang, S. Ren i J. Sun, „Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 1904-1916, 09 01 2015.

[92] C. F. Cadieu, H. Hong, D. Yamins, N. Pinto, N. J. Majaj i J. J. DiCarlo, „The Neural Representation Benchmark and its Evaluation on Brain and Machine,“ *CoRR,* t. abs/1301.3530, 2019.

[93] K. Hao, „MIT Tehnology Review,“ [Na mreži]. Available: https://www.technologyreview.com/2019/09/25/132884/google-has-released-a-giant-database-of-deepfakes-to-help-fight-deepfakes/. [Poslednji pristup 2021].

[94] N. Dufour i A. Gully, „Google AI Blog,“ Google, 24 09 2019. [Na mreži]. Available: https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html. [Poslednji pristup 02 2021].

[95] Y. LeCun, L. Jackel, L. Bottou, C. Cortes, J. S. Denker,, „Learning algorithms for classification: A comparison on,“ u *Neural networks: the statistical*, 1995, pp. 261-276.

[96] A. Krizhevsky, I. Sutskever, and G. E. Hinton, „Imagenet classification with deep convolutional neural networks,“ u *In Advances in neural information processing systems*, 2012, pp. 1097-1105.

[97] M. D. Zeiler and R. Fergus., „Visualizing and understanding convolutional networks,“ u *In Computer Vision–ECCV 2014*, Springer, 2014, pp. 818-833.

[98] K. Simonyan and A. Zisserman, „Very deep convolutional networks for large-scale image recognition,“ u *arXiv preprint arXiv:1409.1556*, 2014.

[99] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, „Going deeper with convolutions,“ u *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[100] S. Ioffe and C. Szegedy, „Batch normalization: Accelerating deep network training by reducing internal covariate shift.,“ u *In Proceedings of The 32nd International Conference on Machine Learning*, 2015.

[101] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, „Imagenet large scale visual recognition challenge,“ 2014.

[102] F. Chollet, „Xception: Deep Learning with Depthwise Separable Convolutions,“ u *IEEE XPLORE*, 2016.

[103] PanZhanga, LingYanga, DaoliangLi, „EfficientNet-B4-Ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment,“ u *Computers and Electronics in Agriculture*, Elseliver, 2020.

[104] Liwei Deng , Hongfei Suo , and Dongjie L, „Deepfake Video Detection Based on EfficientNet-V2 Network,“ u *Harbin University of Science and Technology*, Harbin 150080, China, 2022.

[105] V. Lyashenko i A. Jha, „Cross-Validation in Machine Learning: How to Do It Right,“ Neptune, 18 04 2022. [Na mreži]. Available: https://neptune.ai/blog/cross-validation-in-machine-learning-how-to-do-it-right. [Poslednji pristup 05 05 2022].

[106] E. J. Delp i D. Guera, „Deepfake Video Detection Using Recurrent Neural Networks,“ *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS),* 30 11 2018.

[107] S. Lyu , M.-C. Chang i Y. Li, „In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking,“ *2018 IEEE International Workshop on Information Forensics and Security (WIFS),* 31 01 2019.

[108] F. F. Kharbat, T. Elamsy, A. Mahmud i R. Abdullah, „Image Feature Detectors for Deepfake Video Detection,“ *2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA),* t. 16, p. 3, 2020.

[109] A. Manhendran i A. Vedaldi, „Understanding Deep Image Representations by Inverting Them“.*Computer Vision and Pattern Recognition.*

[110] R. Schowengerdt, Remote Sensing, Models and Methods for Image Processing, Elsevier, 2006.

[111] M. D. Zeiler i R. Fergus, „Stochastic Pooling for Regularization of Deep Convolutional Neural Networks,“ *Neutral Networks,* 16 01 2013.