UNIVERZITET U BEOGRADU

ELEKTROTEHNIČKI FAKULTET

Analiza arhitektura dubokog učenja za verifikaciju potpisa: CNN, sijamske i triplet mreže

SEMINARSKI RAD

Predmet: ALGORITMI ZA DINAMIČKU OPTIMIZACIJU

Mentor:

Prof. Dr Predrag Ivaniš

Kandidat:

Nemanja Vasić 2021/0401

Beograd, 2026.

SADRŽAJ

1. Uvod

Potpis se može predstaviti kao slika koja prenosi određeni šablon piksela koji je karakterističan za tu osobu. Ostavljamo ga na dokumentima, računima, ugovorima kao potvrdu identiteta, pa je zahtev za automatsku verifikaciju potpisa veliki.[1]

U poslednjoj deceniji, razvoj metoda dubokog učenja značajno je unapredio oblast biometrijske autentifikacije, uključujući prepoznavanje i verifikaciju potpisa. Iako su konvolucione neuronske mreže (CNN) i druge duboke arhitekture postigle izuzetne rezultate u zadacima klasifikacije slika, njihova primena u oblasti verifikacije potpisa i dalje predstavlja istraživački izazov. Razlog tome leži u specifičnostima ovog problema: visokoj intraklasnoj varijabilnosti (razlike između potpisa iste osobe), niskoj interklasnoj varijabilnosti kod veštih falsifikata, kao i ograničenoj dostupnosti velikih i dobro anotiranih skupova podataka.

Verifikacija potpisa razlikuje se od klasične klasifikacije jer ne podrazumeva dodeljivanje uzorka unapred definisanoj klasi, već procenu sličnosti između dva potpisa – referentnog i testnog. Upravo zbog toga, savremeni pristupi sve češće koriste arhitekture zasnovane na učenju metrike (metric learning), kao što su sijamske (Siamese) i triplet mreže. Ove arhitekture ne uče eksplicitne

klase, već projekciju potpisa u prostor osobina u kome su potpisi iste osobe međusobno bliski, dok su potpisi različitih osoba ili falsifikati udaljeni.[2]

U radu iz 1994. godine, autori Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Sickinger i Roopak Shah sa AT&T Bell Laboratories predlažu novi pristup verifikaciji potpisa zasnovan na takozvanoj *Siamese* neuronskoj mreži. Cilj njihovog rada bio je razvoj sistema za dinamičku verifikaciju potpisa koji koristi podatke sa digitalne table za potpisivanje i koji omogućava da model potpisa bude sažet na manje od 80 bajtova, kako bi mogao da se skladišti na magnetnoj traci kreditne kartice.

Autori koriste Siamese arhitekturu sastavljenu od dve identične podmreže (bazirane na Time Delay Neural Network – TDNN), koje ekstraktuju karakteristike iz dva potpisa, dok završni sloj meri sličnost između dobijenih vektora osobina pomoću kosinusne mere ugla. Sistem se trenira tako da su pravi potpisi međusobno što sličniji u prostoru osobina, dok su falsifikati udaljeni. Tokom verifikacije, novi potpis se poredi sa statističkim modelom korisnika, a odluka se donosi na osnovu praga verovatnoće. [3]

U radu iz 2022. godine, Chattopadhyay i saradnici predlažu dvostepeni okvir za writer-independent offline verifikaciju potpisa pod nazivom SURDS. Autori prvo primenjuju self-supervised pre-trening zasnovan na encoder-decoder arhitekturi sa 2D attention mehanizmom, čime se uči reprezentacija potpisa bez korišćenja anotacija. Nakon toga, pretrenirani encoder se fino podešava primenom nadgledanog metric learning pristupa zasnovanog na dual triplet loss funkciji. Za razliku od standardnog triplet loss-a, predloženi dual triplet pristup koristi i intra-writer i cross-writer negativne uzorke, čime se dodatno povećava separabilnost između pravih potpisa i falsifikata u prostoru osobina. Eksperimenti na BHSig260 datasetu pokazuju da predloženi metod nadmašuje više postojećih pristupa, čime se potvrđuje efikasnost kombinacije self-supervised učenja i triplet metrike u zadatku offline verifikacije potpisa.[4]

Ovaj rad je organizovan u pet poglavlja. U uvodnom govorimo o značajnim radovima i tehnikama koji se bave ovom temom. U drugom poglavlju govorimo o osnovnim pojmovima koji su vezani metode dubokog učenja. Treće poglavlje govori o prikupljanu podataka, arhitekturama CNN, sijamske i triplet mreže. Četvrto poglavlje iznosi eksperimentalne rezultate kao i kvantitativnu analizu. Peto poglavlje se odnosi na zaključak i mogućnosti za budući rad.

2. Metode dubokog učenja

U ovom poglavlju će se razmatrati teorijska osnova za tri korišćene arhitekture neuralnih mreža. To su konvolucione neuronske mreže, siamske i triplet mreže.

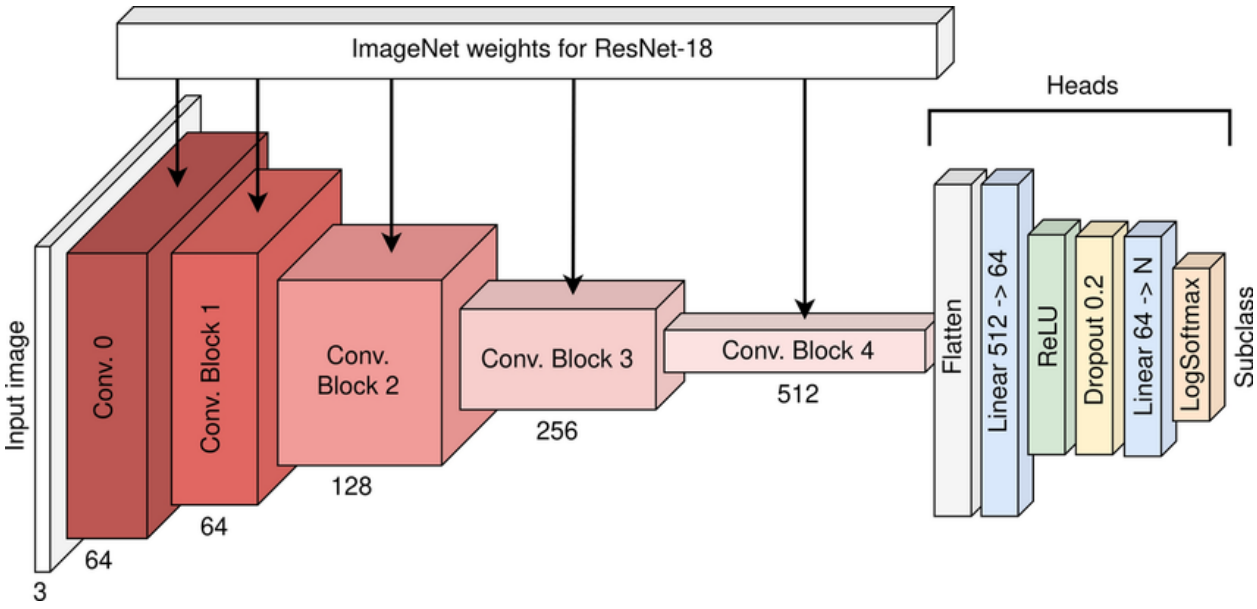
2.1. CNN – CONVOLUTION NEURAL NETWORK

Konvolucione neuronske mreže su napravljene kako bi radile sa mrežastim strukturama kao ulaznim vrednostima (input). Najočigledniji primer ovakvog inputa je 2D slika, ali mreža može raditi i sa tekstom. Ova vrsta podataka, dvodimenzionalna slika, pokazuje prostornu zavisnost jer susedni pikseli često imaju isti intezitet zato što predstavljaju istu boju. Dodatna

dimenzija predstavlja boje što stvara trodimenzionalni ulazni volumen, zbog toga karakteristike u konvolucijskoj neuronskoj mreži zavise od prostorne udaljenosti.

Važna karakteristika ovakvih mreža je operacija konvolucije. Ova operacija je korisna kada se obrađuju podaci koji imaju visoki nivo prostorne ili druge lokalne povezanosti. Konvolucione neuronske mreže koriste konvoluciju u bar jednom sloju.

Tajna uspeha bilo koje neuronske arhitekture leži u projektovanju strukture mreže uz semantičko razumevanje domene na koju se primenjuje. Konvolucione neuronske mreže zasnivaju se upravo na ovom principu, jer koriste retke veze uz visok stepen deljenja parametara na način osetljiv na domenu. Drugim rečima, nije svako stanje u određenom sloju povezano sa stanjima u prethodnom sloju na nasumičan način. Umesto toga, vrednost karakteristike u određenom sloju povezana je samo sa lokalnim prostornim područjem u prethodnom sloju, uz dosledan skup zajedničkih parametara koji se primenjuju na celu prostornu strukturu slike.[5]



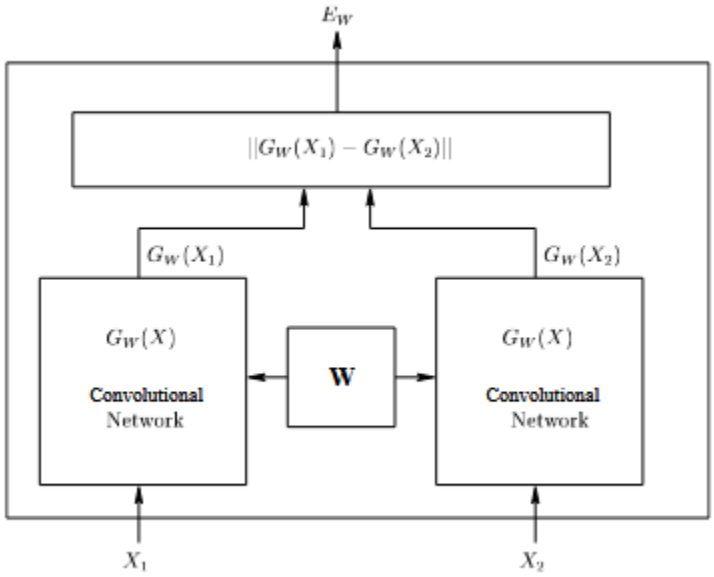
Slika 2.1. Primer klasične arhitekture CNN mreže[1]

2.2. SIAMESE NETWORK

Arhitektura sijamske neuronske mreže specifična je po tome što poseduje dva ulazna polja namenjena poređenju dva različita obrasca, dok izlazna vrednost odgovara nivou njihove sličnosti. Osnovu sistema čine dva identična pod-sistema koji su definisani deljenim vektorom parametara (), što osigurava da oba ulaza prolaze kroz identičnu transformaciju. Ovi pod-sistemi mogu biti zasnovani na mrežama sa vremenskim kašnjenjem (TDNN) koje vrše ekstrakciju obeležja, nakon

čega se udaljenost između dobijenih vektora može izračunati pomoću kosinusa ugla ili slične metrike. [3]

Sistem funkcioniše tako što mapira ulazne slike u prostor niže dimenzije, gde se kompatibilnost između njih meri skalarnom funkcijom energije . U cilju postizanja kompresije u vremenskoj dimenziji, arhitekture često koriste korake pod-uzorkovanja (sub-sampling) ili slojeve usrednjavanja. Tokom procesa učenja, primenjuje se funkcija gubitka koja je dizajnirana da minimizuje energiju kod parova koji pripadaju istom objektu, dok se kod različitih parova energija povećava. Ključno je da funkcija gubitka bude konveksna i da održava određenu marginu između tipova parova, čime se izbegavaju problemi poput gradijentnog platoa koji se može javiti pri korišćenju neadekvatnih normi. [6]



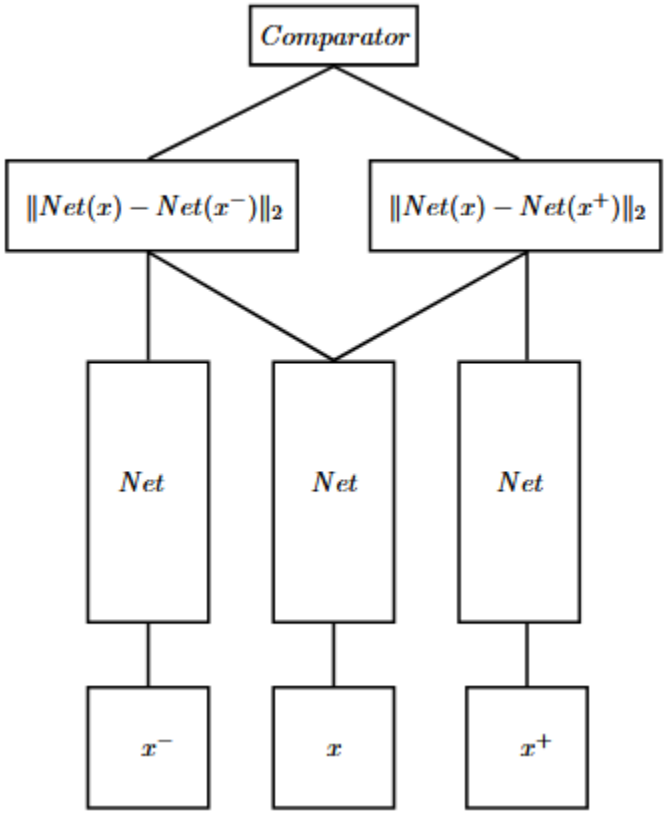
Slika 2.2. Siamese arhitektura[2]

2.3. TRIPLET MREŽE

Teorijska osnova triplet mreža počiva na konceptu učenja poređenjem, gde se model ne obučava da prepozna konkretnu klasu, već da nauči relaciju između uzoraka. Za razliku od siamskih mreža koje koriste parove, triplet mreža se sastoji od tri identična steka neuronskih mreža koje dele iste težine i parametre. Matematički cilj ove arhitekture je da projektuje ulazne podatke u vektorski prostor u kojem će rastojanje između sličnih uzoraka biti minimalno, dok će rastojanje između različitih uzoraka biti maksimalno.

U srcu ove teorije je funkcija gubitka koja operiše nad tri uzorka: sidrom, pozitivnim primerom (koji pripada istoj klasi kao sidro) i negativnim primerom (koji pripada različitoj klasi). Osnovna ideja je uvođenje pozitivne margine, koja služi kao sigurnosna zona koja primorava mrežu da negativni uzorak potisne značajno dalje od pozitivnog. Teorijski gledano, ovo omogućava modelu da nauči fine detalje i suptilne razlike između podataka, jer se učenje ne zasniva na apsolutnim vrednostima, već na relativnom odnosu rastojanja. Ovakva struktura obezbeđuje bolju generalizaciju i omogućava sistemu da uspešno identifikuje i kategoriše nove tipove podataka koji nisu bili prisutni u fazi obučavanja.[7]

U naprednijim okvirima, poput sistema SURDS, uvodi se koncept dualnog triplet gubitka kako bi se model efikasnije nosio sa specifičnim izazovima verifikacije potpisa van mreže. Ovaj pristup simultano tretira dve kategorije negativnih uzoraka: intra-writer negative, koji predstavljaju vešte falsifikate unutar klase istog pisca, i cross-writer negative, koji su autentični potpisi drugih pisaca. Cilj ovakve optimizacije je da se osigura da originalni potpis bude uvek bliži svom pozitivnom paru nego bilo kom negativnom uzorku, bez obzira na to da li se radi o namernoj imitaciji ili slučajnoj sličnosti sa drugim pismom. Ovakva strategija omogućava robusno diskriminativno učenje i bolju separaciju autentičnih potpisa od falsifikata, što je ključno za sisteme koji rade nezavisno od pisca.[4]



Slika 2.3. Arhitektura Triplet mreža[3]

2.4. ODREĐIVANJE PERFORMANSI MODELA

Kada imamo model koji je naš alogoritam za mašinsko učenje napravio koristeći set za treniranje kako da kažemo da je dobar. Koristimo metrike kako bismo odredili da li je model dobar. Za modele koji vrše klasifikaciju najkorišćenije metrike su:

- ***confusion matrix,***
- ***accuracy,***
- ***cost-sensitive accuracy,***
- ***precision/ recall, and***
- ***area under the ROC curve.***

2.4.1. Confusion matrix

The confusion matrix je tabela koja pokazuje kako je uspešan model klasifikacije u tačnom predviđanju primera koji pripadaju različitim klasama.

	original (predikcija)	**fake (predikcija)**
original (realnost)	**TP**	**FN**
fake (realnost)	**FP**	**TN**

Tabela 2.6. Confusion matrix

Tabela gore prikazuje confusion matrix. U slučaju da je predikcija tačna a i realnost tačna onda je to **true positive** (**TP**), a ako je stvarnost netačna onda je **false pozitive (FP)**. Kada je predikcija netačna a stvarnost tačna onda je **false negative** (**FN**), a ako je stvarnost netačna onda je **true negative**(**TN**).

2.4.2. Accuracy

Tačnost je broj ukupno tačnih klasifikovanih primera podeljen sa ukupnim brojem klasifikovanih primera.

$$\boldsymbol{Accuracy = \frac{TP+TN}{TP+TN+FP+FN}} \quad (1)$$

2.4.3. Precision/Recall

Dve najpopularnije metrike za modele su **preciznost** i **odziv**. **Preciznost** je odnos pozitivnih tačnih predikcija u odnosu na ukupan broj pozitivnih predikcija.

$$\boldsymbol{Precision = \frac{TP}{TP+FP}} \quad (2)$$

Odziv je odnost tačnih pozitivnih predikcija sa ukupnim broj pozitivnih primera.

$$\boldsymbol{Recall = \frac{TP}{TP+FN}} \quad (3)$$

2.4.4. Cost-Sensitive Accuracy

Kada imamo klasifikaciju više klasa, koje imaju različito važnost nama, možemo koristiti **cost-sensitive accuracy.** Mi dodeljuje vrednosti za FP i FN. I onda kada popunimo confusion matrix mi vrednosti koje smo dobili za FP i FN množimo sa vrednost koju smo stavili pre nego što iskoristimo jednačinu (1).

2.4.5. Area under the ROC curve

ROC kriva (Receiver Operating Characteristic) prikazuje odnos između **true positive rate (TPR)** i **false positive rate (FPR)** za različite prageve klasifikacije. TPR je definisan kao:

$$TP = \frac{TP}{TP+FN}, \quad FPR = \frac{FP}{FP+TN} \tag{3}$$

Koristi se za evaluaciju klasifikatora koji daju izlaz u vidu verovatnoće (npr. logistička regresija, neuronske mreže).
ROC kriva se dobija tako što se prag klasifikacije menja u opsegu [0, 1], a za svaki prag izračunavaju se TPR i FPR. Na osnovu toga crta se ROC kriva, a **AUC (Area Under the Curve)** meri ukupnu uspešnost klasifikatora što je AUC veći, model je precizniji.[10]

3. Metode i materijali

Skup podatak korišćen u ovom projektu sastoji se od 30 slika potpisa, pri čemu je polovina originalnih, dok je druga polovina falsifikovanih. Projekat je podeljen u dve celine: ***detekciju*** i ***verifikaciju*** potpisa.

Pre procesa detekcije, na ulaznu sliku primenjuje se ***retinex filtar*** kako bi umanjio uticaj senki i neujednačenog osvetljenja. Potom se slika konvertuje u sivu (grayscale) sliku, što omogućava jednostavniju obradu.

Zatim se vrši segmentacija slike pomoću histograma, sa ciljem uklanjanja neželjenih oblasti koje sa velikom verovatnoćom ne sadrže potpis. Na dobijenu binarnu sliku primenjuje se tehnika ***povezane komponente*** (CCL), kako bi se dodatno izolovali segmenti koji mogu predstavljati potpis.

Dobijeni segmenti izdvajaju iz slike koristeci modul ***Cropper***, potom se koristi ***Judger*** koji odlučuje na osnovu određenog kriterijuma da li je segment potpis ili ne.

S obzirom na to da potpisi u ovom skupu često sadrže i ime i prezime, sistem prepoznaje i izdvaja dve regije. Te regije se spajaju u jedan potpis, i u zavisnosti od faze projekta, šalju se kao primer u trening skup ili u test skup podataka.

3.1. MATERIJALI

Materijali za ovaj projekat prikupljani su tokom perioda od sedam meseci, počevši od 12. decembra 2024. godine, zaključno sa 31. julom 2025. godine. U prikupljanju podataka je učestvovalo 13 volontera, od čega 8 muškog i 5 ženskog pola, starosne dobi u rasponu od 20 do 24 godine.

Ukupno je prikupljeno 30 fotografija ručno pisanih potpisa, 15 slika predstavlja prave potpise a ostalih 15 falsifikate. U radu su pravi potpisi označavani kao original a falsifikati kao fake. Za potrebe snimanja korišćeno je ukupno šest kamera. Pet različitih uređaja je korišćeno za slikanje 9 potpisa, dok je preostalih 21 slika zabeleženo jednom kamerom. Tri fotografije su napravljene kamerom visoke rezolucije od 50 megapiksela, dok su ostale snimane uređajima sa rezolucijom od 12 megapiksela i nižom.

Distibucija slika je sledeća: 24% slika je u visokoj rezoluciji od 3768×8160 ili 8160×3768 piksela, 46% u rezoluciji 4000×1848 ili obrnuto, dok je za preostalih 30% korišćena raznovrsna rezolucija

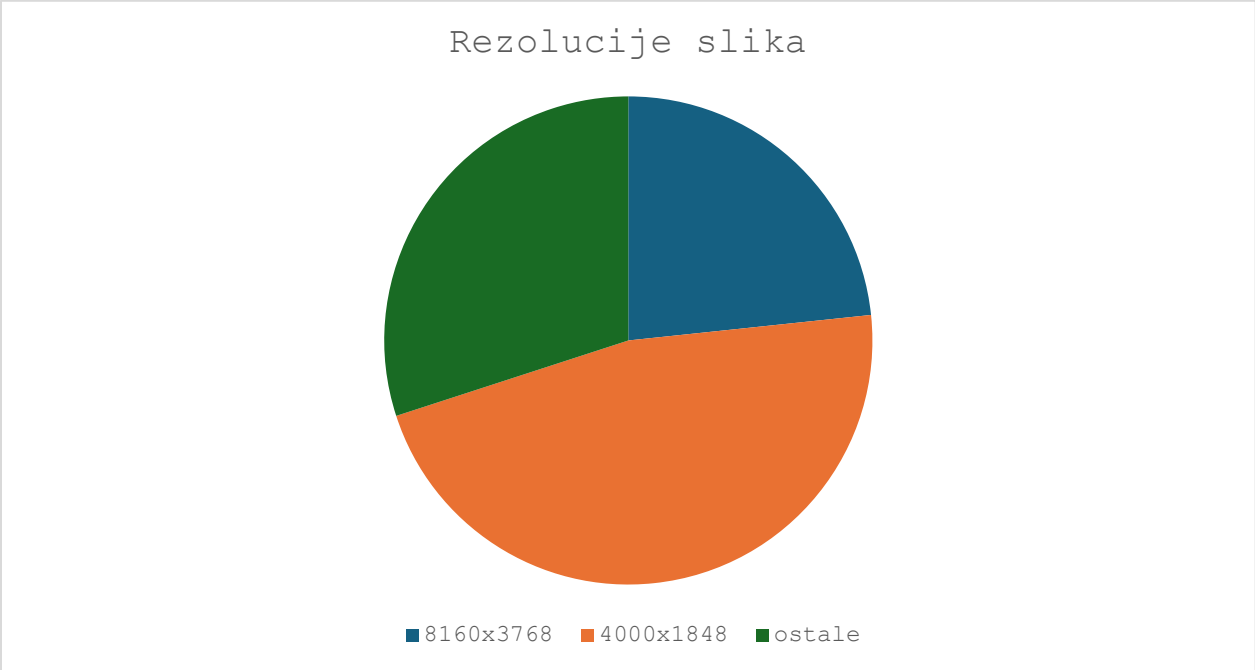
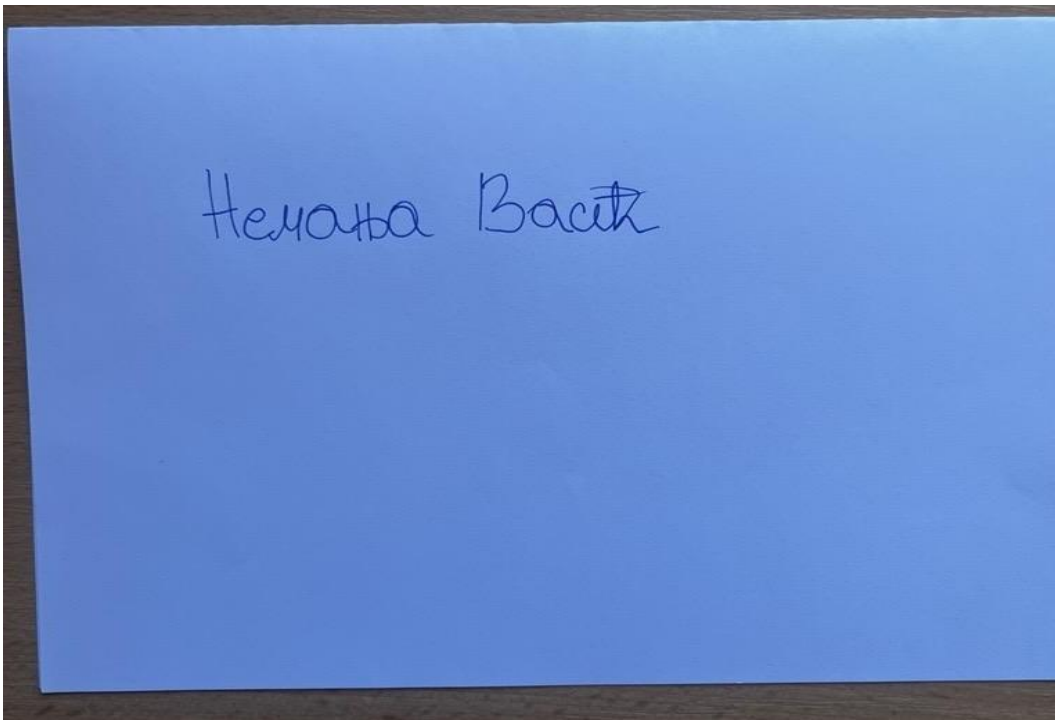


Figura 3.1. Raspodela rezolucija slika potpis

Većina slika (preko 70%) slikano je pod blagim uglom, kako bi se minimizovao uticaj senki. Potpisi su ispisani na belom papiru. U 21 slučaja korišćena je tehnička olovka sa minom poluprečnika 0.75mm, dok je za preostalih 9 korišćena hemijska olovka.

Senke su česta pojava na slikama. Na 6 slika senke prekrivaju više od 20% porvšine. Kod ostalih slika senke su slabog inteziteta ili potpuno izostaju. Dodatno 4 slike imaju plavičasto osvetljenje, što je uticalo na procesiranje tih slika.



Slika 3.1. Slika koja ima izraženu plavičastu pozadinu

3.1.2. Pred procesiranje

U okviru predprocesiranja, primenjuje se ***Multi-Scale Retinex*** (MSR) filter sa ciljem redukcije uticaja senki i neujednačenog osvetljenja koje može otežati detekciju potpisa. Ovaj filter omogućava očuvanje lokalnog kontrasta o poboljšanje kvaliteta slike, naročito u delovima sa jakim senkama. U implementaciji su korišćene tri vrednosti za Gaussovu standardnu devijaciju (σ = 15, 80, 200), čime se postiže balans između globalnog i lokalnog osvetljenja slike.

U ranijim verzijama projekta korišćene su i dodatne metode, kao što su ***gama korekcija*** i morfološka ***dilacija***, međutim ove metode su izbačene nakon eksperimentalne evaluacije.

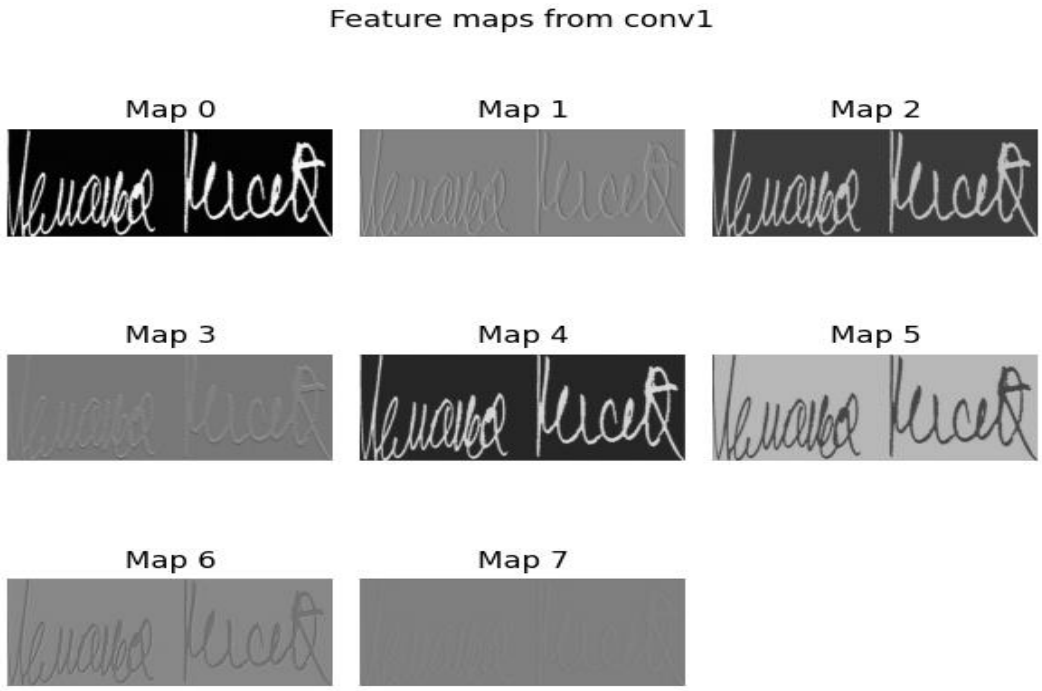
Gama korekcija je uklonjena jer se pokazalo da MSR filter samostalno pruža bolje rezultate na slikam sa izraženim senkama, bez potrebe za dodatnim korekcijama.

Dilacija, iako doprinosi boljoj istaknutosti potpisa, istovremeno je dovodila do narušavanja morfoloških karakteristika linija potpisa, što negatino utiče na kasnije faze validacije i klasifikacije.

3.2. CNN arhitektura

Konvoluciona neuralna mreža koja se koristi u projektu ima jedan ulazni sloj, tri skrivena (hidden) i jedan output layer. Kod ulaznog i sakrivenih slojeva se koristila **RELU** funkcija aktivacija. Dok se kod klasikatora se koristi ***sigmoid*** funkcija kako bi se klasifikovalo da li je potpis original ili fake.

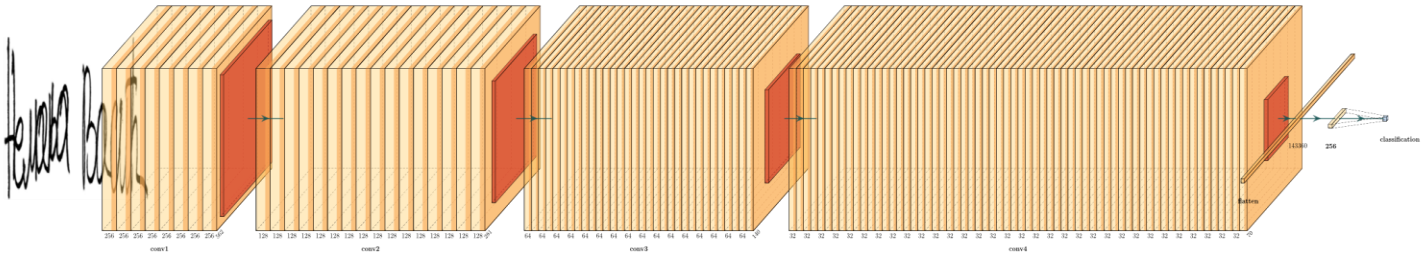
Input layer je tenzor veličine 1124x512x1, 1124x512 je veličina slike na ulazu a 1 je treća dimenzija slike jer je gray scale (da je slika RGB onda bi bilo 1124x512x3). Slika se potom smanjuje i izvlače se karakteristike (feature) mape. U prvom (input) sloju ih je 8.



Slika 3.2.1. Prikaz feature mapa koje izvlači input layer

Drugi sloj prihvata duplo manju rezoluciju slike 562x256. Iz nje izlaze 16 karakterističnih slika. Tako će isto uraditi i treći i četvrti sloj. Četvrti sloj će imati 64 feature mape rezolucija 32x70.

One potom se pretvaraju u niz i nad njima se vrši klasifikacija tako da se na kraju dobijaju vredosti između 0 i 1 koje se pomoću sigmoid funkcije stavljaju na 0 ili 1.



Slika 3.2.2. Arhitektura CNN korišćena u projektu

3.3. Siamese arhitektura

Ova siamska neuronska mreža projektovana je kao simetrična arhitektura koja se sastoji od dva identična pod-sistema koji dele iste težine i parametre. Njena osnovna namena je da nauči mapiranje slika potpisa u zajednički vektorski prostor gde se sličnost dva potpisa može direktno meriti. Arhitektura je podeljena na konvolucioni deo za ekstrakciju vizuelnih karakteristika i potpuno povezani deo za generisanje finalne vektorske ugradnje.

Konvolucioni segment mreže koristi četiri uzastopna bloka koji progresivno povećavaju broj kanala sa 1 do 128. Svaki blok se sastoji od konvolucionog sloja, aktivacione funkcije ReLU i sloja za smanjenje prostornih dimenzija. Ovaj deo mreže funkcioniše kao hijerarhijski ekstraktor obeležja, gde početni slojevi detektuju jednostavne oblike poput ivica i poteza olovke, dok dublji slojevi prepoznaju kompleksnije teksture i specifičnosti rukopisa. Poseban tehnički detalj u kodu je korišćenje probnog prolaza kroz mrežu kako bi se precizno izračunala dimenzija podataka pre ulaska u potpuno povezane slojeve, čime se osigurava stabilnost arhitekture pri različitim ulaznim rezolucijama.

Nakon ekstrakcije obeležja, mreža koristi niz linearnih slojeva sa mehanizmom regularizacije u vidu izbacivanja neurona (Dropout) od 15%, što pomaže u sprečavanju preprilagođavanja modela na specifičnim primercima iz trening skupa. Finalna vektorska reprezentacija potpisa sažeta je u 256 elemenata. U samom procesu poređenja, model izračunava apsolutnu razliku između vektora dva ulazna potpisa, a dobijeni vektor razlike se prosleđuje u izlazni sloj sa sigmoidnom aktivacijom. Rezultat je verovatnoća u rasponu od 0 do 1, koja kvantifikuje stepen sličnosti, odnosno razlikovanja autentičnog potpisa od falsifikata.

Proces obučavanja se oslanja na binarnu unakrsnu entropiju kao funkciju cilja i Adam optimizator. Tokom svake epohe, model ažurira svoje težine na osnovu greške koju pravi pri klasifikaciji parova slika, težeći da smanji ukupni gubitak. Program automatski koristi procesorsku snagu grafičke kartice za ubrzanje računanja, a istorija gubitaka se čuva u spoljnu datoteku kako bi se omogućilo praćenje napretka i konvergencije modela tokom vremena.

3.4. Triplet mreža

Arhitektura implementiranog rešenja bazirana je na konvolucionoj neuronskoj mreži (CNN) koja se sastoji od četiri konvoluciona bloka. U prvom bloku korišćeno je 16 filtera, u drugom 32, u trećem 64 i u četvrtom 128 filtera, čime se postepeno povećava dubina izdvajanja karakteristika. Nakon svakog konvolucionog bloka primenjena je ReLU aktivaciona funkcija i max-pooling sloj za redukciju dimenzionalnosti. Na kraju mreže dodata su dva potpuno povezana sloja koji sliku potpisa preslikavaju u vektor embeddinga dimenzija 128.

Mreža koristi deljene težine (shared weights), što znači da se isti parametri primenjuju za obradu svih ulaznih slika. Za treniranje modela korišćena je Triplet Margin Loss funkcija gubitka sa marginom od 1.0, koja omogućava da se embedding vektori istih potpisnika grupišu, a vektori falsifikata razdvoje. Optimizacija je vršena Adam algoritmom sa početnom stopom učenja od 0.001.

U fazi verifikacije, autentičnost potpisa utvrđuje se izračunavanjem Euklidske distance između embedding vektora dva potpisa. Ukoliko je rastojanje manje od prethodno kalibrisanog praga osetljivosti, potpis se smatra autentičnim. Prag je određen eksperimentalno na validacionom skupu podataka kako bi se postigao optimalan balans između preciznosti i odziva sistema.

4. Rezultati i diskusija

Pošto ima ukupno 30 primera u bazi koristilo se cross-validation kao metod evaluacije ovog modela. Koristio se 5-fold cross-validation. Podeljena je baza primera u 5 grupa, gde će 24 primera ići na treniranje a 6 na testiranje modela. Grupe su opisane po broju original i fake primera, raspodelu možete videti na tabeli (4.1).

	testSet1	testSet2	testSet3	testSet4	testSet5
original	3	4	4	2	2
fake	3	2	2	4	4

Tabela 4.1. Opis datasetova

Postavljena su dva slučaja, jedan kada je epoha stavljena na 10 a drugi kada je onda stavljena na 26. Primenom klasične

4.1. REZULTATI PRIMENOM CNN

CNN arhitekture mi dobijamo sledeće rezultate

EPOHA=10	testSet1	testSet2	testSet3	testSet4	testSet5
Accuracy	50%	83.33%	100%	50%	33%
Precision	50%	80%	100%	40%	33%
Recall	100%	100%	100%	100%	100%

Tabela 4.1.1 Prikaz metrika za EPOHA = 10

EPOHA=26	testSet1	testSet2	testSet3	testSet4	testSet5
Accuracy	100%	100%	100%	67.67%	67.67%
Precision	100%	100%	100%	50%	50%
Recall	100%	100%	100%	100%	100%

Tabela 4.1.1 Prikaz metrika za EPOHA = 26

Kada se stavi veći broj epoha dobijamo bolje rezultate. Vidi se da je testSet4 i testSet5 daju najlošije rezultate to je zato što oni imaju 4 fake primera. CNN klasifikuje sve potpise u originale, što preicision i recall metrike govore. Kao sistem ovo nije najbolje rešenje, ovako bi bilo ko mogao da falsifikuje tuđi potpis.

4.2. REZULTATI PRIMENOM SIAMESE MREŽE

I ovde su podaci podeljeni isto kao i kod CNN. Jedina razlika je što se prave kombinacije slika. Tako da u training fazi ima 48 primera a u test 12. Primeri su raspoređeni tako da bude jedan fake drugi original i tako dalje.

4.2.1. Treniranje sa 10 Epoha

Kada se trenira sa 10 epoha dobija se sledeća confusion matrix za dataSet1

EPOHA = 10	(Prediction) Original	(Prediction) Fake
(Ground Truth) Original	2	3
(Ground Truth) Fake	0	7

Tabela 4.2.1. Confusion matrix za dataSet1 Epoha = 10

Kao što se vidi tačnost je 75% preciznost 100% ali je odziv svega 40%, iako sistem nije nijedan fake primer klasifikovao kao original on je 3 original primera klasifikovao kao fake.

Pri dataSet2 dobija se sledeće

EPOHA = 10	(Prediction) Original	(Prediction) Fake
(Ground Truth) Original	0	6
(Ground Truth) Fake	0	6

Tabela 4.2.2. Confusion matrix za dataSet2 Epoha = 10

Sada imamo tačnost od 50% preciznost je 0 a odziv je takođe nula, ovaj model je potpuno beskoristan jer sve predstavlja kao fake.

Pri dataSet3 dobija se

EPOHA = 10	(Prediction) Original	(Prediction) Fake
(Ground Truth) Original	4	2
(Ground Truth) Fake	0	6

Tabela 4.2.3. Confusion matrix za dataSet3 Epoha = 10

Tačnost je 83% preciznost je 100% a odziv je 60%. Imamo za sada najbolje rezultate u modelu, iako je recall i dalje loš bolje je da model čak i neki originalni potpis klasifikuje kao fake nego obrnuto.

Pri dataSet4 dobija se

EPOHA = 10	(Prediction) Original	(Prediction) Fake
(Ground Truth) Original	6	0
(Ground Truth) Fake	2	4

Tabela 4.2.4. Confusion matrix za dataSet4 Epoha = 10

Tačnost je 83%, preciznost je 75% a odziv je 100%. Pri ovom dataSetu se desilo par neočekivanih stvari pri prvom pokretanju modela dobija se da on klasifikuje sve kao fake, na sledecem treniranju dobio je sve je original na sledeće sve tačno prepozna. Vidi se da moć modela zavisi dosta od početnih vrednost parametara.

Za dataSet5 dobija se

EPOHA = 10	(Prediction) Original	(Prediction) Fake
(Ground Truth) Original	2	4
(Ground Truth) Fake	2	4

Tabela 4.2.5. Confusion matrix za dataSet5 Epoha = 10

Tačnost i preciznost su 50% a odziv je 40%.

Srednja tačnost modela za epohu 10 je 68.33% preciznost je 77% a odziv je 49%.

4.2.2. Treniranje sa 26 epoha

Kada se trenira dataSet1 dobija se sledeće

EPOHA = 26	(Prediction) Original	(Prediction) Fake
(Ground Truth) Original	5	0
(Ground Truth) Fake	0	7

Tabela 4.2.1. Confusion matrix za dataSet1 Epoha = 26

Dobija se savršeno predviđanje. Za dataSet2 se dobija sledeće

EPOHA = 26	(Prediction) Original	(Prediction) Fake
(Ground Truth) Original	4	2
(Ground Truth) Fake	0	6

Tabela 4.2.2. Confusion matrix za dataSet2 Epoha = 26

Za tačnost se dobija 83.33% preciznost 100% i odziv 60%. Za dataSet3 i dataSet4 se dobija

EPOHA = 26	(Prediction) Original	(Prediction) Fake
(Ground Truth) Original	6	0
(Ground Truth) Fake	0	6

Tabela 4.2.3. Confusion matrix za dataSet3 Epoha = 26

Za dataSet5 se dobija

EPOHA = 26	(Prediction) Original	(Prediction) Fake
(Ground Truth) Original	1	5
(Ground Truth) Fake	2	4

Tabela 4.2.4. Confusion matrix za dataSet5 Epoha = 26

DataSet5 je najlošiji dataSet za ovaj model dobija se tačnost od samo 42% preciznost je 33,33% a odziv je svega 18%. Srednja tačnost je 85% odziv je 75.8% a preciznost je 91%

4.3. REZULTATI PRIMENOM TRIPLET MREŽE

Kod triplet mreže podaci se moraju drugačije organizovati u svakom dataSet-u broj primera zavisi od broja fake primera, koliko fake primera toliko primera imamo u daSetu. Isto važi i za test primera.

4.3.1 Treniranje sa 10 Epoha

Kada se koristi dataSet1 dobijaju se sledeće distance

Distanca original primer	Distanca fake primer
0.1973	1.53
0.22	1.48
0.12	0.5775

Tabela 4.3.1.1. Distance koje model daje za Epoha = 10

Kao što se vidi model daje male distance do original primera dok dve od tri fake primera stavlja da imaju velike distance veće od 1.3 dok za jedan fake primer daje 0.6 distancu. Iako daje ovakve rezultate mi bismo mogli ako bismo stavili granicu (sve manje je original sve više fake) da dobijemo tačnostod 83% preciznost 100% i odziv 83%. A ako bismo smanjili distancu za ovaj model na 0.3 tačnost za ove slučajeve bi bila veća. Kada se ovaj model primeni na drugim datasetovima dobijaju se jako dobri rezultati. Čak i kada smanjimo broj epoha na 2 dobijaju se korisni rezultati

Distanca original primer	Distanca fake primer
0.0425	0.1064
0.0351	0.0455
0.0268	0.053

Tabela 4.3.1.2. Distance koje model daje za Epoha = 2

Kao što se vidi model daje male distance za original prikmera ali i za jedan fake primer 0.053. Razlika između distanca fake i original je dosta mala, moralo bi da se napravi granica koja bi bila 0.0450 da bi dobijao 100% tačnost. Kada se ovaj model primeni na drugim dataSetovima dobijaju se lošiju razultati granica postane toliko slepljena da se desi da model za fake kaže da je bliži nego original.

Distanca original primer	Distanca fake primer
0.25	0.3035
0.1505	0.0745
0.1244	0.1410

Tabela 4.3.1.3. Distance koje model daje za Epoha = 2

4.3.2 Treniranje sa 26 Epoha

Rezultati su bolji nego sa 10 epoha. Odnos između distanci fake i original primera je oko u najgorem slučaju tri puta a u najbolje 18 puta je distanca fake primera veća od original primera

Distanca original primer	Distanca fake primer
0.1131	1.79
0.1073	1.75
0.0813	0.48
0.482	1.7536
0.1067	1.8986
0.1332	1.9027
0.3992	1.8695
0.0318	0.2739
0.0337	0.3320
0.0691	0.3330

Tabela 4.3.2.1. Distance koje model daje za Epoha = 26 kod raznih datasetova

Kod nekih datasetova distance se razlikuju za jedan i nešto dok kod drugih tistance se razlikuju za 0.2, mana ovoga je što bi moralo da se pravi granica na licu mesta kada se vide podaci. Iako na većini primera da se granica stavi da bude 0.5 bi bilo zadovoljavajuće kao što se vidi u nekim slučajevima, ali u drugim bi pravio problem.

4.4. UPOREDA REZULTATA

Može se odmah primetiti da CNN kaska u odnosu na rezultate Siamese i Triplet mreža. Razlog za to leži u činjenici da CNN model rešava problem kao zadatak višeklasne klasifikacije, pri čemu se optimizuje odluka o pripadnosti klasi, a ne međusobna sličnost potpisa.

EPOHA = 26	CNN	Siamise
Accuracy	86.67%	85%
Precision	78%	91%
Recall	100%	75.8%

Tabela 4.4.1. Prikaz rezultata CNN i Siamese za EPOHA = 26

Glavna razlika između CNN i Siamise modela je što CNN klasifikuje određeni broj fake primera u original, dok Siamese je propustio u 60 primera svega dva. Iako je odziv sistema dosta manji nego kod CNN model je korisniji. Nama ima vrednost onaj model koji će što više falsifikata klasifikovati kako treba, čak iako deo originala bude klasifikovan kao fake.

Za Triplet mreže glavni izazov je dobar izbor sidra. Kada se dobro izabere dobija se jasna razlika između distanci fake i original primera. Kao što se vidi u slučaju kada je anchor primer loš dobija se mala razlika između fake i original primera. U slučaju kada se napravi dobar dataset onda triplet mreže daju najbolje rezultate.

Distanca original primer	Distanca fake primer
0.5689	1.77
0.49	1.82
0.43	1.395
0.55	0.975
0.42	1.38
0.52	0.82

Tabela 4.4.2. Distance koje model daje za Epoha = 100 kod modifikovanog dataseta

4.5. Diskusija

Rezultati pokazuju da klasična CNN arhitektura ostvaruje najslabije performanse u poređenju sa modelima zasnovanim na učenju metrike. Razlog za to leži u činjenici da CNN model rešava problem kao zadatak višeklasne klasifikacije, pri čemu se optimizuje odluka o pripadnosti klasi, a ne međusobna sličnost potpisa. S obzirom na veliku intraklasnu varijabilnost rukopisa, ovakav pristup se pokazuje nedovoljno robusnim, naročito u prisustvu veštih falsifikata.

Sijamska mreža pokazuje značajno poboljšanje u odnosu na CNN. Korišćenjem kontrastivne funkcije gubitka, ovaj model uči direktnu metriku sličnosti između parova potpisa. Na taj način, problem verifikacije se modeluje prirodnije, jer sistem odlučuje da li su dva potpisa slična ili ne. Iako sijamska mreža ostvaruje bolje rezultate u pogledu tačnosti i EER vrednosti, njeno ograničenje leži u činjenici da se u svakom koraku posmatraju samo parovi uzoraka, bez eksplicitnog uvida u globalnu strukturu prostora osobina.

Najbolje performanse u eksperimentima ostvaruje triplet mreža. Ova arhitektura koristi triplet loss funkciju koja istovremeno uzima u obzir referentni potpis (anchor), originalni potpis istog autora (positive) i potpis drugog autora ili falsifikat (negative). Takva formulacija omogućava

efikasnije razdvajanje klasa u embedding prostoru, jer se pozitivni uzorci privlače, dok se negativni potiskuju na veću udaljenost. Kao rezultat, dobija se kompaktniji i bolje separabilan prostor osobina, što direktno utiče na smanjenje EER vrednosti i povećanje tačnosti sistema.

Na osnovu dobijenih rezultata može se zaključiti da pristupi zasnovani na metric learning paradigmi predstavljaju znatno pogodnije rešenje za problem offline verifikacije potpisa u poređenju sa klasičnim klasifikacionim CNN modelima. Posebno se ističe triplet mreža, koja se pokazala kao najefikasniji pristup u pogledu robusnosti, generalizacije i otpornosti na vešte falsifikate.

5. Zaključak

U ovom radu izvršena je eksperimentalna analiza tri različite arhitekture dubokog učenja za zadatak offline verifikacije potpisa: klasične konvolucione neuronske mreže (CNN), sijamske mreže i triplet mreže. Cilj poređenja bio je da se ispita uticaj različitih paradigmi učenja na tačnost i pouzdanost sistema za verifikaciju potpisa. Najbolje rezultate daje triplet mreža pa sijamska i na kraju klasična CNN mreža.

Neki budući radovi bi bili prikupljanje više potpisa radi verodostojnije procene rezultata, promena detekcije potpisa, korišćenje modifikovanih triplet arhitektura radi boljeg postavljanja praga.

LITERARURA:

[1] T. A. Farag and A. A. El-Sayed, "Offline signature recognition using convolutional neural network," *Procedia Computer Science*, vol. 170, pp. 398–403, 2020. [Online]. https://doi.org/10.1016/j.procs.2020.03.091

[2] L. G. Hafemann, R. Sabourin and L. S. Oliveira, "Learning Features for Offline Handwritten Signature Verification using Deep Convolutional Neural Networks," *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2016.

[3] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger and R. Shah, "Signature Verification Using a 'Siamese' Time Delay Neural Network," in *Advances in Neural Information Processing Systems (NeurIPS 6)*, Morgan Kaufmann, 1994, pp. 737–744.

[4] S. Chattopadhyay, S. Bhattacharya, S. Manna and U. Pal, "SURDS: Self-Supervised Attention-guided Reconstruction and Dual Triplet Loss for Writer Independent Offline Signature Verification," arXiv preprint arXiv:2201.10138, 2022.

[5]] M. A. Nielsen, *Neural Networks and Deep Learning*, 2015. pp. 305-322

[6] Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a Similarity Metric Discriminatively, with Application to Face Verification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, 539-546. doi: 10.1109/CVPR.2005.202

[7] Hoffer, E., & Ailon, N. (2015). Deep metric learning using Triplet network. *In Workshop at the International Conference on Learning Representations (ICLR)*. arXiv:1412.6622

SLIKE:

[1] Gómez, Rodrigo & Lorentz, Joe & Hartmann, Thomas & Goknil, Arda & Singh, Inder & Halaç, Tayfun & Boruzanli Ekinci, Gülnaz. (2024). An AI pipeline for garment price projection using computer vision. Neural Computing and Applications. 36. 1-21. 10.1007/s00521-024-09901-w.

[2] Chopra, S., Hadsell, R., & LeCun, Y. (2005). Learning a Similarity Metric Discriminatively, with Application to Face Verification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, 539-546. doi: 10.1109/CVPR.2005.202

[3] Hoffer, E., & Ailon, N. (2015). Deep metric learning using Triplet network. *In Workshop at the International Conference on Learning Representations (ICLR)*. arXiv:1412.6622.