SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 5117

Interaktivna semantička segmentacija rukom pisanih znakova

Katarina Blažić

Zagreb, lipanj 2017.

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA
ODBOR ZA ZAVRŠNI RAD MODULA

Zagreb, 10. ožujka 2017.

ZAVRŠNI ZADATAK br. 5117

Pristupnik: **Katarina Blažić (0036485710)**
Studij: Računarstvo
Modul: Računarska znanost

Zadatak: **Interaktivna semantička segmentacija rukom pisanih znakova**

Opis zadatka:

Semantička segmentacija rukom pisanog teksta je neriješen problem računalnog vida s mnogim zanimljivim primjenama. U posljednje vrijeme najbolji rezultati u tom području postižu se pristupima utemeljenima na dubokim konvolucijskim modelima. Međutim, za učenje i evaluiranje konvolucijskih modela potrebno je pribaviti veliki broj slika označenih na razini piksela. Tema ovog rada je razvoj interaktivnog programa za učinkovito pribavljanje takvih oznaka.
U okviru rada, potrebno je proučiti biblioteke programskog jezika Python za rukovanje i vizualizaciju matrica i slika. Izraditi alat za rijetko nepotpuno označavanje rukopisa te vizualizaciju i popravljanje postojećih oznaka. Izraditi jednostavan klasifikator pojedinačnih piksela odnosno njihovih malenih susjedstava. Ispitati intenzitet označavanja za postizanje vizualno prihvatljive segmentacije rukopisa. Prikazati i ocijeniti ostvarene rezultate.
Radu priložiti izvorni i izvršni kod razvijenih postupaka, ispitne slijedove i rezultate, uz potrebna objašnjenja i dokumentaciju. Citirati korištenu literaturu i navesti dobivenu pomoć.

Zadatak uručen pristupniku: 10. ožujka 2017.
Rok za predaju rada: 9. lipnja 2017.

Mentor:
Izv. prof. dr. sc. Siniša Šegvić

Djelovođa:
Doc. dr. sc. Tomislav Hrkać

Predsjednik odbora za završni rad modula:
Prof. dr. sc. Siniša Srbljić

Zahvaljujem mentoru prof. dr. sc. Siniši Šegviću na smjernicama i pomoći pri izradi ovog rada. Veliko hvala i mojoj obitelji na bezuvjetnoj podršci tijekom cijelog školovanja.

SADRŽAJ

1. Uvod 1

2. Histogram boje 2

3. Bayesova formula 4

4. Bayesov klasifikator 5

5. Umjetne neuronske mreže 7

6. Konvolucijske neuronske mreže 9

6.1. Konvolucijski sloj 9

6.2. Sloj sažimanja 11

6.3. Potpuno povezani sloj 11

6.4. Mreža VGG16 12

7. Označavanje slika 13

8. Eksperimentalni rezultati 14

8.1. Rezultati statističke Bayesove klasifikacije 15

8.2. Rezultati s konvolucijskim neuronskim mrežama 21

9. Zaključak 23

1. Uvod

Računalni vid je područje umjetne inteligencije kojem je cilj prepoznavanje predmeta u slikama i razumijevanje slika. Obuhvaća metode za stjecanje, obradu, analizu i razumijevanje slike s ciljem prikupljanja informacija koje se mogu iskoristiti u učenju i donošenju zaključaka. Ljudi s lakoćom percipiraju i razlikuju stvari u svojoj okolini, ali za računala je to težak zadatak. U računalnom vidu zato nastojimo opisati svijet koji vidimo na slikama i iz opisa rekonstruirati njegova svojstva poput oblika, distribucije. Računalni vid danas ima široke primjene poput medicinske obrade slika u svrhu donošenje dijagnoza, određivanja sadrži li slika određeni objekt, čitanja rukom pisanih brojeva, prepoznavanja lica.

Segmentacija je zadatak traženja grupe piksela koji “ide zajedno”, odnosno zadatak grupiranja piksela u klase. Semantička segmentacija rukom pisanog teksta je neriješeni problem računalnog vida. Cilj ovog rada je napraviti segmentaciju slika i grupirati piksele na slikama u dvije grupe, znakove i pozadinu. Segmentacija znakova je jedan od prvih koraka daljnje analize i obrade teksta.

Kako bi segmentacija bila moguća bilo je potrebno označiti slike na kojima bi se moglo učiti. S tim ciljem je u sklopu ovog rada izrađen alat za označavanje slika. Nakon prikupljanja oznaka analizirana su dva pristupa klasifikaciji: statistički pristup i pristup utemeljen na dubokim neuronskim mrežama. Prikazani su i komentirani rezultati oba pristupa i iz njih je vidljivo da su oba podjednako primjenjiva za segmentaciju teksta.

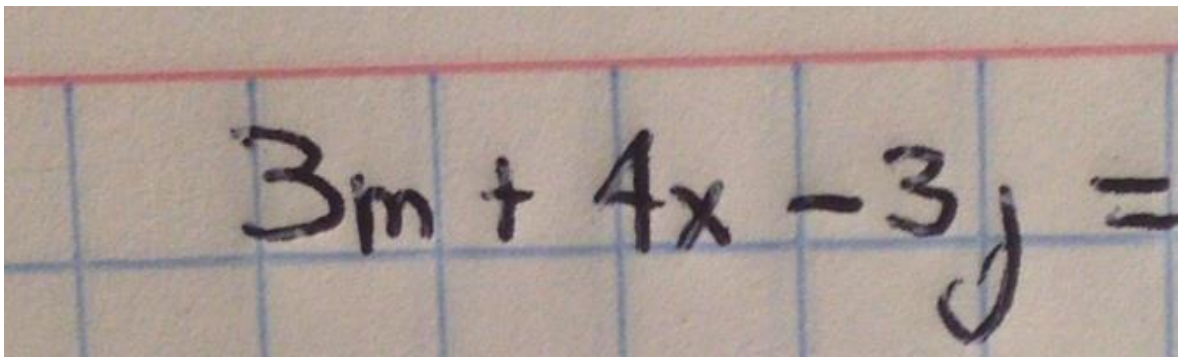
2. Histogram boje

Histogram je neparametarski prikaz distribucije boja na slici. Kod digitalnih slika histogram se može predstaviti kao niz pretinaca (eng. bin) koji sadrže broj piksele koji se nalaze u određenom rasponu vrijednosti. Primjerice ako imamo raspon boja 0-255 i 8 pretinaca u 1. pretincu će se nalaziti broj boja u rasponu 0-31.

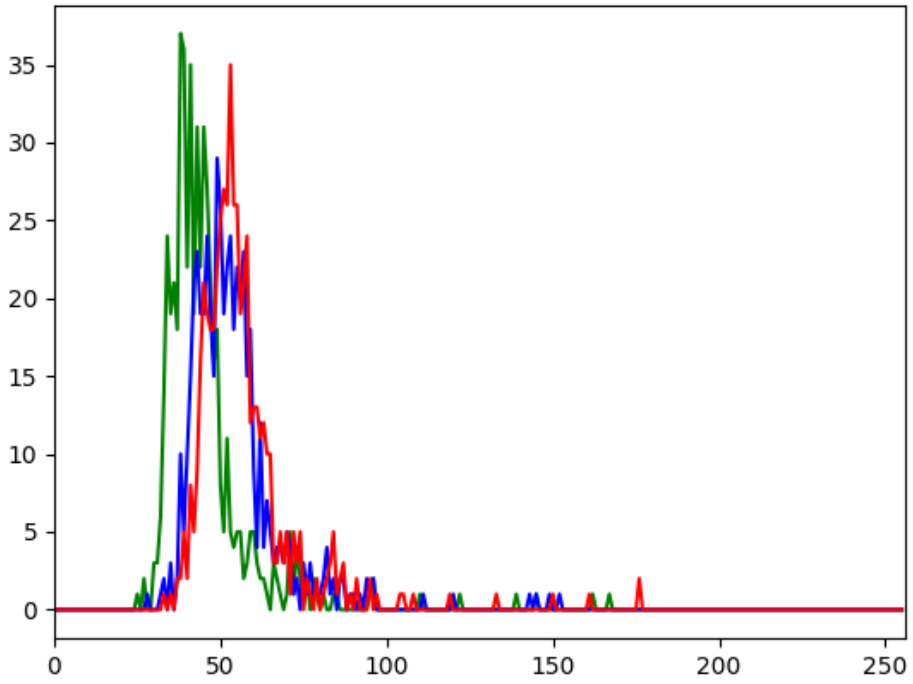
U ovom radu je korišten RGB prostor boja koji ima 3 kanala boja, jedan za crvenu, jedan za zelenu i jedan za plavu boju. Za pohranu boja korišteno je 8 bitova što nam daje 256 vrijednosti za svaku boju. Na slikama 2.2. i 2.3. su prikazani 2D histogrami texta i pozadine za sliku 2.1. Na horizontalnoj osi su naznačeni redni brojevi pretinaca, a na vertikalnoj osi broj piksela u pretincima. Svi pretinci su veličine 1.

U histogramima se može uočiti da je većina piksela teksta grupirana oko pretinca s rednim brojem 50 a pikseli pozadine su grupirani oko pretinca s rednom brojem 175. Ti pikseli će se dati lako klasificirati, a problem će stvarati pikseli koji se nalaze na prijelazu tih dviju regija, u ovom konkretnom primjeru pikseli u pretincima oko rednog broja 100.

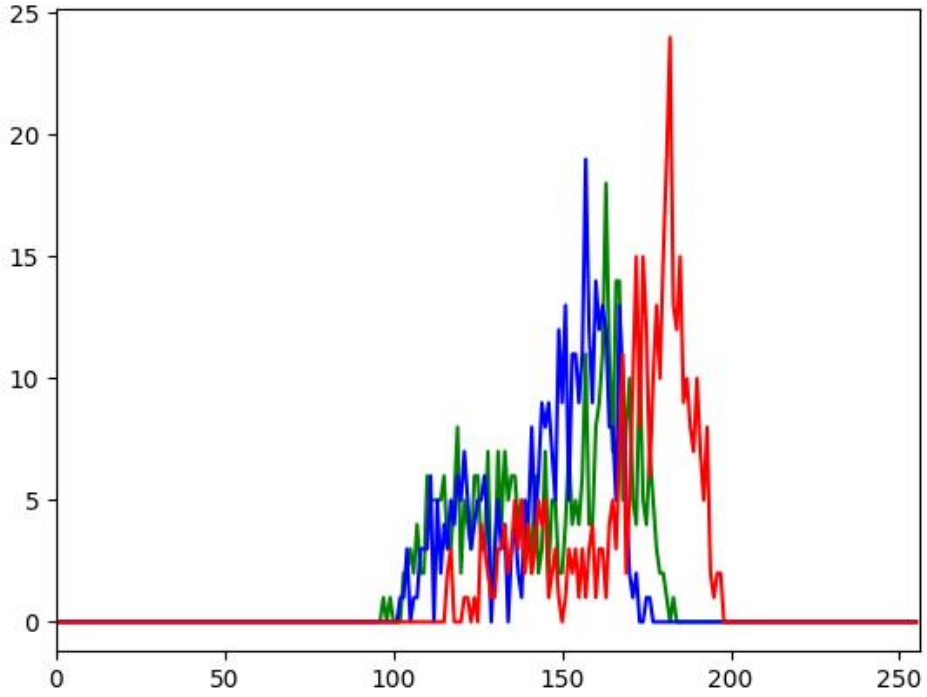
Za izradu histograma je korištena biblioteka OpenCV (Open Source Computer Vision).



Slika 2.1: Slika iz ulaznog skupa slika



Slika 2.2: Histogram teksta



Slika 2.3: Histogram pozadine

3. Bayesova formula

Neka je (H_i, i ∈ I), I ⊆ N, potpun sustav događaja u vjerojatnosnom prostoru (Ω, F, P) i neka je A ∈ F događaj takav da je P(A) > 0. Tada za svaki i ∈ I, I ⊆ N, vrijedi Bayesova formula:

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A)} = \frac{P(A|H_i)P(H_i)}{\sum_{j=1}^{n} P\left(A\middle|H_j\right)P(H_j)} \qquad (3.1)$$

Pri čemu su:

- $P(H_i)$ apriorna vjerojatnost hipoteze H_i,
- $P(A|H_i)$ izglednost hipoteze H_i,
- $P(H_i|A)$ aposteriorna vjerojatnost hipoteze H_i.

Bayesova formula služi za računanje aposteriornih vjerojatnosti hipoteza ako znamo da se dogodio događaj A. Pri tome je važno da hipoteze moraju biti međusobno isključive. Bayesova formula je dobro teorijski utemeljena ali zahtjeva veliku količinu podataka te a priori vjerojatnosti i izglednosti moraju biti unaprijed određene.

U našem primjeru hipoteze su H_1="piksel pripada razredu tekst" i H_2="piksel pripada razredu pozadina". Te hipoteze čine potpuni sustav događaja jer su tekst i pozadina jedini razredi koji se pojavljuju na ulaznom skupu slika te su ujedno i isključivi jer piksel ne može ujedno pripadati tekstu i pozadini. Kao događaje uzimamo RGB vrijednosti piksela za koje računamo aposteriorne vjerojatnosti. Jednadžba (3.1) prelazi u:

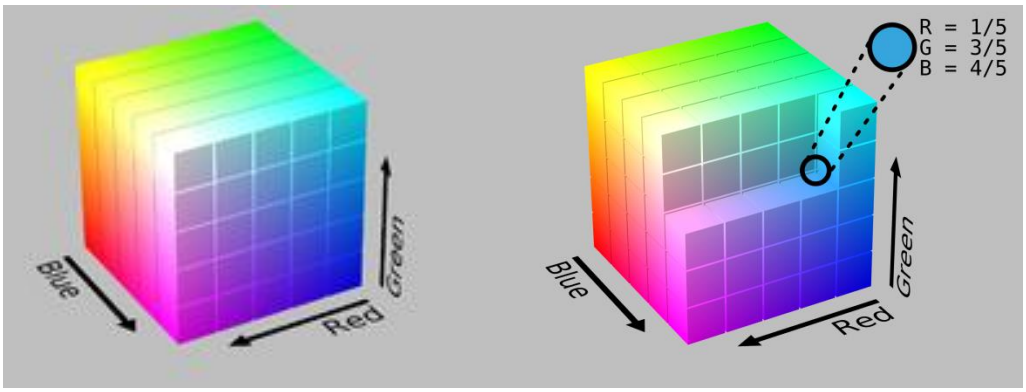
$$P(tekst|rgb) = \frac{P(rgb|tekst)P(tekst)}{P(rgb|tekst)P(tekst) + P(rgb|pozadina)P(pozadina)} \qquad (3.2)$$

$$P(pozadina|rgb) = \frac{P(rgb|pozadina)P(pozadina)}{P(rgb|tekst)P(tekst) + P(rgb|pozadina)P(pozadina)} \qquad (3.3)$$

4. Bayesov klasifikator

Želja nam je bila napraviti jednostavan klasifikator piksela koji će na temelju aposteriornih vjerojatnosti $P(H_i|rgb)$ klasificirati piksele kao dio tekst ili kao dio pozadine. Za izračun aposteriornih vjerojatnosti po Bayesovoj formuli najprije je potrebno odrediti apriorne vjerojatnosti i izglednosti hipoteza. Kao pomoć pri izračunu tih vjerojatnosti konstruirali smo 3D histograme teksta i pozadine.

3D histograme možemo predočiti kao kocke s tri osi. Svaka os predstavlja jedan od tri kanala boje RGB prostora. Kocka je duž svake osi popločena s pretincima koji čuvaju broj piksela čija vrijednost je u rasponu vrijednosti pretinca.



Slika4.1: 3D histogram boje [10]

Da bi se piksel klasificirao kao tekst mora vrijediti:

$$P(tekst|rgb) > P\ (pozadina|rgb) \tag{4.1}$$

Možemo raspisati te aposteriorne vjerojatnosti po Bayesovoj formuli i preoblikovati nejednadžbu tako da na lijevoj strani budu izglednosti hipoteza, a na desnoj apriorne vjerojatnosti:

$$\frac{P(rgb|tekst)P(tekst)}{P(rgb)} > \frac{P(rgb|pozadina)P(pozadina)}{P(rgb)} \tag{4.2}$$

$$P(rgb|tekst)P(tekst) > P(rgb|pozadina)P(pozadina) \tag{4.3}$$

$$\frac{P(rgb|tekst)}{P(rgb|pozadina)} > \frac{P(pozadina)}{P(tekst)} \tag{4.4}$$

Iz konstruiranih histograma možemo izračunati izglednosti hipoteza uz sljedeće oznake:

- t[rgb] broj piksela u rgb pretincu histograma teksta
- N_{tekst} ukupan broj piksela u histogramu teksta
- p[rgb] broj piksela u rgb pretincu histograma pozadine
- $N_{pozadina}$ ukupan broj piksela u histogramu pozadine

Dobivamo:

$$P(rgb|tekst) = \frac{t[rgb]}{N_{text}} \qquad P(rgb|pozadina) = \frac{p[rgb]}{N_{pozadina}} \tag{4.5}$$

Nadalje, iz histograma bismo mogli izračunati i apriorne vjerojatnosti hipoteza kao:

$$P(tekst) = \frac{N_{tekst}}{N_{tekst} + N_{pozadina}} \qquad P(pozadina) = \frac{N_{pozadina}}{N_{tekst} + N_{pozadina}} \tag{4.6}$$

Međutim, vrijednosti N_{tekst} i $N_{pozadina}$ ovise samo o tome koliko je piksela kojeg razreda korisnik označio i ne daju nam točne vjerojatnosti. Možemo uzeti da je omjer vjerojatnosti P(pozadina) i P(tekst) konstantan i iznosi Ɵ:

$$\frac{P(pozadina)}{P(tekst)} = \theta \tag{4.7}$$

Time smo problem klasifikacije sveli na provjeru istinitosti jednadžbe:

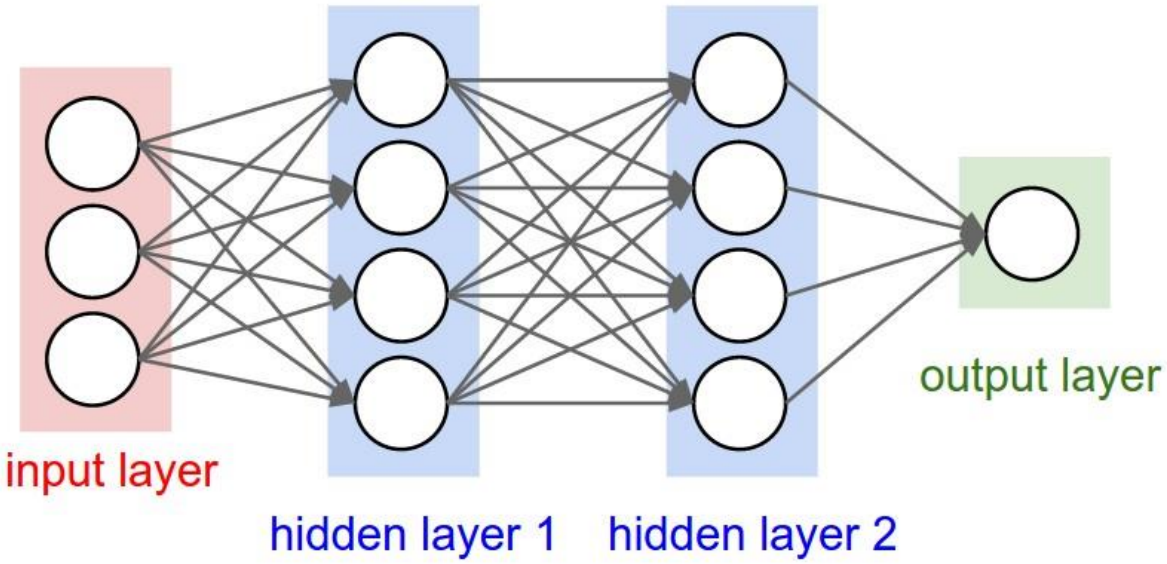
$$\frac{P(rgb|tekst)}{P(rgb|pozadina)} > \theta \tag{4.8}$$

Parametar Ɵ nazivamo prag (eng. threshold). Mijenjajući njegovu vrijednost možemo podešavati performanse klasifikacije.

5. Umjetne neuronske mreže

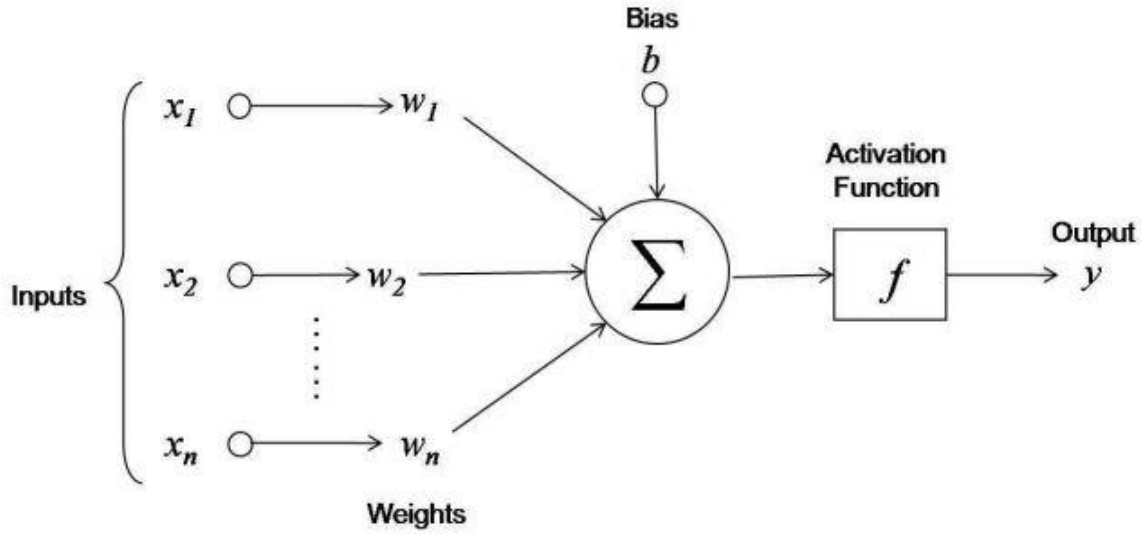
Umjetne neuronske mreže nastale su po uzoru na ljudski mozak koji možemo promatrati kao složeno računalo koje paralelno obavlja više nelinearnih operacija. Neuronske mreže su zamišljene da samostalno uče na temelju iskustva a ne da ih se programira. Prednost neuronskih mreža pred ekspertnim sustavima je da mogu raditi s nejasnim ili manjkavim podacima te da su dobre u procjeni nelinearnosti i prilagođavaju se okolini.

Umjetna neuronska mreža se obično sastoji od ulaznog sloja, skrivenih slojeva i izlaznog sloja. Mreža koja se sadrži 1 ili više skrivenih slojeva se naziva duboka neuronska mreža.



Slika 5.1: Duboka neuronska mreža [6]

Ljudski mozak se sastoji od oko 10 milijardi živčanih stanica – neurona. Građevna jedinica umjetnih neuronskih mreža je umjetni neuron prikazan na slici 5.2. Analogija između biološkog i umjetnog neurona je sljedeća: signali su numeričke vrijednosti, jakost sinapse opisuje težinski faktor w, tijelo stanice je zbrajalo a akson je aktivacijska funkcija f.



Slika 5.2: Umjetni neuron

Ulaz u neuronsku mrežu, odnosno u ulazni sloj, je skup vrijednosti [x_1, x_2, …, x_n], a ulaz u svaki sljedeći sloj je izlaz iz prethodnog sloja. Izlaz iz neurona računamo po formuli:

$$o = f(w_1x_1 + w_2x_2 + \cdots + w_nx_n + b) = f\left(\sum_{i=0}^{n} w_ix_i\right) \qquad (5.1)$$

Svaki neuron množi podatke s ulaza x_i s pripadnim težinama w_i (engl. weights) i umnošku pridodaje prag b (engl. bias). Na tu vrijednost primjenjuje aktivacijsku funkciju f koja daje izlaz iz neurona. Aktivacijska funkcija je najčešće nelinearna sigmoidalna funkcija: logistička funkcija ili tangens hiperbolni. Linearne aktivacijske funkcije nisu dobre iz razloga što su operacije u neuronima linearne i kada bi na njih primijenili linearnu funkciju ponovno bi dobili linearnu funkciju, a znamo da mozak obavlja nelinearne operacije.

Znanje mreže pohranjeno je implicitno u težinama veze između neurona. Te težine je potrebno učiti ili trenirati da bi ispravno okarakterizirale podatke s ulaza mreže. Kada se kaže učenje mreže misli se na iterativno predočavanje ulaznih podataka i eventualno očekivanih vrijednosti na izlazu. Metoda učenja kod koje se uz ulaz predočava i očekivani izlazi naziva se nadzirano učenje.

6. Konvolucijske neuronske mreže

Konvolucijske neuronske mreže uvode pretpostavku da je ulaz u mrežu slika što omogućava kodiranje određenih svojstava u mrežu, pojednostavljuje aktivacijsku funkciju te značajno smanjuje broj parametara u mreži. Konvolucijske mreže daju izrazito dobre rezultate prilikom klasifikacije objekata na slikama. Nedostatak potpuno povezane mreže je činjenica da zbog karakteristika mreže broj parametara naglo raste što može dovesti do prenaučenosti. Iz tog razloga su konvolucijske mreže prikladnije za rad sa slikama.

Neuroni su unutar slojeve konvolucijske mreže organizirani u 3 dimenzije: širinu, visinu i dubinu. Za razliku od potpuno povezanih slojeva kod kojih je neuron povezan sa svim neuronima iz prethodnog sloja, neuroni su u konvolucijskim slojevima povezani samo s malim brojem neurona iz prethodnog sloja.

Slojevi od kojih je konvolucijska mreža najčešće sastavljena su: konvolucijski sloj, sloj sažimanja i potpuno povezani sloj. Svaki sloj obavlja istu zadaću: transformira 3D ulaz u 3D izlaz primjenjujući pritom neku diferencijabilnu aktivacijsku funkciju.

6.1. Konvolucijski sloj

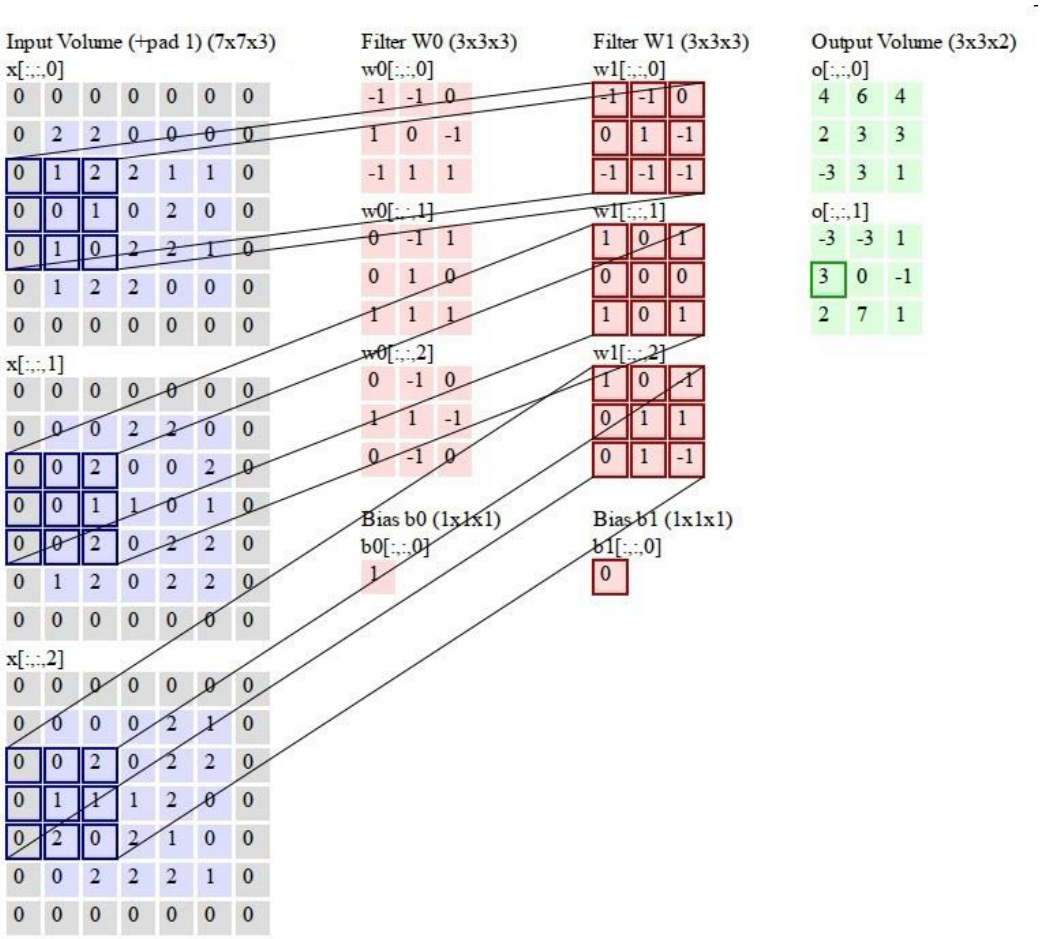
Konvolucijski sloj je dobio naziv po tome što koristi operaciju konvolucije umjesto matričnog produkta. Parametri konvolucijskog sloja su filteri koji sadrže težine koje je potrebno naučiti. Filteri obično imaju malu visinu i širinu, a dubine su jednake kao i ulaz u mrežu. Pri unaprijednom prolazu filter pomičemo s nekim korakom (engl. stride) po visini i širini ulaza i računamo skalarni produkt, odnosno konvoluciju između elemenata filtra i ulaza. Operacija konvolucije proizvodi 2-dimenzionalne aktivacijske mape koje daju odzive filtra. Za svaki filtar dobijemo po jednu aktivacijsku mapu, a izlaz iz mreže dobijemo tako da posložimo aktivacijske mape u dubinu. Konvolucija ima zanimljivo svojstvo ekvivarijantnosti s obzirom na pomak. Intuitivno to možemo tumačiti kao da se konvolucijski filtri “aktiviraju” na značajke koje su nam interesantne, a konvolucijski algoritmi uče na koje značajke

se filtri trebaju aktivirati. Izlaz mreže ne ovisi o tome gdje se značajke nalaze već samo o tome jesu li prisutne.

Izlaz iz konvolucijskog sloja računamo po formuli:

$$conv(I,K)_{xy} = \sigma\left(b + \sum_{i=0}^{h-1}\sum_{j=0}^{w-1}\sum_{k=o}^{d-1} K_{ijk} \cdot I_{x+i,y+j,k}\right) \tag{6.1}$$

Gdje I predstavlja ulaz (engl. input), K jezgru(engl. kernel) ili filtar dimenzija h×w×d, b prag filtra, σ aktivacijsku funkciju. Primjer izračuna izlaza iz konvolucijskog sloja s 2 filtera dan je na slici 6.1.

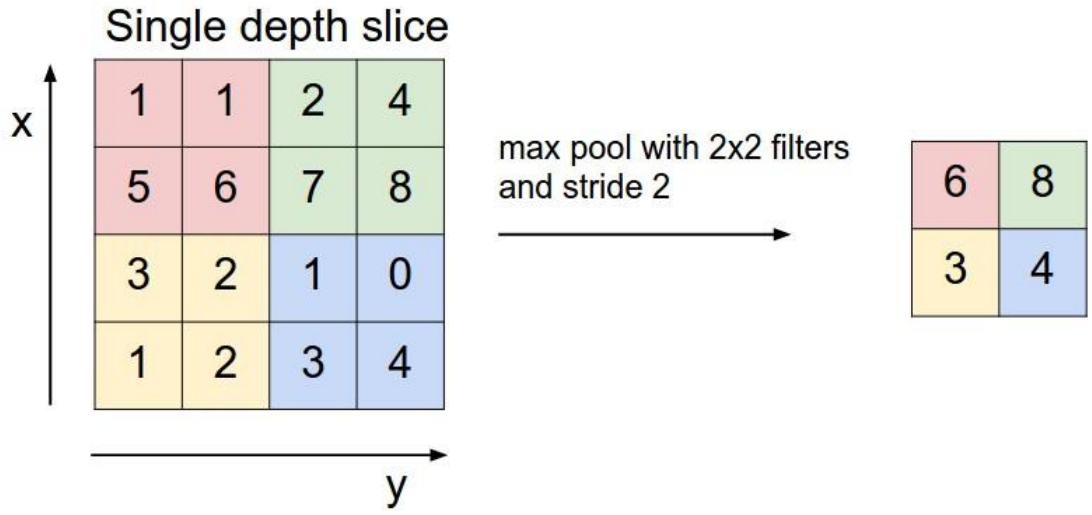


Slika 6.1: Konvolucijski sloj [6]

6.2. Sloj sažimanja

Uobičajeno se periodički između uzastopnih konvolucijskih slojeva umeće sloj sažimanja (engl. pooling layer). Njegova uloga je smanjivanje veličine aktivacijskih mapa i pomoć pri sprječavanju prenaučenosti mreže.

Sloj sažimanja mapira skup prostorno bliskih značajki na ulazu u jednu značajku na izlazu. Pri tome koristi neku nelinearnu funkciju, poput maksimalne vrijednosti ili aritmetičke sredine vrijednosti značajki. Najčešće se koriste slojevi sažimanja s regijama sažimanja veličine 2×2 i korakom 2. Takvi slojevi smanjuju dimenzije ulaza 2 puta po visini i 2 puta po širini, dok dubina ulaza ostaje nepromijenjena. U novijim klasifikacijskim modelima se koristi i globalno sažimanje koje za svaku mapu značajki daje po jednu vrijednost na izlazu. Sloj sažimanja pridonosi i invarijantnosti na pomak, što su veće regije sažimanja to je i invarijantnost veća. Primjer sažimanja s max funkcijom prikazan je na slici 6.2.



Slika 6.2: Sloj sažimanja [6]

6.3. Potpuno povezani sloj

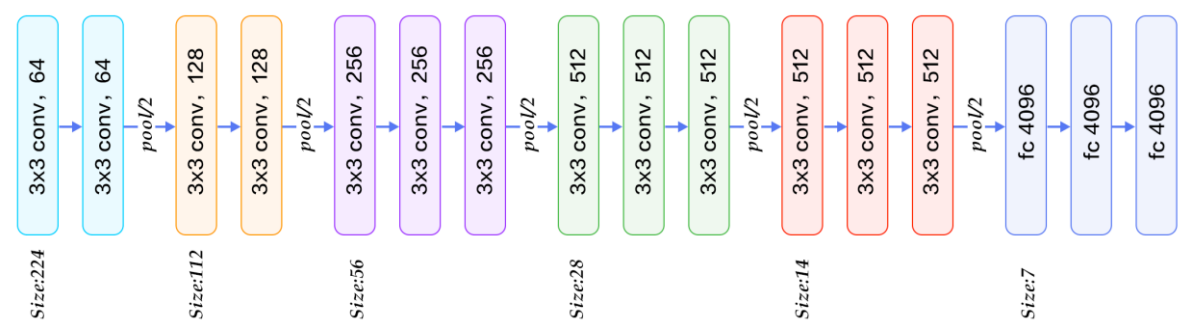
Potpuno povezani sloj obično dolazi na kraju konvolucijske mreže. Neuroni u potpuno povezanom sloju su povezani sa svim aktivacijskim funkcijama prethodnog sloja. Izlaz iz potpuno povezanih slojeva je N-dimenzionalni vektor, gdje je N broj različitih klasifikacijskih razreda. Nakon primjene neke

normalizacijske funkcije, primjerice softmaks, svaka vrijednost u vektoru predstavlja vjerojatnost pojedinog razreda. Nakon što konvolucijski sloj i sloj sažimanja detektiraju najvažnije značajke razreda zadaća potpuno povezanog sloja je da odredi korelaciju između značajki i razreda.

6.4. Mreža VGG16

Mrežu VGG (Visual Geometry Group) je razvila grupa istraživača sa Sveučilišta u Oxfordu. Mreža VGG je konvolucijska mreža trenirana na preko milijun slika iz izuzetno teškog seta slika ImageNet, koji sadrži 14 milijuna slika i 1000 različitih razreda slika. Ostvarila je klasifikacijske greške na 5 najčešćih razreda 7,5% na validacijskom skupu i 7,4% na testnom skupu.

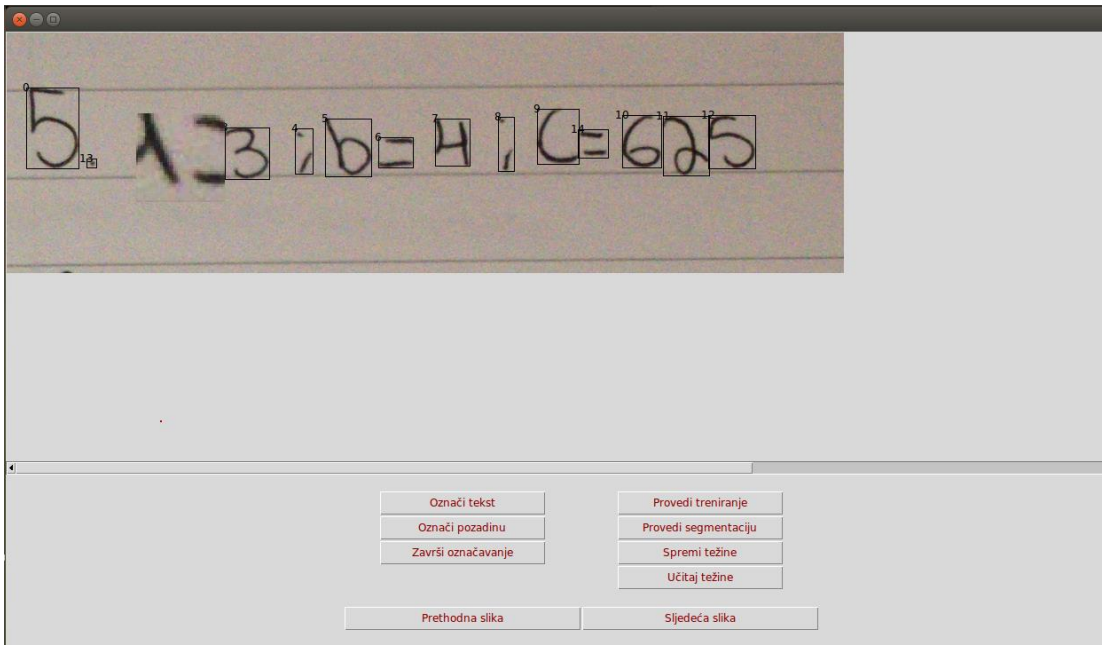
Autori VGG mreže Simonyan i Zisserman su pokazali da je dubina konvolucijske mreže ključna za dobre performanse. Najbolje rezultate su dobili s konvolucijskom mrežom od 16 slojeva: 13 konvolucijskih i 3 potpuno povezana sloja. Arhitektura mreže je izrazito homogena s filtrima veličine 3×3 i sažimanjima s regijama veličine 2×2 od početka do kraja. Loša strana VGG je ta da koristi mnogo memorije i parametara, oko 140M. Većina parametara se nalazi u potpuno povezanim slojevima.



Slika 6.3: Arhitektura VGG16 [11]

7. Označavanje slika

Za učenje i evaluaciju rezultata bilo je potrebno pribaviti velik broj slika označenih na razini piksela. U sklopu ovog rada napravljen je alat koji vizualizira postojeće okvire oko rukom pisanih znakova i omogućava označavanje piksela teksta i pozadine. Ideja nije u potpunosti označiti sliku nego ju rijetko označiti i tako ubrzati postupak prikupljanja oznaka. Program omogućuje zumiranje slika za veću preciznost označavanja. Na slikama ispod je prikazan postupak označavanja i jedna rijetko označena slika.



Slika 7.1: Postupak označavanja

Slika 7.2: Primjer rijetko označene slike

8. Eksperimentalni rezultati

Ulazni skup slika je podijeljen u 2 skupa: skup za učenje koji sadrži 30 slika i skup za testiranje koji sadrži 27 slika. Sljedeći rezultati su dobiveni na skupu od 30 rijetko označenih slika. Ukupno je označeno 8052 piksela teksta i 4992 piksela pozadine.

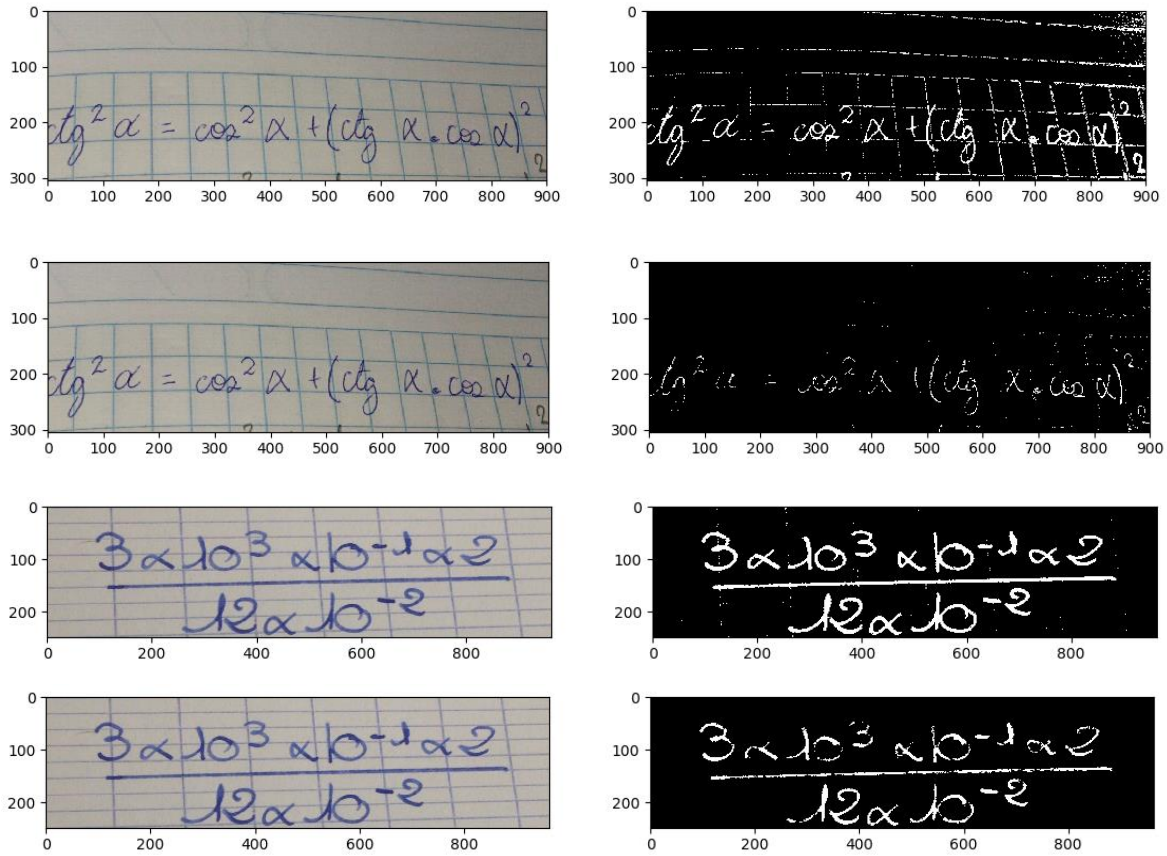
Rezultati segmentacije se obično vrednuju primjenom Jaccardovog indeksa piksela poznatog i pod nazivom presjek povrh unije (eng. intersection over union). Računa se za svaki razred po formuli:

$$točnost(X) = \frac{stvarni\ pozitivi(X)}{stvarni\ pozitivi(X) + lažni\ pozitivi(X) + \ lažni\ negativi(X)} \qquad (8.1)$$

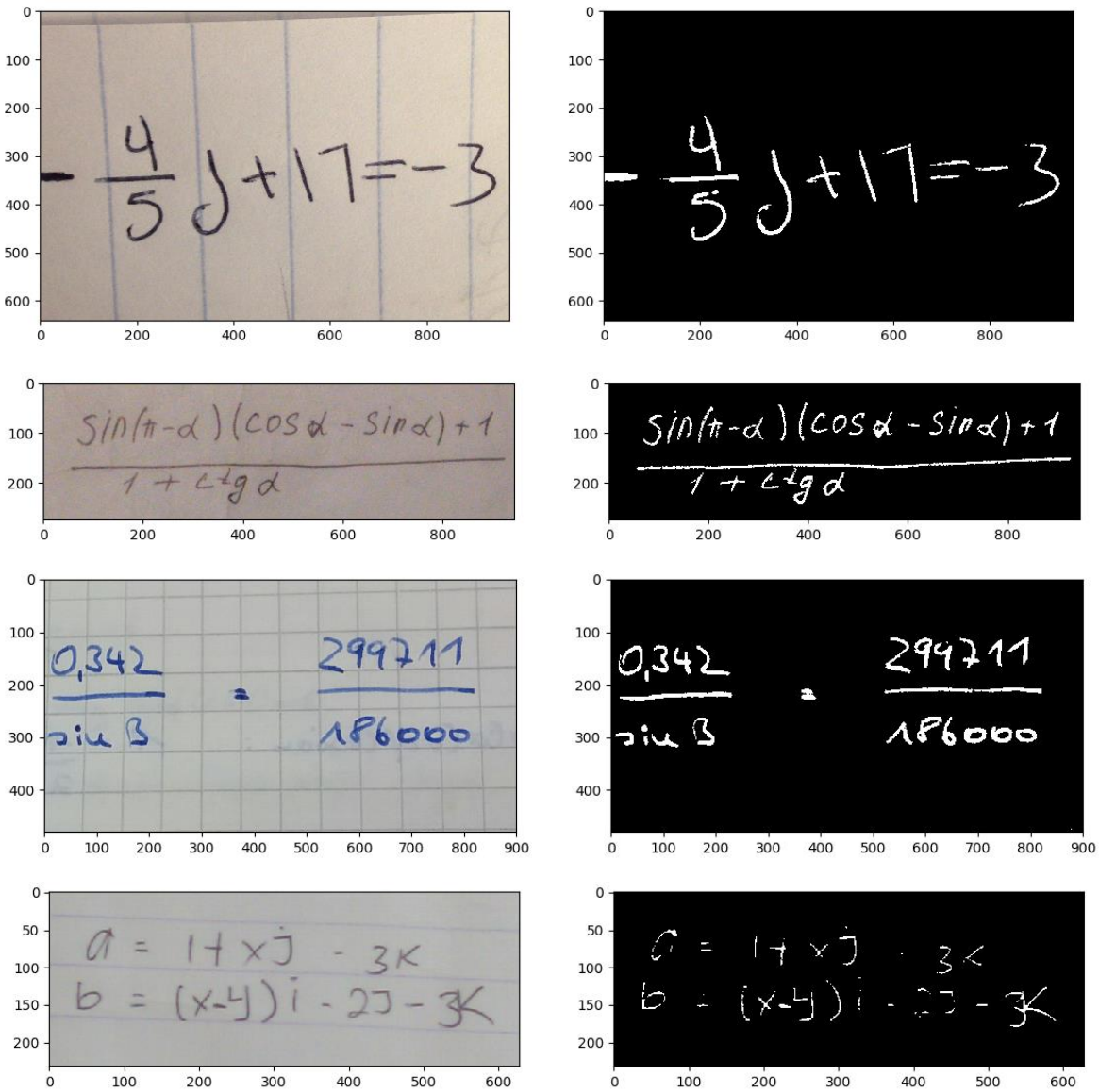
Gdje su stvarni pozitivi broj piksela razreda X koji je ispravno klasificiran, lažni pozitivi(X) broj piksela drugog razreda koji je su klasificirani kao razred X, lažni negativi(X) broj piksela razreda X koji je klasificiran kao neki drugi razred. Ukupna točnost segmentacije računa se kao prosječna vrijednost točnosti svih klasifikacijskih razreda. Međutim naše rezultate ne možemo vrednovati na taj način jer je jako teško označiti tekst i pozadinu na razini piksela zato ćemo samo komentirati dobivene rezultate.

8.1. Rezultati statističke Bayesove klasifikacije

Da bismo dobili najbolje rezultate bilo je potrebno odabrati adekvatnu veličinu pretinaca histograma i prilagoditi prag Θ. S obzirom na svega 13000 označenih piksela pretinci veličine 1 nisu imali pretjeranog smisla jer bi se moglo dogoditi da u nekim pretincima nema nijednoga piksela. Krenuli smo s pretincima veličine 32 i minimalnim pragom. Dobiveni rezultati prikazani su na slikama 8.1 i 8.2.

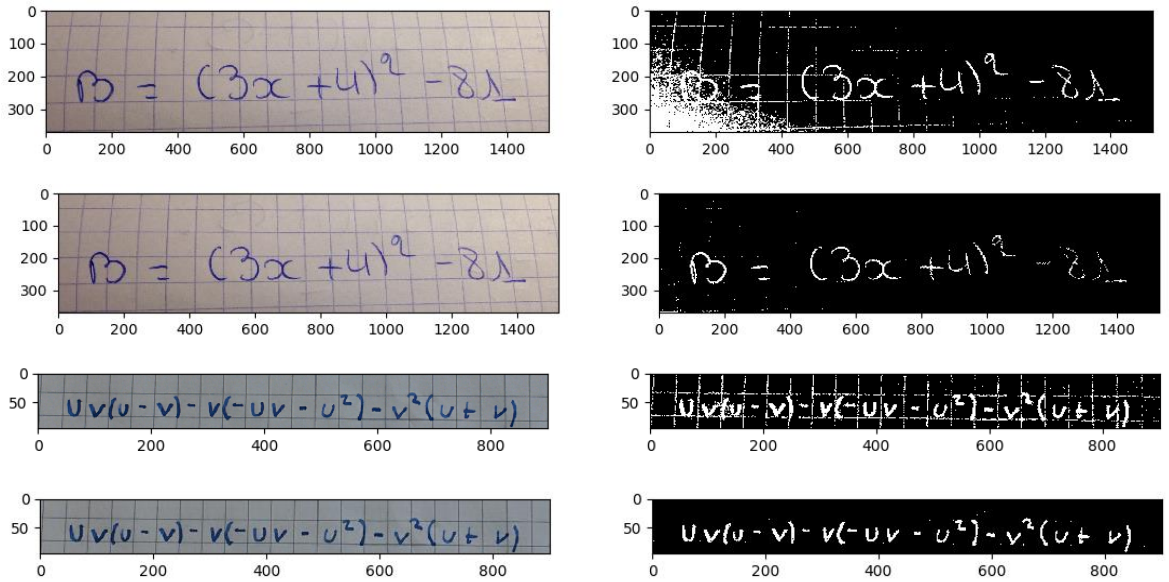


Slika 8.1: Primjer lažnih pozitiva uz pretince veličine 32

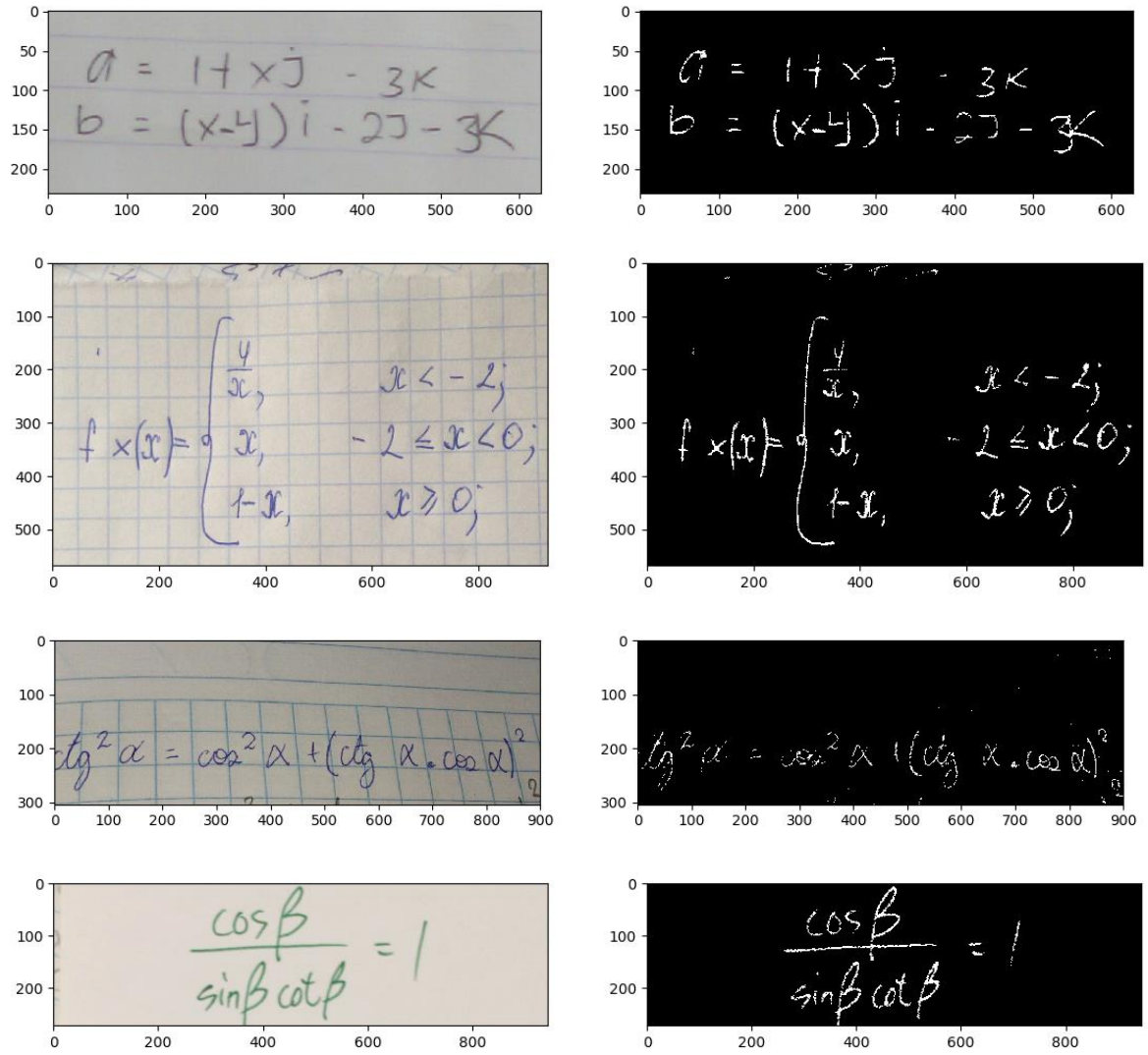


Slika 8.2: Rezultati s pretincima veličine 32

Rezultati klasifikacije su dosta dobri, no na nekim slikama se javlja popriličan broj lažnih pozitiva, odnosno piksela pozadine koji se pogrešno klasificiraju kao pikseli teksta. To su najčešće pikseli nijansi plave boje. Nijanse crne, plave i sive boje se na ulaznim slikama pojavljuju i rukom pisanim znakovima i u linijama na pozadini pa se vrlo lako mogu pogrešno klasificirati tako da nas pojava lažnih pozitiva ne iznenađuje. Pokušali smo ih se riješiti povećanjem praga klasifikacije. To smo donekle uspjeli ali smo time ujedno dosta smanjili količinu prepoznatog teksta kao što se može vidjeti na slici 8.1. Nakon što smo minimalno povećali prag klasifikacije dobili smo dosta lošije rezultate za većinu slika. Pogledajmo što se događa kada veličinu pretinaca smanjimo na 16.



Slika 8.3: Primjer lažnih pozitiva uz pretince veličine 16



Slika 8.4: Rezultati s pretincima veličine 16

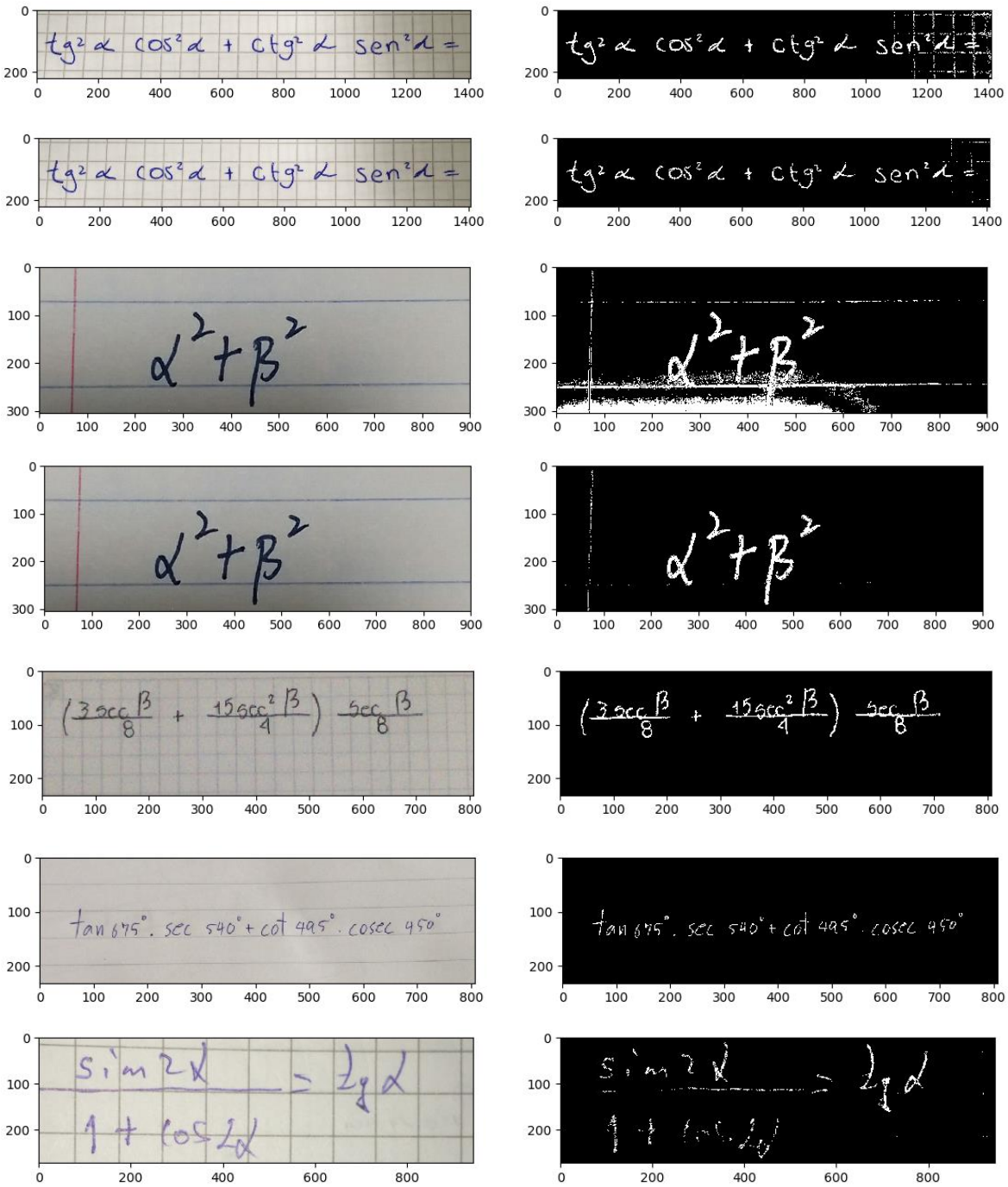
Smanjujući veličinu pretinaca uz minimalan prag na slici 8.3 uočavamo mnogo više lažnih pozitiva. Međutim tih se pozitiva ponovno možemo riješiti povećanjem praga, ali ovoga puta puno uspješnije. Za razliku od prethodnog slučaja kada smo prag mogli tek malo povećavati da ne bi ugrozili rezultate klasifikacije sada prag možemo dosta varirati a da rezultati i dalje budu dobri. Povećavajući prag uklanjamo većinu lažnih pozitiva a da se pritom količina prepoznatog teksta tek neznatno smanji. To nam omogućava da prag klasifikacije bude veći i da klasifikacija bude neovisnija o pragu, odnosno da daje podjednako dobre rezultate za veći raspon praga.

Na slici 8.4 su primjeri slika za koje je klasifikacija s pretincima veličine 16 dala dosta bolje rezultate nego klasifikacija s pretincima veličine 32. Veća je količina prepoznatog teksta, manje je lažnih pozitiva i iako se slične nijanse boja pojavljuju i u tekstu i u pozadini razlikovanje je uspješnije. Još jedna bitna činjenica koja se može uočiti ja da dijelovi slika koji su u sjeni, odnosno koji su tamniji također imaju mnogo manje pogrešno klasificiranih piksela. Osvjetljenje ima veliki utjecaj na boju jer boja može uvelike varirati u ovisnosti o raznim uvjetima osvjetljenja. Uočava se i bolja generalizacija, primjerice za najdonji primjer na slici 8.4 gdje su prepoznati zeleni znakovi ikako su zeleni pikseli rijetko viđeni u skupu označenih slika, tek na jednoj slici.

U daljnjim testiranjima kada smo smanjili veličinu pretinaca na 8 broj lažnih pozitiva je postao još veći na manjim pragovima, a klasifikacija je postala još neovisnija o pragu. Međutim zaključeno je kako naglo povećanje lažnih pozitiva ozbiljno ugrožava klasifikaciju i da se ne može tako lako ukloniti povećanjem praga te da ovo i daljnja smanjivanja veličine pretinaca nemaju pretjeranog smisla. Nadalje smanjujući veličinu pretinaca povećavamo broj pretinaca i memorijsko zauzeće histograma, te u konačnici i vrijeme izvođenja.

Zanimljivo je pogledati kakvi su rezultati s pretincima veličine 128, odnosno kada imamo samo 8 pretinaca u svakom od histograma. Uz minimalan prag rezultati su iznenađujuće dobri. Ali uz bilo kakvo povećanje praga svi pikseli se klasificiraju kao pozadina što nam onemogućava da smanjimo broj lažnih pozitiva.

Svi do sada navedeni rezultati su provedeni na skupu slika za učenje. Sljedeći rezultati dobiveni su na skupu za testiranje i to uz veličinu pretinaca 16 jer je ta veličina postigla najbolje rezultate u prethodnim razmatranjima.



Slika 8.5: Rezultati na skupu za testiranje

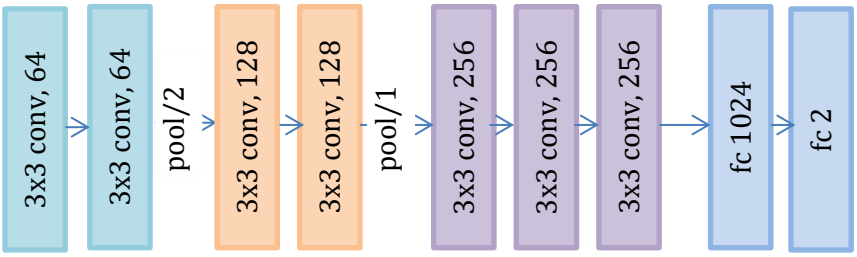
Rezultati su nešto lošiji nego na skupu za učenje. Količina prepoznatog teksta je manja, ali količina lažnih pozitiva nije značajno veća. S povećanjem praga nismo se u potpunosti uspjeli riješiti lažnih pozitiva. Zanimljivo kako su crvene linije u pozadini uvijek klasificirane kao dio teksta iako je u označenim slikama bilo sličnih primjera. Analizirajući crvene piksele i vizualizirajući

zastupljenost crvene, zelene i plave komponente ustanovili smo da crveni pikseli nisu u potpunosti crveni, tj. crvena komponenta nije maksimalna a plava i zelena komponenta nisu jednake 0. Sve tri komponente su zastupljene dok je crvena tek neznatno više zastupljena. Zbog toga se može dogoditi da pri većim veličinama pretinaca ti pikseli ne budu dobro odijeljeni i da završe u pretincima kod kojih su sve tri komponente u jednakom rasponu. Time je veća zastupljenost crvene tek neznatno naglašena ili u potpunosti zanemarena te se pikseli pogrešno klasificiraju.

8.2. Rezultati s konvolucijskim neuronskim mrežama

Za ovakve klasifikacijske probleme najčešće se primjenjuju modeli temeljeni na dubokim neuronskim mrežama. Stoga smo odlučili usporediti rezultate statističke metode s jednim konvolucijskim modelom. Konvolucijski model je bio prikladan iz razloga što ulazni skup slika sadrži slike različitih dimenzija. Kao arhitektura je odabrana prethodno trenirana mreža VGG16. Kod s cjelokupnom implementacijom mreže je preuzet iz projekta studenta Ivana Fabijanića [2] i prilagođen našem problemu.

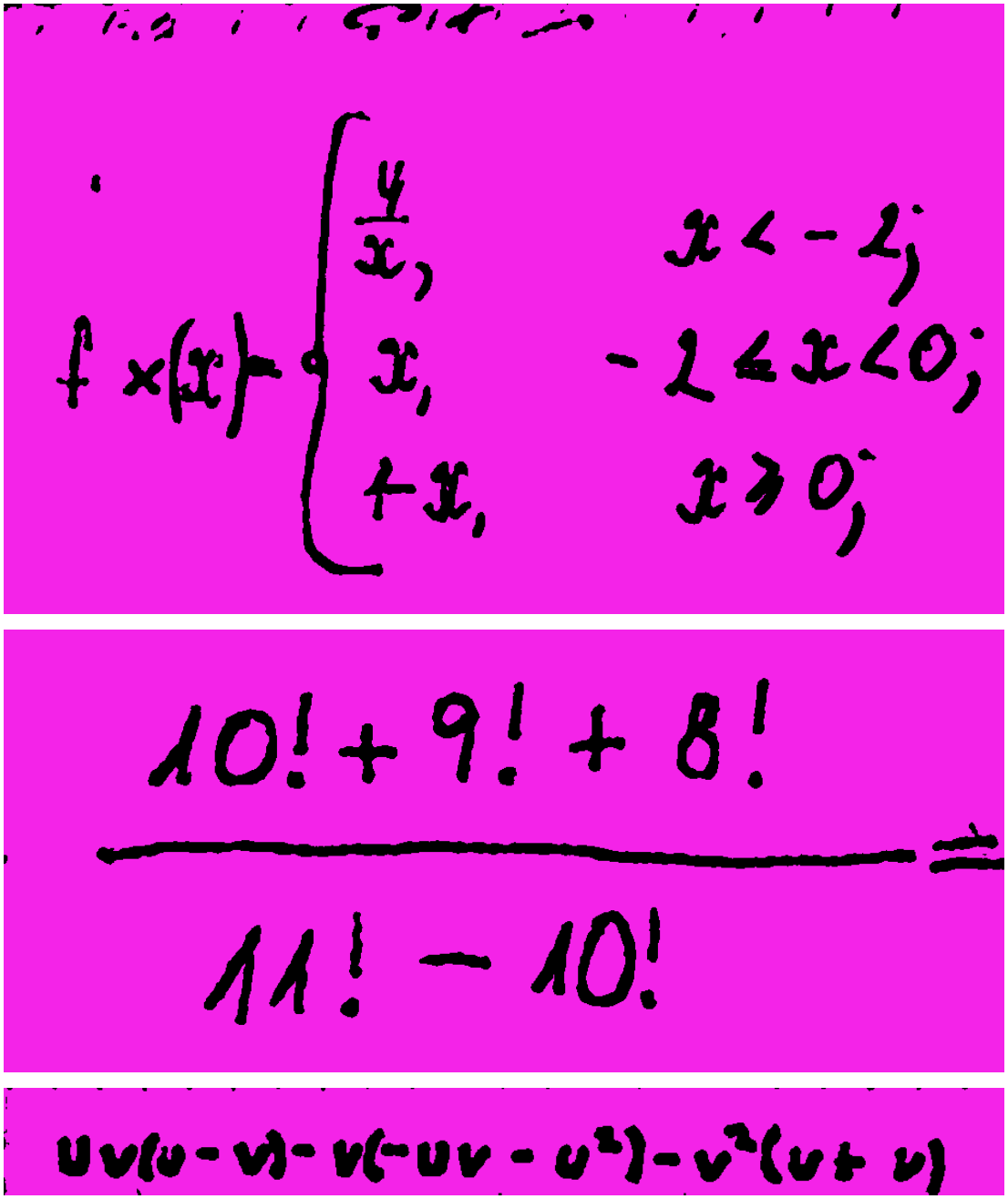
Cilj je bio maksimalno pojednostaviti mrežu jer je naš klasifikacijski zadatak mnogo jednostavniji od zadatka za koji je mreža građena. Iz mreže su već bili uklonjeni potpuno povezani slojevi i umjesto njih uvedena dva konvolucijska sloja koja glume potpuno povezane slojeve. Prvi konvolucijski sloj ima 1024 neurona a drugi 2 neuron što odgovara broju klasifikacijskih razreda. Time je značajno smanjen broj parametara mreže i vjerojatnost da će doći do prenaučenosti do koje dolazi ako mnogo parametara učimo na malo podataka. Od 13 konvolucijskih slojeva koji su inicijalno postojali u mreži ostavljeno je samo sedam konvolucijska sloja i 2 sloja sažimanja. Pojednostavljena arhitektura je prikazana na slici 8.6.



Slika 8.6: Pojednostavljena arhitektura mreže

Mreža je trenirana sa skupom od 10 rijetko označenih slika kroz 25 epoha. Algoritam učenja mreže bio je Adam sa stopom učenja 0,0001 i eksponencijalnim padom u daljnjim epohama. Rezultati su prikazani na slici 8.6.

Vidimo da je model dobro naučio razlikovati tekst i pozadinu. Najveći problem mu predstavljaju područja uz same rubove slike koja imaju najlošiju klasifikaciju. To je iz razloga što je korištena konvolucija s nadopunom nula na rubovima.



Slika 8.7: Rezultati s konvolucijskim modelom

9. Zaključak

U problemima u kojima je potrebno klasificirati piksele u svega jedan ili dva razreda statističke metode poručuju jako dobre rezultate. Prednost statističkih metoda je da imaju malen broj parametara i ne zahtijevaju dugotrajno učenje, a nedostatak da imaju smanjenu mogućnost generalizacije i zaključivanja. Praktične su u problemima u kojima je potrebna brza klasifikacija i koji imaju veću toleranciju na pogreške. Također vrijeme izvođenja i memorijsko zauzeće podataka je mnogo manje nego kod klasifikacije s neuronskim mrežama. Međutim konvolucijske neuronske mreže imaju mnogo veću ekspresivnost od klasičnih metoda i bolje se prilagođavaju uvjetima u okolini. Neuronske mreže su sposobne donositi bolje zaključke i predviđanja jer bolje akumuliraju znanje iz podataka za učenje.

U ovom radu prikazani su rezultati statističke klasifikacije temeljene na histogramima boja i pokazano je kako veličina histograma i način grupiranja piksela imaju utjecaja na klasifikaciju. Analizirana je pojava lažnih pozitiva i njihova ovisnost o pragu klasifikacije.

Kao drugi pristup rješavanju problema semantičke segmentacije odabran je konvolucijski model. Pojednostavljena je postojeća arhitektura istrenirane mreže VGG i pokazano da se uz smanjen broj parametara i malo treniranja može prilagoditi za razne klasifikacijske probleme.

Za problem semantičke segmentacije rukom pisanog teksta u budućnosti bi se mogle nastaviti razvijati statističke metode zbog svoje jednostavnosti, ali i neuronski modeli zbog svoje robusnosti.

LITERATURA

[1] Richard Szeliski. Computer Vision: Algoritams and Applications, 2010.
URL http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf
[2] Ivan Fabijanić. Interaktivna semantička segmentacija sportskih susreta. Projekt, 2017.
[3] Tomislav Babić. Primjena histograma boje u računalnom vidu. Seminar, 2010.
[4] S. Ribarić, B. Dalbelo Bašić. Modeliranje neizvjesnosti, službeni materijali s predmeta Inteligentni sustavi, 2002.
URL http://www.zemris.fer.hr/predmeti/is/nastava/Modeliranje_neizvjesnosti.pdf
[5] Josip Krapac: Duboke unaprijedne, prezentacija sa predmeta Duboko učenje, 2016. URL http://www.zemris.fer.hr/~ssegvic/du/du1feedforward.pdf
[6] Convolutional Neural Networks for Visual Recognition.
URL http://cs231n.github.io/convolutional-networks/
[7] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
[8] Michael J. Jones, James M. Rahg. Statistical Color Models with Application to Skin Detection, 1998.
[9] S. Šegvić, K. Brkić, Z. Kalafatić, V. Stanisavljević, M. Ševrović, D. Budimir, I. Dadić. A computer vision assisted geoinformation inventory for traffic infrastructure, 2010.
[10] Gaurav Manek. 3D Histogram.
URL http://gauravmanek.com/blog/2011/01/31/3d-histogram.html
[11] Image Classification.
URL http://book.paddlepaddle.org/03.image_classification/

Interaktivna semantička segmentacija rukom pisanih znakova

Sažetak

U ovom radu su obrađena dva pristupa semantičkoj segmentaciji rukom pisanog teksta. Prvi pristup je statistički temeljen na histogramima boja i uvjetnim vjerojatnostima. Pritom je analiziran utjecaj koji veličina histograma i prag imaju na klasifikaciju. Drugi pristup je temeljen na konvolucijskim neuronskim mrežama. Pronađena je pojednostavljena arhitektura konvolucijske mreže VGG koja uz malo treniranja daje dobre rezultate. Prikazan je način i intenzitet označavanja ulaznih slika te su oba pristupa testirana na ulaznom skupu slika pri čemu su dali zadovoljavajuće rezultate.

Ključne riječi: semantička segmentacija teksta, histogram boja, konvolucijske neuronske mreže, VGG, OpenCV.

Interactive semantic segmentation of handwritten text

Abstract

This paper analyzes two approaches to the semantic segmentation of a handwritten text. The first approach is statistically based on color histograms and conditional probabilities. Thereby the effect of histogram size and threshold on the classification is analyzed. The second approach is based on convolutional neural networks. We simplified the architecture of the VGG convolution network that gives good results with little training. The way and intensity of labeling input images is shown and both approaches are tested and provide satisfactory results.

Keywords: semantic segmentation, handwritten text, color histogram, convolution neural networks, VGG, OpenCV.