

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

ZAVRŠNI RAD br. 6391

**Detekcija i klasifikacija tekstovnih
elemenata na slici koristeći duboke
neuronske mreže**

Lukas Šestić

Zagreb, travanj 2019.

*Umjesto ove stranice umetnite izvornik Vašeg rada.
Da bi ste uklonili ovu stranicu obrišite naredbu \izvornik.*

Zahvaljujem se prof. dr. sc. Domagoju Jakoboviću na pruženoj pomoći i sredstvima danim u svrhu uspješne izrade završnog rada.

Također se zahvaljujem svojim roditeljima na prilici za studiranje i stalnoj podršci.

SADRŽAJ

1. Uvod	1
1.1. Uvod	1
1.2. Računalni vid	1
1.2.1. Pregled	1
1.2.2. Konvolucijske neuronske mreže	2
2. Generiranje seta podataka	4
2.1. Značaj podataka u dubokom učenju	4
2.2. Generiranje slika	5
2.2.1. Generalizacija postupka	5
2.2.2. Prikupljanje fontova	6
2.2.3. Generiranje simbola	6
2.2.4. Transformacije	6
2.2.5. Kreiranje cjelovitih slika	6
Literatura	7

1. Uvod

1.1. Uvod

Računalna moć uređaja koje gotovo neprestano nosimo sa sobom kao što su pametni telefoni i prijenosna računala je unazad deset godina eksponencijalno narasla. Naravno, sa modernim alatima dolaze i moderni problemi.

Jedan najraširenijih alata, koji se sve više i više koristi za rješavanje problema koji su do nedavno bili nemogući, ili izrazito algoritamski komplicirani za riješiti je *umjetna inteligencija*. Umjetna inteligencija podrazumjeva skup načina i metoda koje računalu opisuju početno i konačno stanje do kojeg mora doći sam.

Ovaj rad, bavit će se najkorištenijom metodom umjetne inteligencije, *dubokim učenjem*, i njegovim podskupom *računalnim vidom*. Kroz rad i programsku implementaciju, prihvatiti ću se problema detekcije napisanog teksta na slici i daljnom obradom istog.

Detaljno ću kroz poglavlja obraditi postupke koje sam primjenio za generiranje raznolikih slika koje imitiraju rukopis i proces potreban da računalo nauči prepoznavati isti na slici.

Na kraju, izlučeni tekst sa slike, biti će moguće obraditi na željeni način. Način koji ću ja predstaviti biti će primjena jednostavne matematike, slično onome što pruža *Photomath, Inc.*. Na primjer, za sliku na kojoj je napisan tekst " $2 + 2$ ", izlaz će biti slika sa kvadratima oko prepoznatih simbola, i rješenje obrađenog teksta, u ovom slučaju " 4 ".

1.2. Računalni vid

1.2.1. Pregled

Na najvišoj razini, *računalni vid* su metode koje računalima daju mogućnost razumjevanja slike na visokoj razini, najčešće s ciljem automatiziranja ljudskih

poslova. Osnovni zadatak je raspoznavanje veze između obrazaca na slici i rješenja na problem koji želi riješiti. Svi procesi koji koriste strojno učenje, u konačnici se svode na detekciju i klasifikaciju elemenata na slici. Metode računalnog vida temelje se na geometriji, statistici, fizici i teoriji učenja.

Danas, se velika količina problema rješava uz pomoć računalnog vida, često da ljudi za to nisu ni svjesni:

- Prepoznavanje znakova (Slika 1.1)
- Prepoznavanje lica
- Kompresija i restauracija slike
- Prepoznavanje elemenata na slici
- Analiza medicinskih snimki u svrhu detaljnije analize
- Itd.



Slika 1.1: Maskiranje elemenata na slici prometa

1.2.2. Konvolucijske neuronske mreže

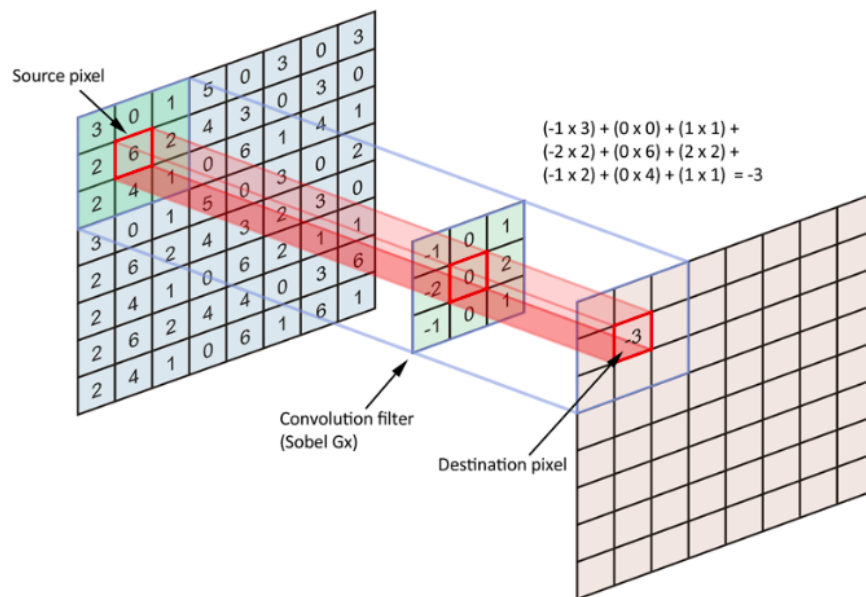
Konvolucijske neuronske mreže danas se koriste kao najefektivniji način postizanja računalnog vida. Glavna prednost nad potpuno povezanim neuronskim mrežama je manji broj *težina* za treniranje što znatno ubrzava treniranje. Ipak, ono što je možda najvažnije za napomenuti je to što pozicija traženog elementa na slici *konvolucijskoj neuronskoj mreži* ne igra ulogu.

Svaki sloj *duboke konvolucijske neuronske mreže* funkcionira kao filter koji se kreće po slici, pamteći što ga je najviše aktiviralo. Najčešće se koristi filter veličine 3×3 . (Slika 1.2)

Dalje, uz konvolucijski sloj, nerijetko se postavlja *max pooling sloj*. Na apstraktnoj razini, princip rada max pooling sloja je sljedeći: Ako uzmemo veličinu pooling filtra kao onu koja se najčešće koristi, to jest 2×2 , on izlaz iz prethodnog sloja raspodjeli na kvadrate iste veličine. Zatim, filter se postavi između 4 kvadrata i sebi za vrijednost stavi najveću iz svakog u pripadajuće polje.

Prirodno je pitati se zašto se to koristi i zašto to radi.

Pooling filter jednostavno smanjuje "rezoluciju" prethodnog sloja, ne mijenjajući važne čimbenike potrebne za daljnji rad mreže. Na primjer, vertikalna linija, krug, ili elipsa, ostaje ono što je, jedino manje razlučivo. Bitno je napomenuti da smanjivanjem rezolucije dobivamo puno manje parametara za treniranje. Stavimo to u brojeve. Slike unutar *mnist* seta podataka su veličine 28×28 . To znači da bi se treniralo 28×28 parametara. Primjenom *Max pooling sloja* veličine 2×2 , treniralo bi se $\frac{1}{4} \times (28 \times 28)$ parametara.



Slika 1.2: Klizeći konvolucijski filter

Spomenute prednosti, referenciraju se na glavnu značajku *konvolucijskih mreža*. Cilj je ići dublje, ne šire. Za sliku veličine 100×100 , potpuno povezanoj neuronskoj mreži u prvom sloju treba 10 000 čvorova, svaki sa svojim parametrom za treniranje, dok konvolucijskoj to ne treba.

Svaki sljedeći sloj ima drugu ulogu. Prvi najčešće ima ulogu raspoznavanja najosnovnijih elemenata slike kao što su različiti rubovi, dok sve dublji koriste podatke od prošlih i osnovne elemente grupiraju u apstraktne strukture koji predstavljaju značajnije elemente slike (O'Shea i Nash (2015)).

2. Generiranje seta podataka

2.1. Značaj podataka u dubokom učenju

Prvi i najdulji praktični korak treninga predstavlja priprema podataka. Sve ovisi o zadatku koji mreža mora riješiti, ali, generalno je pravilo da je više podataka bolje. Konačna kvaliteta rješenja osim o arhitekturi mreže koju dizajniramo, ovisi o kvaliteti podataka kojom ju usmjeravamo. Priprema podataka vrši se u 3 glavna koraka (Gonfalonieri (2019)):

1. Prikupljanje
2. Klasifikacija
3. Označavanje

Prikupljanje podataka

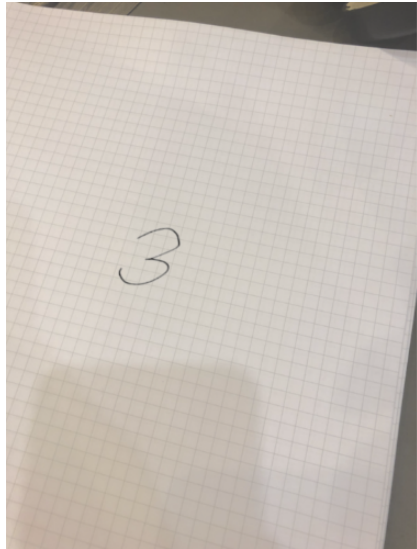
Prikupljanje podataka mora biti sustavan i smislen proces jer može otežati i olakšati daljnje korake. Najpreporučeniji način za prikupljanje je dugoročno i postepeno spremanje podataka jer rezultira velikim brojem objektivnih i kvalitativnih podataka. Ja sam se ipak odlučio na metodu računalnog generiranja vlastitog seta podataka. Razlog tome je raznolikost elemenata koje mreža mora moći detektirati i fleksibilnost koju dobivam jednomo kada ustanovim sve potrebe.

Klasifikacija i označavanje podataka

Generirani podaci na određeni način moraju biti prikazani mreži. Iako u mrežu slika ulazi kao vektor dimenzija (**visina x širina x kanali**) mreži su potrebni i podaci za uspoređivanje rezultata i računanje uspješnosti. U ovom radu koristio sam .csv datoteku za dohvaćanje i opisnik slika. Postupak automatskog generiranja slika uvelike je olakšao klasifikaciju i označavanje jer je cijeli postupak ostvaren kao "cjevovod". Pri izlasku, slika bi bila prikazana kao na slici 2.1.

Datoteka bi upisano imala ime slike, simbol na slici, širinu, visinu i točan položaj elementa na slici. Prednost ovog pristupa je i u tom što slika nije zadana absolutnom putanjom, što znači, da sam slike mogao kreirati na vlastitom računalu, prenjeti ih na udaljeni server za treniranje i bez komplikacija koristiti iste.

Veličina opisnika je također bila zanemariva. Nakon raspodjele 80:20 za trening i validaciju na 15 000 slika, veličine su bile 440kB i 110kB dok je direktorij sa slikama bio veličine 6,7GB.



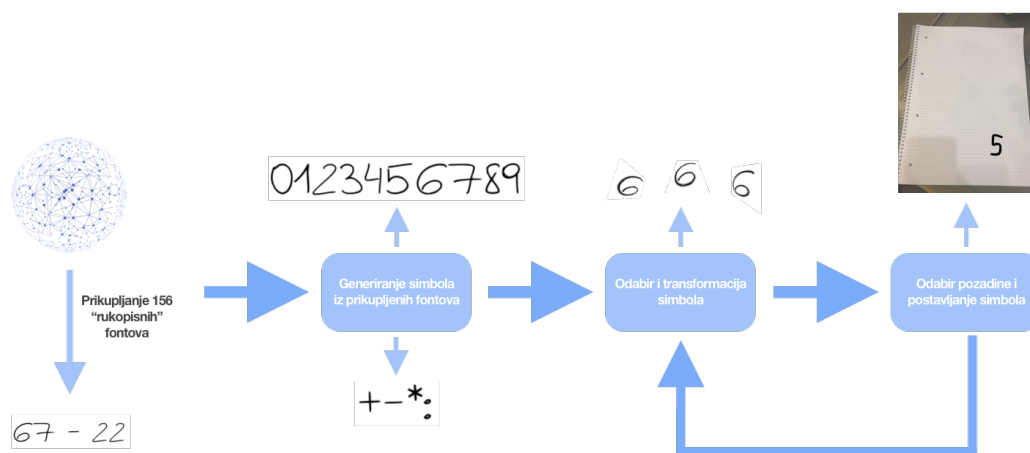
filename	class	imwidth	imheight	xmin	xmax	ymin	ymax
eqjodd.png	:	540	720	175	189	263	361
arnqxq.png	:	540	720	254	277	510	624
tvxdvf.png	*	540	720	394	455	579	669
fapsod.png	:	540	720	32	67	104	243
qardct.png	-	540	720	257	292	246	375
thzsfh.png	-	540	720	211	253	149	212
vdtpni.png	3	540	720	34	157	311	436
uyekux.png	3	540	720	116	169	542	636
iyjrea.png	4	540	720	142	300	131	206
uqqaqt.png	0	540	720	250	340	444	580
khhwvh.png	0	540	720	280	384	121	332
yqumug.png	5	540	720	349	428	454	586

Slika 2.1: Slika i pripadajuća referenca u .csv datoteci

2.2. Generiranje slika

2.2.1. Generalizacija postupka

Za relativan uspjeh treniranja mreže za detekciju i klasifikaciju 14 tekstovnih elemenata (0-9, +, -, *, :) potrebno je minimalno 10 000 slika. Ne samo zbog broja elemenata već i zbog složenosti i raznolikosti između njih. Postupak koji sam razvio primjenjuje sve taktike (Chollet (2017)) potrebne za stvaranje raznovrsnog i kvalitetnog seta podataka. Zbog transformacija, opisanih u daljnjim djelovima poglavlja, gotovo je nemoguće da iako se isti font stavlja na pozadinu, nastane isti oblik. Na slici 2.2 prikazana je topologija cjevovoda koja kreira slike.



Slika 2.2: Prikaz visoke razine cjevovoda za generiranje slika

2.2.2. Prikupljanje fontova

2.2.3. Generiranje simbola

2.2.4. Transformacije

Skaliranje

Rotacija

Afine transformacije

2.2.5. Kreiranje cjelovitih slika

LITERATURA

F. Chollet. *Deep Learning with Python*. Manning Publications Company, 2017. ISBN 9781617294433. URL <https://books.google.hr/books?id=Yo3CAQAACAAJ>.

Alexandre Gonfalonieri. How to build a data set for your machine learning project, 2019. URL <https://towardsdatascience.com/how-to-build-a-data-set-for-your-machine-learning-project-5b3b871881ac>.

Keiron O'Shea i Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.