

Očitavanje rukom pisanih identifikacijskih brojeva temeljeno na dubokim modelima

Tomislav Božurić

Fakultet elektrotehnike i računarstva

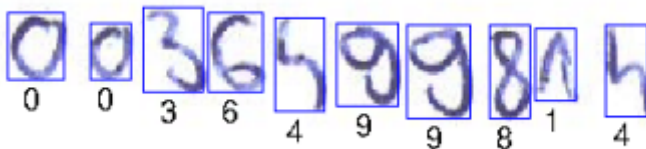
Diplomski rad

Mentor: doc.dr.sc. Marko Čupić

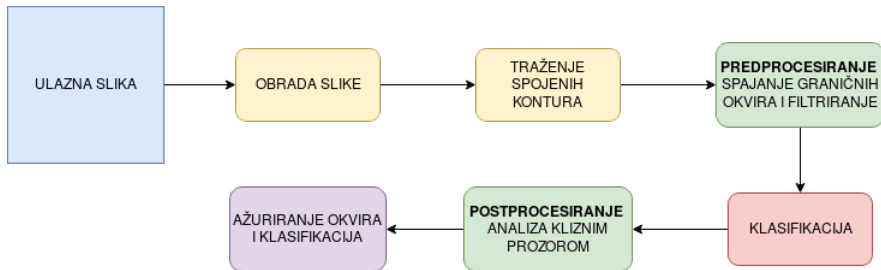
10. srpnja 2019.

- 1 Uvod
- 2 Pristup problemu
 - Obrada slike
 - Pronalazak spojenih kontura
- 3 Predprocesiranje
- 4 Klasifikacija
 - Arhitektura neuronske mreže
 - Rezultati
- 5 Postprocesiranje
- 6 Izgrađeni sustav
- 7 Rezultati
- 8 Zaključak

- **Očitavanje rukom pisanih identifikatora** → problem lokalizacije i klasifikacije znamenaka koje sadrži pojedini identifikator



Slika: Pregled problema

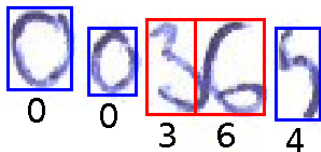


Slika: Tok obrade ulazne slike kroz izgrađeni sustav

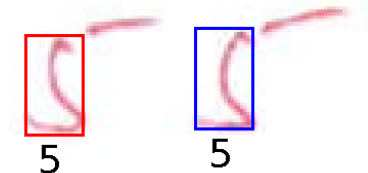
- Pretpostavka: ulazna slika sastoji se od dvije klase slikovnih elemenata
- Koraci:
 - **pretvorba slike u sive nijanse** temeljeno na ljudskoj percepciji boja
 - **binarizacija** → Otsuova metoda
 - **dilatacija** odnosno operacija podebljavanja kontura
 - **interpolacija** → metoda najbližeg susjeda
 - **translacija** → metoda centra mase

Pronalazak spojenih kontura

- Lokalizacije pojedine znamenke vrši se pronalaskom spojenih kontura → *Connected Component Labeling*
- Problem preklapajućih i spojenih kontura znamenaka (**analiza kliznim prozorom**) i problem podjele konture na više dijelova (**dilatacija**)



(a) Primjer spojenih kontura



(b) Primjer pogrešne lokalizacije

Cilj

- **eliminiranje šumova**
- **spajanje okvira** koji vjerojatno čine jednu znamenku

Opis

- u potpunosti temeljen na *a priori* znanju o tome kako čovjek piše identifikatore
- na temelju **udaljenosti između okvira, omjeru širina i visina** te na temelju odnosa **visine i širine potencijalno spojenog okvira i prosječne visine i širine svih okvira** vrši spajanja
- eliminira šumove na temelju **prosječne visine svih okvira**

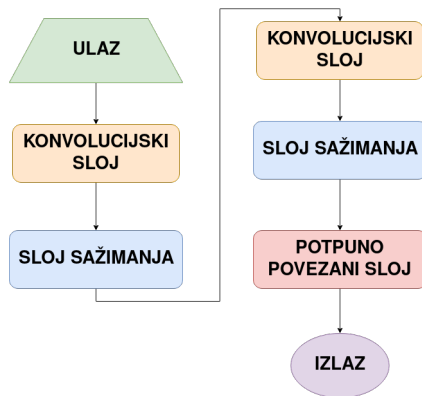
- implementiran koristeći radni okvir za duboko učenje *DeepLearning4j*
- baza rukom pisanih identifikatora je izgrađena iz uzoraka od desetak različitih autora
- skup za učenje se sastojao od 2371 slike, a skup za ispitivanje od 402 slike
- učenje je provedeno kroz 30 epoha, u mini-grupama po 54 uzorka
- težine su inicijalizirane koristeći **Xavierovu** inicijalizaciju:

$$W_{i,j} \sim \mathcal{N}(0, \frac{2}{n_{in} + n_{out}}) \quad (1)$$

- učenje stohastičkim gradijentnim spustom (SGD) s Nesterovim momentom
- L2 regularizacija

Arhitektura dubokog konvolucijskog modela

- Ukupno 6 slojeva i 1 385 130 parametara



Slika: Arhitektura izgrađenog konvolucijskog modela

Matrica zabune i evaluacije (makro uprosječena)

	0	1	2	3	4	5	6	7	8	9
0=0	39	0	0	0	0	0	0	0	1	0
1=1	0	41	0	0	0	0	0	0	0	0
2=2	0	0	40	0	0	0	0	0	0	0
3=3	0	0	0	37	0	0	0	0	0	3
4=4	0	0	0	0	40	0	0	0	0	0
5=5	0	0	0	0	0	36	4	0	0	0
6=6	0	0	0	0	2	0	38	0	0	0
7=7	0	0	0	0	0	0	0	41	0	0
8=8	0	0	0	0	0	0	0	0	40	0
9=9	0	0	0	0	3	0	0	0	0	37

Točnost	0.9677
Preciznost	0.9694
Odziv	0.9675
F1	0.9676

Cilj

- predviđanje i razrješavanje pogrešno lokaliziranih/klasificiranih znamenaka

Uvjeti da bi se okvir razmatrao kao pogrešno lokaliziran/klasificiran:

- širina okvira mora biti barem 1.5 puta veća od prosječne širine svih okvira
- sigurnost (engl. *confidence*) klasifikatora u odluku mora biti manja od 50% i zbunjenost (engl. *perplexity*) mora biti veća od eksperimentalno dobivene vrijednosti

Zbunjenost modela

Zbunjenost diskretne vjerojatnosne distribucije p definiramo preko mjere nesigurnosti ili neodređenosti modela, odnosno preko entropije kao

$$2^{H(p)} = 2^{-\sum_x p(x) \log_2 p(x)}. \quad (2)$$

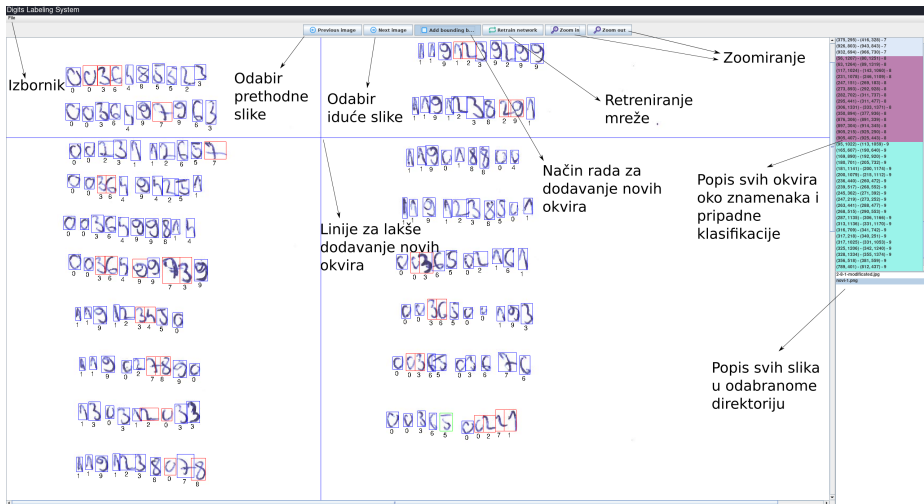
Opis algoritma razrješavanja pogrešno lokaliziranih/klasificiranih okvira

- **metodom klznog prozora** vrši se analiza okvira
- veličine klznog prozora su **dinamičke**
- različitim veličinama klznog prozora pokušava se podijeliti okvir kako bi postigli minimalnu zbunjenost modela

Izraz za traženje podjele okvira s najmanjom zbunjenosti

$$Utjecaj = \sum_{i=0}^{broj\ okvira - 1} zbunjenost_i^2 \cdot e^{-i}. \quad (3)$$

Izgrađeni sustav



Slika: Prikaz glavnog ekrana aplikacije

- Na ulaznim uzorcima gdje su znamenke razmaknute sustav s visokim postotkom uspješnosti vrši lokalizaciju i klasifikaciju
- Izradom postprocesora pokazuje se značajan napredak na ulaznim slikama s gusto napisanim brojevima
- Na uzorku od 230 znamenaka (23 studentska identifikatora) 8 znamenaka je sustav krivo segmentirao (3.48%), dok je 6 znamenaka djelomično krivo segmentirao (2.61%), no dovoljno dobro za uspješnu klasifikaciju

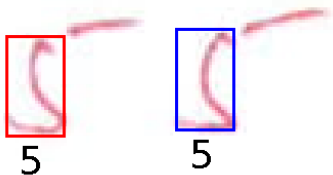
Primjer pogrešaka prilikom segmentacije i klasifikacije



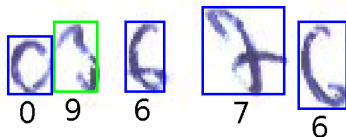
(a) Primjer pogrešne separacije okvira na više manjih



(b) Primjer okvira koji sadrži dvije znamenke



(c) Primjer pogrešne segmentacije zbog načina pisanja



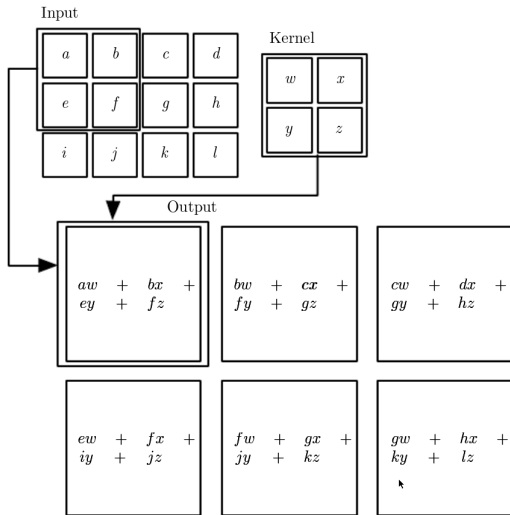
(d) Primjer pogrešne klasifikacije

Slika: Primjer pogrešaka prilikom segmentacije i klasifikacije

- Mješavinom zaključaka stečenim **učenjem klasifikatora** i **matematičkim pretpostavkama o problemu** sustav se vrlo dobro nosio s problemima lokalizacije i klasifikacije
- Daljnja unapređenja treba usmjeriti ka sustavu za **postprocesiranje (metoda kliznog prozora)**
- Povećanjem skupa uzoraka za učenje klasifikator postaje još robusniji i kvalitetnije se nosi s problemima
- Eksperimentiranjem nad matematički modeliranim znanjem možemo fino ugađati sustav i dodavanjem novog znanja poboljšati rad sustava

Hvala na pažnji!

Konvolucija



Slika: Konvolucija 2D ulaza i jezgre

- Koristeći **lokalnu interakciju** potrebno je naučiti manje parametara s istom količinom označenih podataka
- **Dijeljenje parametara** omogućuje da svi izlazi dijele jedan te isti skup parametara → manji model koji je teže prenaučiti
- **Sažimanjem** dobivamo reprezentaciju koja je približno invarijantna na male translacije u ulazu (fokusiramo se na postojanje objekta, a ne na točnu lokaciju)

- **Točnost** definiramo kao omjer ispravno klasificiranih primjera i ukupnog broja primjera: $A_{cc} = \frac{TP+TN}{N} = \frac{TP+TN}{TP+TN+FP+FN}$
- **Preciznost** je definirana izrazom: $P = \frac{TP}{TP+FP}$
- **Odziv** je definiran omjerom broja primjera koje je klasifikator označio kao pozitivne i ukupnog broja pozitivnih primjera: $R = \frac{TP}{TP+FN}$
- **F1** mjeru definiramo kao harmonijsku sredinu između preciznosti i odziva: $F1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2PR}{P+R}$

Makro i mikro mjera

		Stvarno			
		C_1	C_2	C_3	
Predviđeno	C_1	1	1	0	$N=13$
	C_2	2	2	3	
	C_3	0	0	4	

$$\begin{array}{lcl} TP_1=1, FP_1=1, FN_1=2 & \Rightarrow & P_1=0.5, R_1=0.33 \Rightarrow F_{1,1}=0.4 \\ TP_2=2, FP_2=5, FN_2=1 & \Rightarrow & P_2=0.29, R_2=0.66 \Rightarrow F_{1,2}=0.4 \\ TP_3=4, FP_3=0, FN_3=3 & \Rightarrow & P_3=1, R_3=0.57 \Rightarrow F_{1,3}=0.73 \end{array}$$

$$\Downarrow \quad \Downarrow \quad \Downarrow$$
$$TP=7 \quad FP=6 \quad FN=6$$

$$\Downarrow$$
$$P=R=0.54$$

$$F_1^{micro} = \frac{2 \cdot 0.54 \cdot 0.54}{0.54 + 0.54} = 0.54$$

$$\Downarrow$$
$$F_1^{macro} = \frac{1}{3}(0.4 + 0.4 + 0.73)$$
$$= 0.51$$

Makro vs. mikro mjera

- Makro- F_1 sve klase tretira jednako
 - zbog toga primjeri iz malih klasa imaju veći utjecaj na mjeru nego što bi imali kod mjere mikro- F_1
 - razlika je vidljiva kod neuravnoteženih skupova

Npr.

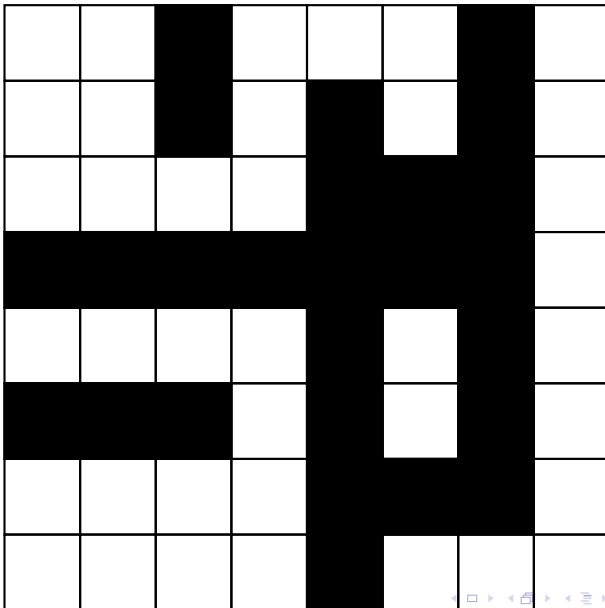
	C_1	C_2	C_3
C_1	100	10	1
C_2	2	2	3
C_3	8	5	400

$$F_1^{micro} = 0.95$$

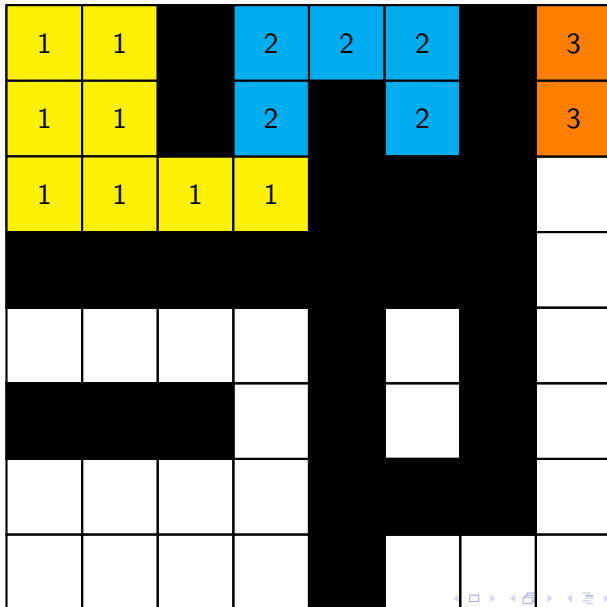
$$F_1^{macro} = 0.69$$

- makro- F_1 je tipično manji od mikro- F_1
 - zato jer je na slabo zastupljenim klasama klasifikacija tipično lošija

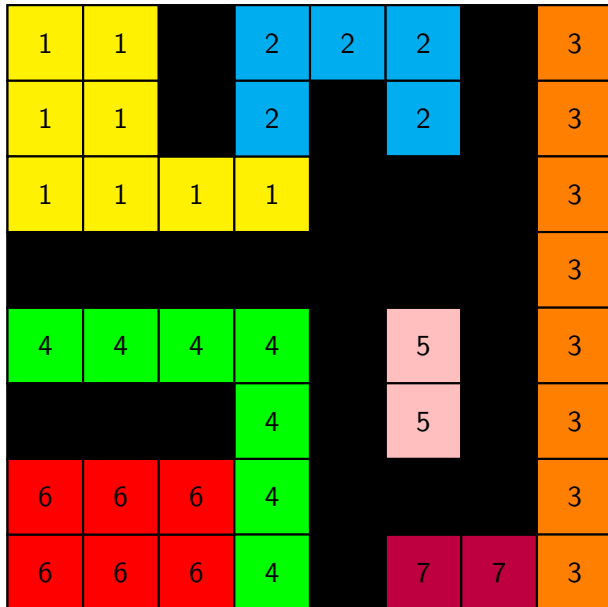
Pronalazak spojenih kontura



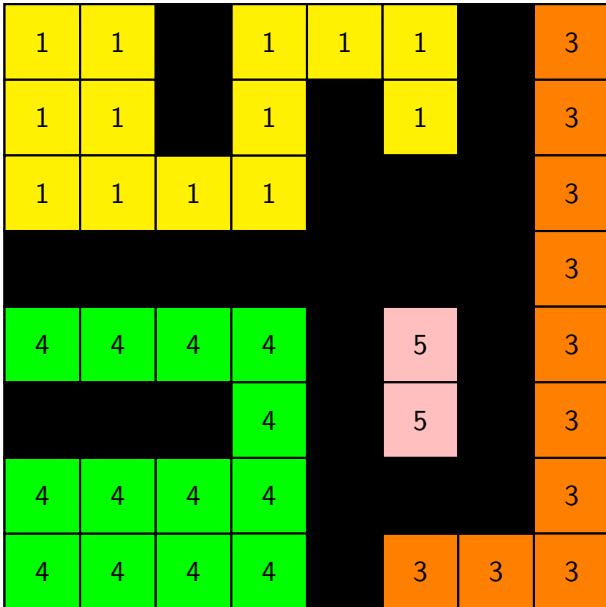
Pronalazak spojenih kontura



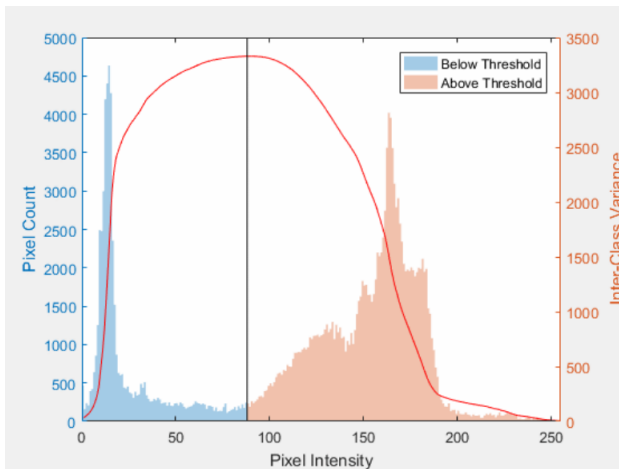
Pronalazak spojenih kontura



Pronalazak spojenih kontura



Otsuova metoda



Slika: Prikaz određivanja praga binarizacije koristeći maksimizaciju varijance između razreda

- Ubrzavanje učenja kod šumovite procjene gradijenata i kod malih ali konzistentnih gradijenata
- Postupak akumulira eksponencijalno umanjujući prosjek prethodno izračunatih gradijenata i nastavlja u tome smjeru

$$\mathbf{v} \leftarrow \alpha \cdot \mathbf{v} - \epsilon \cdot \nabla_{\theta} \left(\frac{1}{m} \sum_{i=1}^m L(f(\mathbf{x}^{(i)}; \theta), y^i) \right) \quad (4)$$

$$\theta \leftarrow \theta + \mathbf{v} \quad (5)$$