

**JEDNO REŠENJE SISTEMA ZA OPTIČKO PREPOZNAVANJE KARAKTERA
UPOTREBOM DUBOKIH NEURONSKIH MREŽA****ONE SOLUTION OF OPTICAL CHARACTER RECOGNITION USING DEEP NEURAL
NETWORKS**Dejan Ikonić, *Fakultet tehničkih nauka, Novi Sad***Oblast – ELEKTROTEHNIKA I RAČUNARSTVO**

Kratak sadržaj – U okviru ovog rada pokazana je praktična primena dubokih konvolucijskih neuronskih mreža u oblasti detekcije karaktera od značaja na slikama sa identifikacionim oznakama bandera ulične rasvete. Zadatak je realizovan u tri iteracije sa dobijenom tačnošću od 93,04%.

Ključne reči: Duboko učenje, Neuronske mreže, Veštačka inteligencija

Abstract – This paper presents one approach of automated optical character recognition on photographs with streetlights identification using deep Convolutional Neural Network. The task was realized in three iterations, with an accuracy of 93.04%.

Keywords: Deep learning, Artificial intelligence, Neural networks

1. UVOD

U ovom radu opisano je jedno rešenje za detekciju petocifrenih identifikacionih brojeva na identifikacionim listovima sa fotografijama stubova ulične rasvete. Cilj rada je razvoj sistema koji radi na principu najnovijih tehnologija i koji se uz jednostavnu pripremu ulaznog skupa i konfiguraciju parametara sistema može primeniti za rešavanje različitih slučajeva optičkog prepoznavanja karaktera (eng. Optical Character Recognition - OCR).

Alati veštačke inteligencije su pogodni za ovakve probleme jer nude rešenja segmentacije i klasifikacije na šta se OCR problem može i svesti.

Primere primene možemo naći od ranih radova kao što je opisano u [1] iz 1996. godine do novih rešenja kao što nudi članak [2] iz 2019. godine. Implementacije zasnovane na korišćenju veštačkih neuronskih mreža u prepoznavanju teksta i simbola, prvi je iskoristio Dan Ciresan i njegove kolege na takmičenju [3]. Njihove neuronske mreže su bile prvi veštački prepoznavajući šablona, koji su mogli da dostignu performanse koje mogu da pariraju čoveku kod prepoznavanja saobraćajnih znakova i rukopisa [4].

Sistem koji je predstavljen u ovom radu realizovan je upotrebom duboke konvolucijske neuronske mreže koja se koristi za segmentaciju karaktera, i daje ulaze za

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Vladimir Bugarski, docent.

neuronsku mrežu sa povratnim vezama za klasifikaciju identifikovanih karaktera.

Opisan sistem je kreiran, a rezultati obučavanja i testiranja su predstavljeni u ovom radu.

2. BAZA ULAZNIH PARAMETARA MREŽE

Na fotografijama u bazi prikazani su stubovi ulične rasvete sa istaknutim nalepnicama koje nose njihov identifikacioni broj. Cilj sistema je da za datu fotografiju detektuje i prikaže njen identifikacioni broj. Bazu sačinjava skup od 11.000 fotografija. Baza slika je podeljena tako da je za obuku mreže odvojeno 9.000 fotografija, dok je ostatak od 2.000 služio kao baza za testiranje mreže. Primer fotografije je vidljiv na slici 1.



Slika 1. Primer fotografije za prepoznavanje

Korišćene su dve metode preprocesiranja: smanjivanje dimenzija i augmentacija.

2.1. Iteracija 1: 64x64 bez augmentacije

Fotografije za trening i testiranje, su sa originalne veličine smanjene na veličinu 64x64 piksela. Osim smanjenja veličine, osetno je smanjen i kvalitet fotografija, što je uticalo na kvalitet same obuke.

2.2 Iteracija 2: 128x128 bez augmentacije

Originalne fotografije u iteraciji 2 smanjene su na veličinu 128x128 piksela.

2.3 Iteracija 3: 128x128 sa augmentacijom

Cilj, u iteraciji 3, je bio da se poveća broj nekarakterističnih fotografija u skupu, kao što je slika 2, a da se pri tome ne preoptereti obuka mreže sa velikim skupom podataka. To je postignuto koristeći operaciju rotacije iz [5] koja rotira fotografiju na ulazu za nasumičnu vrednost između zadatih granica, od -45° do 45°. Na ovaj način je dobijen skup od 18.000 fotografija kao trening skup podataka.



Slika 2. Primer slabo vidljive fotografije za prepoznavanje

3. ARHITEKTURA MREŽE

Implementiran je sistem za prepoznavanje koji sadrži konvolucijsku neuronsku mrežu i povratnu neuronsku mrežu sa dugotrajnim pamćenjem (eng. Long Short-Term Memory - LSTM) sa modelom zapažanja.

Konvolucijska neuronska mreža je zadužena za segmentaciju fotografija i učenje regiona od interesa. Ova mreža detektuje delove na fotografiji koji sadrže karaktere za prepoznavanje, od njih pravi jednodimenzionalne sekvencijalne regione koje predaje LSTM delu mreže čiji je zadatak da date regione analizira, prepozna karaktere i prosledi rezultate na izlaz. Korišćena konvolucijska mreža ima sledeće slojeve:

1. Sloj sa ispravljačkim mehanizmom praćen slojem sažimanja maksimalnom vrednošću po obe dimenzije;
2. Sloj sa ispravljačkim mehanizmom praćen slojem sažimanja maksimalnom vrednošću po obe dimenzije;
3. Sloj sa normalizacijom i ispravljačkim mehanizmom;
4. Sloj sa ispravljačkim mehanizmom praćen slojem sažimanja maksimalnom vrednošću po širini;
5. Sloj sa normalizacijom i ispravljačkim mehanizmom;
6. Sloj sa ispravljačkim mehanizmom praćen slojem sažimanja maksimalnom vrednošću po širini;
7. Sloj sa normalizacijom i ispravljačkim mehanizmom praćen slojem sažimanja maksimalnom vrednošću po širini.

Kao enkoder i dekodeer u implementiranom sistemu koristi se LSTM mreža i model zapažanja. Broj skrivenih slojeva u mrežama korišćenim za enkoder i dekodeer je 128. Enkoder kao ulaz prima vektor karaktera za prepoznavanje koji daje prethodna mreža. Dekodeer koristi model zapažanja koji omogućava da naučene karakteristike karaktera koji se dekodiraju. Zapažanje funkcioniše tako što računa kontekst vektor sa težinama svakog elementa na ulazu po važnosti za određenu operaciju na izlazu. Pored ovoga dekodeer koristi pretragu snopom koja pomoću jezičkog modela određuje vrednost svakog karaktera za prepoznavanje. U ovom slučaju jezički model su cifre 0-9.

4. PROCES UČENJA SISTEMA

Za treniranje mreža u ovom radu korišćen je algoritam prostiranja greške unazad. Na ulazni sloj mreže se dovode ulazni podaci da bi se započelo učenje. Sledeći korak je propuštanje signala kroz mrežu i generisanje krajnjeg rezultata na izlaznom sloju. Greška predstavlja razliku između generisanih i željenih rezultata. Tokom učenja

izlaz mreže se približava željenom, a greška se smanjuje. Prvo se određuje greška u izlaznom sloju, zatim se za prethodni sloj računa koliko je uticaja imao svaki neuron na grešku sledećeg sloja [6]. Greška izlaza neurona se računa kao suma razlike kvadrata dobijenog izlaza g_k i željenog izlaza o_k i data je formulom (1):

$$J(w) = \frac{1}{2} \cdot \sum_{k=1}^c (g_k - o_k)^2 = \frac{1}{2} \|g - o\| \quad (1)$$

U formuli (1) g i o su dobijeni i željeni izlazni vektori dužine c , a w predstavlja težine unutar mreže. Težine su na početku nasumično inicijalizovane i vrednosti težina se menjaju u toku treniranja, u cilju smanjivanja greške, po izrazu (2):

$$\Delta w = -\eta \frac{\partial J}{\partial w} \quad (2)$$

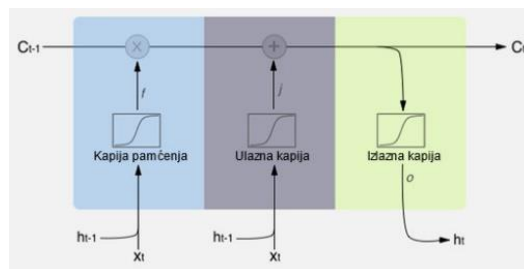
Relativna stopa učenja je predstavljena sa η . Jednačina (1) pokazuje da J nikada ne može dostići negativnu vrednost, što osigurava da će proces učenja u jednom trenutku da se zaustavi, po nekom od zadatih kriterijuma. Izračunata greška trenutnog sloja se zatim propagira na prethodni sloj. Ovo se postiže sumiranjem uticaja neurona iz prethodnog sloja j na neurone iz trenutnog sloja i , kao što je prikazano izrazom (3):

$$\frac{\partial J}{\partial y_i} = \sum_j \frac{\partial J}{\partial net_j} \frac{\partial net_j}{\partial y_i} = \sum_j w_{ij} \frac{\partial J}{\partial net_j} \quad (3)$$

Parametar net predstavlja aktivacijsku funkciju. Sledeći korak je računanje parcijalne derivacije po težinama, po jednačini (4):

$$\frac{\partial J}{\partial w_{ij}} = \frac{\partial J}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} \quad (4)$$

Kod LSTM mreže postoje kapije za propuštanje znanja i stanje ćelije. Stanje ćelije u ovoj mreži služi za propagiranje znanja iz prošlih iteracija, dok kapije nose informaciju da li određenu osobinu u ćeliji mreže treba uzeti u obzir ili ne. Kapije predstavljaju sigmoidne aktivacijske funkcije sa izlaznom vrednošću između 0 i 1. Postoje tri kapije u jednoj ćeliji povratne LSTM mreže (sika 3): ulazna i , kapija pamćenja f i izlazna o .



Slika 3. LSTM ćelija

Jednačina za ulaznu kapiju (5) nam govori koje nove informacije će uticati na stanje date ćelije. Kapija pamćenja u svojoj jednačini (6) opisuje koje informacije ne treba uzeti u obzir za datu ćeliju. Izlazna jednačina (7) se koristi kao aktivacijska funkcija konačnog izlaza LSTM ćelije.

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad (5)$$

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (6)$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad (7)$$

gde je:

- σ - aktivacijska sigmoidna funkcija
- w_x - težina neurona na datoj kapiji (x)
- h_{t-1} - izlaz prethodnog LSTM bloka
- x_t - tekući ulaz
- b_x - gradijent date kapije (x)

Stanje LSTM ćelije, kandidat stanja i konačni izlaz dobijaju se po formulama (8), (9) i (10).

$$\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_c) \quad (8)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (9)$$

$$h_t = o_t * \tanh(c^t) \quad (10)$$

gde je:

- c_t - stanje ćelije u trenutku t
- \tilde{c}_t - kandidat za stanje ćelije u trenutku t

Jednačina (9) pokazuje da u svakom trenutku LSTM ćelija zna šta treba da zanemari iz prethodnog stanja, a šta da uzme u obzir iz tekućeg ulaza. Nakon ovoga, stanja ćelija se filtriraju i propuštaju kroz aktivacijsku funkciju koja predviđa šta se treba pojaviti kao izlaz u datom trenutku t .

Konvolucijska mreža i LSTM mreža su povezane. Izlaz konvolucijske korišćen je kao ulaz LSTM. Obučavanje je podeljeno na manje celine: epohe i gomile. Epoha predstavlja proces u kome je kompletan skup podataka za trening propušten kroz mrežu jedanput, što predstavlja jednu iteraciju gradijentnog spusta. Gomila predstavlja broj podataka za treniranje nakon koga se model obučavanja osvežava.

4.1 Iteracija 1: 64x64 bez augmentacije

Za obuku sistema je korišćeno 9.000 fotografija veličine 64x64 piksela. Podaci su podeljeni u gomile od 150 fotografija i 200 epoha. Ovakva podela znači da se nakon obrađenih 150 fotografija model učenja ažurira, a svih 9.000 fotografija je prošlo kroz mrežu 200 puta. Kriterijum zaustavljanja je postavljen tako da se obuka završi nakon svih 200 epoha ili kada greška padne ispod 10^{-4} . Treniranje je završeno nakon 200 epoha, sa greškom 10^{-2} .

4.2 Iteracija 2: 128x128 bez augmentacije

Za drugi pokušaj, 9.000 fotografija za treniranje mreže veličine 128x128 piksela, podeljeno je u 120 gomila, obučavanih u 220 epoha. Kriterijum zaustavljanja je postavljen na 220 epoha ili dostignutu vrednost greške od 10^{-4} . Greška je u ovom slučaju smanjena na 10^{-3} , nakon što je završeno svih 220 epoha treniranja.

4.3 Iteracija 3: 128x128 sa augmentacijom

Za treću iteraciju skup za obučavanje je, nakon augmentacije, imao 18.000 fotografija veličine 128x128 piksela. Fotografije su podeljene u gomile od po 80, dok je definisani broj epoha bio 250. Kriterijum zaustavljanja je podešen na završetak svih 250 epoha ili vrednost greške 10^{-4} . Obučavanje je prekinuto kada je greška pala ispod 10^{-4} , da bi se izbegla loša generalizacija podataka.

5. TESTIRANJE MREŽE I REZULTATI

Testiranje je izvršeno nad skupom slika koje nisu učestvovala u trening fazi da bi se proverilo da mreža nije pretrenirana i da nije naučila neke odlike specifične samo za podatke u okviru trening skupa [7].

Skup od 2.000 fotografija za testiranje je propušten kroz prethodno obučenu mrežu, jedna po jedna, zajedno sa očekivanim vrednostima za svaku od njih. Klasifikovane vrednosti karaktera su zatim sačuvane zajedno sa očekivanim. Nad rezultatima testiranja je izvršena analiza koja govori o uspešnosti predloženog sistema.

Testiranih 2.000 fotografija su sadržale po 5 karaktera kao identifikacioni broj bandere, što znači da je sistem trebao da prepozna ukupno 10.000 karaktera. Odnos prepoznatih karaktera i ukupnog broja za prepoznavanje uzet je kao parametar uspešnosti sistema. Pored ovoga, analizirane su još neke karakteristike koje su se istakle kao bitne prilikom treniranja i testiranja: uspešnost prepoznavanja kompletnog identifikacionog broja, prepoznavanje u odnosu na cifru koja treba da se prepozna i u odnosu na poziciju u identifikacionom broju.

Kada se uporede rezultati treniranja i testiranja mreže (Tabela 1) može se uočiti da je svakom od izmena koje su napravljene, poboljšana obučenos sistema da uspešno identifikuje fotografije u fazi testiranja. To se najbolje vidi u broju prepoznatih karaktera, koji se kreće od 66,53% u prvoj iteraciji do 93,04% u trećoj.

Tabela 1. Uspešnost prepoznavanja

	Očekivano	Prepoznato	Procenat uspešnosti
Iteracija 1	10000	6653	66,53%
Iteracija 2		8744	87,44%
Iteracija 3		9304	93,04%

Sistem je u iteraciji 2, povećanjem slika za trening i test, obučen da mnogo bolje uočava cifre na fotografijama koje su napravljene sa veće udaljenosti ili imaju neku smetnju. To se može jasno videti i u smanjenju broja fotografija sa velikim brojem promašaja kroz iteracije (Tabela 2).

Tabela 2. Uspešnost prepoznavanja kompletnog identifikacionog broja kroz iteracije

Prepoznato cifara	Iteracija 1	Iteracija 2	Iteracija 3
0	29	0	0
1	276	28	0
2	294	280	12
3	405	321	142
4	406	539	376
5	590	832	1470
Ukupno:	2000	2000	2000

U iteraciji 3, gde je za trening korišćen veći skup podataka, postignuto je to da je sistem mnogo bolje naučio osobine brojeva koji se ređe javljaju, jer nisu na prve dve pozicije, kao 2 i 3. Ovi podaci su jasno vidljivi u uporednom prikazu uspešnosti u odnosu na broj koji se prepoznaje (Tabela 3).

Tabela 3. *Uspješnost prepoznavanja u odnosu na broj koji se prepoznaje*

Očekivana cifra	Broj cifara	Iteracija 1	Iteracija 2	Iteracija 3
0	816	569	678	742
1	743	448	674	714
2	1993	1636	1940	1973
3	1579	686	1418	1538
4	845	221	717	836
5	753	568	656	692
6	673	496	589	632
7	733	630	642	665
8	895	612	639	680
9	970	787	791	832
Ukupno:	10000	6653	8744	9304

Analiza uspješnosti, u odnosu na poziciju broja koji se prepoznaje, pokazuje da se i dalje cifre slabije prepoznaju što im je veća pozicija u identifikacionom broju. Uzrok je kombinacija dva pomenuta problema, u većim pozicijama se češće pojavljuju brojevi koji nisu 2 i 3, a fotografije su takve da su često te pozicije nejasno vidljive. Popravkom parametara obučavanja kroz 3 iteracije uspeli smo da i u ovom kontekstu postignemo veliku tačnost (Tabela 4).

Tabela 4. *Uspješnost u odnosu na poziciju broja koji se prepoznaje*

Pozicija	Broj cifara	Iteracija 1	Iteracija 2	Iteracija 3
1	2000	569	1921	1990
2	2000	448	1842	1975
3	2000	1636	1830	1929
4	2000	686	1674	1816
5	2000	221	1477	1594
Ukupno:	10000	6653	8744	9304

Krajnji rezultat uspješnosti sistema se smatra zadovoljavajućim za skup podataka koji je korišćen.

6. ZAKLJUČAK

Glavni zadatak ovog rada bio je da da jedno rešenje problema koji je dosta rasprostranjen u modernom svetu: optičko prepoznavanje karaktera (OCR).

Opisani sistem sastoji se od konvolucijske neuronske mreže koja za zadatak ima da vrši segmentaciju fotografija i izdvaja regione od interesa, koje sadrže karaktere za prepoznavanje, i predaje ih LSTM mreži sa modelom zapažanja koja radi klasifikaciju dobijenih regiona i kao rezultat daje karaktere sa slike u digitalnoj formi.

Glavna prednost ovakvog sistema su prilično jednostavno preprocesiranje, jer ne zahteva obeležavanje regiona od interesa u cilju pripreme za učenje. Zbog specifičnosti skupa podataka odrađen je i korak augmentacije ulaznog skupa podataka u cilju isticanja osobina prisutnih na manjem delu skupa.

Sistem opisan u ovom radu je moguće primeniti na bilo koji OCR problem. Ovo je moguće postići jednostavnim treniranjem nad novim skupom uz eventualnu izmenu karaktera skupa za prepoznavanje.

Dobijeni rezultati na test skupu su prilično dobri za ati problem i korišćeni skup podataka. Korišćen je veoma veliki broj fotografija za prepoznavanje i dobijena je tačnost od 93,04%. S obzirom na raznolikost skupa podataka, gde su fotografije pravljene sa različitih udaljenosti od znaka, pod različitim uglovima, a sami objekti bili zakrivljeni, ovakav rezultat je zadovoljavajući. Uspješnost bi se mogla povećati dodatnim filtriranjem skupa podataka, gde bi se izbacile nejasne fotografije ili primenom namenskog sistema koji se bazira na detekciji i zahteva ručnu pripremu trening skupa.

7. LITERATURA

- [1] C. Tanprasert, T. Koanantakool: "Thai OCR: a neural network application", Proceedings of Digital Processing Applications, vol. 1, pp. 90-95, Perth, Australia, 1996.
- [2] S. Ali, Z. Shaukat, M. Azeem, Z. Sakhawat, T. Mahmood, K. ur Rehman: "An efficient and improved scheme for handwritten digit recognition based on convolutional neural network", SN Applied Sciences, Springer Nature Switzerland, vol. 1, art. 1125, pp. 1-9, 2019.
- [3] D.C. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber: "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images", Advances in neural information processing systems, Lake Tahoe, Nevada, United States, 2012.
- [4] S. Gavran: "Veštačke neuronske mreže u istraživanju podataka: pregled i primena", Univerzitet u Beogradu, Matematički fakultet, Master rad, 2016.
- [5] <https://github.com/mbloice/Augmentor> (pristupljeno u septembru 2021.)
- [6] A. Krizhevsky, I. Sutskever, G.E. Hinton: "Imagenet classification with deep convolutional neural networks", Communications of the ACM, Association for Computing Machinery, vol. 60, issue 6, pp. 84-90, 2017.
- [7] S. Hassanpour, N. Tomita, T. DeLise, B. Crosier, L. A. Marsch: "Identifying substance use risk based on deep neural networks and Instagram social media data", Neuropsychopharmacol, Springer Nature, vol. 44, pp. 487-494, 2019.

Kratka biografija:



Dejan Ikonić rođen je 27.07.1991. god. u Sarajevu. Osnovne studije iz oblasti elektrotehnika i računarstvo završio je 2015. na Fakultetu tehničkih nauka u Novom Sadu.