

Duboke konvolucijske neuronske mreže – koncepti i aktuelna istraživanja

Marko M. Dabović, Igor I. Tartalja

Apstrakt — U radu su opisani koncepti i arhitektura konvolucijskih neuronskih mreža, čija primena je široko rasprostranjena. Značaj primene ove vrste neuronskih mreža je veliki, te su i istraživanja u ovoj oblasti od posebnog interesa. Postignuti su već značajni rezultati, te su napori oko sistematizacije u oblasti opravdani. U radu su predstavljeni važni rezultati istraživanja u oblasti, predložena je konzistentna terminologija i ukazano na izazove daljih istraživanja.

Gljučne reči — Konvolucijske neuronske mreže; CNN; mašinsko učenje; duboko učenje

I. UVOD

U radu je predstavljen detaljan pregled literature u oblasti konvolucijskih neuronskih mreža (eng. *Convolutional Neural Networks – CNNs*) [1], koje predstavljaju aktuelan i istraživački atraktivan domen u širokoj oblasti veštačke inteligencije (eng. *Artificial Intelligence – AI*). CNNs su biološki inspirisana verzija višeslojnog perceptrona (eng. *Multilayer Perceptron – MLP*). Prema svojoj arhitekturi CNNs se mogu svrstati u duboke neuronske mreže (eng. *deep neural networks*), za koje se vezuje pojam dubokog učenja (eng. *Deep Learning – DL*). DL je relativno nova oblast mašinskog učenja, koja je u ekspanziji. Posle dugog niza godina zastoja razvoja u oblasti AI, duboke arhitekture neuronskih mreža, koje se mogu primenjivati u različitim domenima, pokazale su veliki značaj i dale doprinos u rešavanju naizgled, za jednu mašinu, nerešivih problema. CNN sačinjava jedan ili više konvolucionih slojeva i opcionalno jedan ili više potpuno povezanih (eng. *Fully Connected – FC*) slojeva, koji se mogu sresti u konvencionalnim višeslojnim neuronskim mrežama. CNNs su projektovane tako da prednost postižu u radu sa 2D strukturama, kao što su slike ili ulazi poput govornog signala (eng. *speech signal*) [2], dok najnovije studije [3] pokazuju da postižu značajne rezultate i sa 3D strukturama, čak i na nivou eksperta. CNNs su u osnovi implementacije mnogih *state-of-the-art* rešenja u oblasti računarske vizije (eng. *Computer Vision – CV*). Smatra se da su od 2014. godine duboke CNNs postale „glavni tok“ i predmet istraživačkih napora i zapaža se da od tada postižu značajne prednosti u raznim testovima [4].

II. KARAKTERISTIKE

CNN je dobila ime po konvoluciji, široko korišćenom operatoru, koji se primenjuje u obradi slike i signala (eng. *image and signal processing*) [5]. U ovim oblastima, reč je o

konvolucionim filterima koji se koriste za izoštravanje i zamućenje slika, kao i za detektovanje ivica. Sa aspekta CNNs, značajna je primena algoritama za detektovanje ivica, kojima se otkrivaju objekti na slici. Tamo gde klasične neuronske mreže pokazuju svoje nedostatke, tu CNNs pokazuju svoje prednosti. Na primer, kada se kao ulazni podaci koriste slike, slika dimenzija 200×200 piksela, pretvara se u 40000 neurona u ulaznom sloju, što je ogroman broj i praktično je nemoguće treniranje potpuno povezanih neuronskih mreža. Za sliku u boji treba uzeti u obzir i 3 kanala kojima su predstavljene osnovne boje (RGB), što daje 120000 neurona u ulaznom sloju. Ideja CNNs jeste da se postavi veći broj slojeva za otkrivanje bitnih osobina ulaznih podataka. U skladu sa tom idejom se konvolucionni filteri primenjuju na sliku kako bi se izvukle korisne karakteristike i kreirale njihove mape (eng. *feature maps*). Primenom filtera na ulaznu sliku, pored otkrivanja značajnih karakteristika, vrši se i redukcija rezolucije. Nedostatak CNNs je potreba za značajnim hardverskim resursima, kao i to što su dosadašnji modeli obučavani velikim brojem slika, ali noviji rezultati [6] sugerišu da se visoke performanse mogu postići i sa relativnom malim brojem uzoraka.

III. KONCEPTI CNNs

A. Ulazna zapremina

CNNs se najčešće primenjuju na slike (eng. *image data*) koje su matrice piksela i predstavljaju se svojom širinom, visinom i vrednostima piksela. Za slike u boji, svaki od 3 kanala (RGB) se uobičajeno predstavlja vrednostima piksela u opsegu 0-255. RGB slika se predstavlja 3D strukturom, koja se zove ulazna zapremina (eng. *input volume*).

B. Karakteristike

Jedan od koncepta koji se koristi u kontekstu CNNs jeste i karakteristika ili bitna odlika (eng. *feature*) slike. U literaturi se sreće i izraz koristan obrazac (eng. *useful pattern*). Karakteristika se dobija iz slike koja se u mrežu uvodi putem ulaznog sloja (eng. *input layer*). Ono što CNNs uče jesu karakteristike ulaznih slika (eng. *input images*) kojima se neuronska mreža obučava.

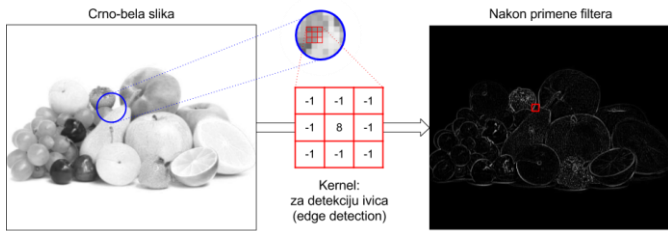
C. Filteri

Konvolucionni filter u bazičnim CNNs je generalizovan linearni model (eng. *Generalized Linear Model – GLM*) za region slike (eng. *image patch*) na koji se primenjuje [7]. Pored naziva filter, u kontekstu CNNs, u literaturi se koristi i naziv konvolucionni kernel. Filter se predstavlja dvodimenzionalnom matricom, malih dimenzija, u poređenju sa slikom na koju se primenjuje i sastoji se od realnih vrednosti. Filteri se odnose na konvolucionni sloj arhitekture

Marko M. Dabović – Univerzitet u Beogradu – Elektrotehnički fakultet, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: marko.dabovic@outlook.com).

Igor Tartalja – Univerzitet u Beogradu – Elektrotehnički fakultet, Bulevar Kralja Aleksandra 73, 11020 Beograd, Srbija (e-mail: tartalja@rcub.bg.ac.rs).

CNN, u kome se primenjuju na sliku koja se obrađuje u datom sloju. Sam postupak primene filtera na sliku, zove se konvolucija. O postupku konvolucije biće više reči u delu IV. U zavisnosti od filtera koji se primenjuje na sliku, dobijaju se različiti rezultati. Na Sl. 1. je primenjen jedan od filtera za detektovanje ivica (eng. *edge detection*).



Sl. 1. Sa leve strane je originalna slika, u sredini je *filter* označen crvenom bojom, primenjen na *receptivno polje*. Sa desne strane je rezultujuća slika nakon primene filtera.

D. Receptivno polje

Pojam receptivno polje (eng. *receptive field*) potiče iz medicinskih nauka i povezan je sa idejom za nastanak CNNs. U domenu AI pojam se pojavljuje u kontekstu veštačkih neuronskih mreža, naročito CNNs. Ideja da oponašaju način na koji mozak životinje (mačke) funkcioniše je osnova na kojoj je zasnovan dizajn CNNs. Rad [8] ukazuje na to da se vizuelni korteks mačke sastoji od kompleksno raspoređenih ćelija. Ove ćelije su osetljive na male pod-regione vidnog polja, koje zovemo receptivna polja ili lokalni receptori, koji su poređani tako da prekrivaju vidno polje [1]. Konkretno, umesto da svaki veštački neuron u nekom sloju bude povezan sa svim neuronima u susednom sloju, kao što je to kod MLP, neuroni se raspoređuju u 3D strukturu (širina×visina×dubina), pri čemu se svaki neuron povezuje sa određenim regionom (receptivnim poljem). Povezivanje svih neurona sa svim mogućim regionima ulazne zapremine nije praktično, jer može dovesti do prevelikog broja težina (eng. *weights*) za treniranje, što implicira preveliku kompleksnost izračunavanja. Pojam receptivnog polja u kontekstu CNNs upućuje na deo slike koji je filteru vidljiv u datom trenutku.

E. Korak i dopuna

Napredovanje ili korak (eng. *stride*) i dopuna (eng. *padding*) su dva važna parametra koja direktno zavise od kreatora arhitekture CNN, što implicira da se mogu menjati kao i dimenzije filtera, kako bi se modifikovalo ponašanje svakog sloja arhitekture. *Korak* (S) se odnosi na korak pomeranja filtera preko slike sa ulaznog sloja. Kako bi se izvukle karakteristike niskog nivoa (eng. *low-level features*), kao što su pikseli slike, cilj je da se u prvim slojevima CNN očuva što veći broj informacija. S tim ciljem, se uvodi drugi pojam – dopuna (P). Parametar P se odnosi na oivičenje koje se sastoji od konstantnih vrednosti piksela. Najčešća vrednost piksela kojima se oivičava je nula, odatle pojam dopune nulama (eng. *zero-padding*). Ako je, na primer, u ulaznom sloju slika 7×7 , kao ulazna zapremina, na koju će se primeniti filter 3×3 , pri čemu će *korak* biti 1, kao izlazna zapremina dobija se slika dimenzija 5×5 . Primenom konvolucije, dimenzije slike se brzo smanjuju. Smanjenjem dimenzija

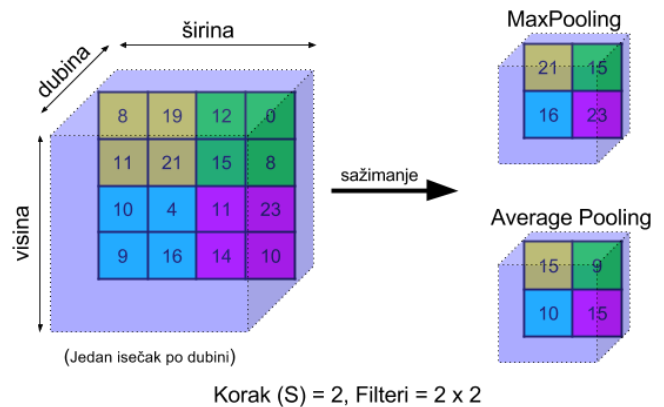
originalne slike gube se mnoge informacije, a to nije poželjno, bar ne na početku postupka.

F. Sažimanje maksimumom

Još jedan važan koncept CNNs je i sažimanje maksimumom (eng. *max pooling*), koji je forma nelinearnog smanjivanja broja piksela (eng. *down-sampling*) [2]. U zavisnosti od tipa sažimanja bira se piksel sa određenom vrednošću u određenom regionu. Kod operacije sažimanja maksimumom se bira piksel sa najvećom vrednošću. Sloj sažimanja (eng. *pooling layer*) koristi se sa ciljem progresivnog smanjenja veličine slike, samim tim i broja karakteristika, što dovodi do smanjenja kompleksnosti izračunavanja. Najčešće korišćena operacija sažimanja je sažimanje maksimumom, čiji primer je dat na Sl. 2.

Sažimanje maksimumom je u računarskoj viziji korisno iz dva razloga [2]:

1. Uklanjanjem nemaksimalnih vrednosti (eng. *non-maximal values*), smanjuje se izračunavanje.
2. Pruža invarijantnost translacije (eng. *translation invariance*) – prepoznavanje objekta neće zavistiti od eventualne transformacije translacije objekta unutar slike.



Sl. 2. Eliminisanjem nemaksimalnih vrednosti iz matrice oblika 4×4 , primenom operacije sažimanja maksimumom (veličinom filtera 2×2 i korakom 2), dobijamo matricu oblika 2×2 , koja sadrži maksimalne vrednosti. Uz operaciju sažimanja maksimumom, prikazana je i matrica dobijena primenom operacije sažimanja usrednjavanjem (eng. *average pooling*).

G. Hiper-parametri

Ono što arhitekturu neke CNN čini specifičnom, jeste izbor hiper-parametara (eng. *hyperparameters*). Glavni hiper-parametri su receptivno polje (R), dopuna nulama (P), dimenzije ulaznog volumena ($\dot{S} \times V \times D$) i korak (S) [9]. Izbor navedenih parametara je od velike važnosti za arhitekturu neke CNN. Tri parametra koja utiču na veličinu izlazne zapremine (eng. *output volume*) su: *dubina* (eng. *depth*), *korak* i dopuna nulama [10].

IV. ARHITEKTURA I ALATI CNNs

Iako mnogi radovi ukazuju na različite tipove arhitekture, tri su osnovne gradivne komponente (sloja) koje su sastavni deo osnovne (eng. *basic*) konvolucione neuronske mreže. To su: konvolucioni, sloj sažimanja i FC sloj. Primer ovakve

arhitekture CNN je LeNet-5 predstavljen u radu [11]. Reč je o eksperimentalnom modelu koji su autori koncipirali tako da se sastoji od sedam slojeva, ne uključujući ulazni sloj. Pored ulaznog sloja, mrežu čine i tri konvoluciona sloja, između kojih se nalazi po jedan sloj sažimanja (ukupno dva sloja) i dva FC sloja. LeNet-5 je arhitektura, nastala sa ciljem prepoznavanja rukom pisanih karaktera. Iz simulacije [12] se može videti način rada LeNet-5 arhitekture. Uz pomenuta tri sloja svake CNN, može se naći i sloj aktivacione funkcije, čija je uloga značajno poboljšanje performansi CNN u obavljanju određenih zadataka.

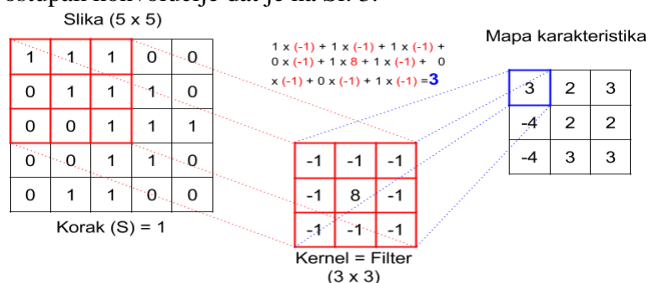
A. Generalni opis arhitekture

1) Ulazni sloj

Ulazni sloj se ne odnosi direktno na arhitekturu CNN. Pojam ulazni sloj je primenljiv na svaku neuronsku mrežu. To je sloj putem koga se podaci uvode u mrežu. Kod CNNs ulazni podatak je najčešće slika. Koncept ulazna zapremina je u direktnoj vezi sa ulaznim slojem.

2) Sloj konvolucije

Razlika CNN u odnosu na druge tipove arhitekture neuronskih mreža ogleda se u sloju gde se vrši konvolucija, po kome je i dobila ime. Konvolucioni sloj (eng. *convolutional layer* ili *conv-layer*), osnovni je gradivni blok CNN i obavlja mnoga veoma zahtevna izračunavanja (eng. *computational heavy lifting*) [13]. Konvolucioni sloj realizuje osnovnu operaciju obučavanja neurona mreže. U svom najopštijem obliku, konvolucija je operacija dve funkcije, čiji su argumenti realne vrednosti [15]. U [15] dat je slikovit primer praćenja putanje svemirskog broda, koji objašnjava operaciju konvolucije. U komparaciji sa pomenutim primerom, dat je osvrt na primenu operacije konvolucije u kontekstu CNNs, gde se kao prvi argument navodi ulaz, a kao drugi filter, pri čemu se izlaz odnosi na mapu karakteristika. Postupak konvolucije dat je na Sl. 3.



Sl. 3. Postupak konvolucije. Na monohromatsku sliku (I) [5×5] se primenjuje filter (K) [3×3], pri čemu je korak S=1. U svakom koraku, vrednosti na istoj poziciji (npr. $I_{1,1}=1$ i $K_{1,1}=8$) se množe i sabiraju sa proizvodima drugih parova poziciono korespondentnih vrednosti.

3) Sloj sažimanja

U kontekstu CNNs, sažimanje je značajan koncept. Uobičajeno je da se u arhitekturi CNN periodično, između sukcesivnih konvolucionih slojeva, umetne sloj sažimanja (eng. *pooling layer*) [13]. Kao što je objašnjeno u delu III.B, primenom operacije sažimanja smanjuje se broj parametara, samim tim i izračunavanja unutar mreže. Smanjenjem broja parametara, može se kontrolisati i pretreniranost (eng. *overfitting*) mreže [13]. Pored najčešće korišćenih sažimanja (usrednjavanjem [16] i maksimumom [17]) postoji veći broj metoda sažimanja, koji se primenjuju u CNNs, a to su: *Mixed*

Pooling [18], *Lp Pooling* [19], *Spatial Pyramid Pooling (SPP)* [20], *Stochastic Pooling* [21] i druge.

4) Sloj aktivacione funkcije (ReLU)

Pravi izbor aktivacione funkcije (eng. *activation function*) može značajno poboljšati performanse CNN, koja se primenjuje za određene zadatke [7]. LeNet-5 arhitektura CNN ne podrazumeva upotrebu aktivacione funkcije iza svakog konvolucionog sloja, ali primetno je da novije arhitekture CNNs koriste aktivacione funkcije na takav način. Jedna od najčešće korišćenih aktivacionih funkcija je ReLU (eng. *Rectified Linear Unit*), pa se u literaturi može naći pojam „ReLU sloj“ (eng. *ReLU layer*), koji upućuje na sloj u kome se primenjuje ReLU aktivaciona funkcija. U okviru rada [7] predstavljene su aktivacione funkcije koje se mogu sresti u skorijim radovima. Pored ReLU, neke od njih su: *Leaky ReLU (LReLU)*, *Parametric ReLU (PReLU)*, *Randomized ReLU (RReLU)*, *Exponential Linear Unit (ELU)* i druge.

5) Potpuno povezan sloj

Naziv potpuno povezanog sloja (eng. *FC layer*) upućuje na njegovu konfiguraciju: svi neuroni u ovom sloju su povezani sa svim izlazima prethodnog sloja. Tipično za FC sloj jeste da se koristi kao poslednji.

B. Različiti pristupi – ILSVRC

Dobar pregled različitih pristupa i arhitektura CNNs, može se naći analizom *ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)* [22]. Reč je o godišnjem takmičenju, pokrenutom 2010. godine sa ciljem poređenja najnovijih i *state-of-the-art* softverskih rešenja u domenu detekcije objekata i klasifikacije slika. *ImageNet* [23] je projekat nastao sa ciljem da istraživačima obezbedi jednostavan pristup velikoj bazi podataka, koja trenutno ima 14,197,122 slika. Može se napraviti analogija ovog takmičenja sa Olimpijadom u domenu računarske vizije. U nastavku su opisani pristupi koji su predstavljeni na ILSVRC-u, a u Tabeli I sumirani rezultati postignuti ovim pristupima.

1) AlexNet (2012)

Radom [24], u kome je Alex Krizhevsky sa koautorima predstavio svoj pristup u rešavanju problema klasifikacije na ILSVRC-u [25] 2012. godine, započelo je novo poglavlje u domenu CNNs. U prilog tome govori i činjenica da je to jedan od najuticajnijih radova u oblasti CNNs, sa citiranošću [26] od preko 10140 puta do sada. Svojim pristupom, Alex Krizhevsky je na ILSVRC-u odneo pobjedu, i to postizući „*top 5 test error rate*“ od 15.4%. Drugoplasirani model je postigao rezultat od 26.2%. U radu [24] predstavljena je *AlexNet* arhitektura neuronske mreže koja ima 60 miliona parametara i 650000 neurona, a čini je 8 slojeva od kojih je 5 konvolucionih slojeva, neki su praćeni slojevima maksimalnog sažimanja, i 3 FC sloja, od kojih je u prva dva korišćena tehnika napuštanja (eng. *dropout*). Reč je o regularizacionom metodu [27] koji ukazuje na potvrđenu visoku efikasnost u smanjenju pretreniranja u FC slojevima.

2) ZF Net (2013)

Oslanjajući se na rad [24] i *AlexNet* arhitekturu CNN, Matthew D. Zeiler i Rob Fergus identifikovali su i pristupili rešavanju dva problema koja su vrlo značajna. Oni zaključuju da Alex Krizhevsky u svom radu nije dovoljno razjasnio kako se ti značajni rezultati postižu i kako se oni mogu poboljšati.

Zeiler i Fergus su u svom radu [28] predstavili mrežu koja je na ILSVRC-u 2013. godine, imala značajan rezultat postižući procenat grešaka u klasifikaciji (eng. *classification error rate*) od 13.5%. Naime kreirali su CNN veoma sličnu *AlexNet* arhitekturi, sa izvesnim modifikacijama. Neuronska mreža *ZF Net* sastoji se, kao i *AlexNet*, od 8 slojeva. Ono što je specifično, jeste da mape karakteristika prolaze kroz *ReLU*, podvrgavaju se sažimanju i normalizovanju kontrasta i tako kroz 5 slojeva. Kako bi izvršili vizualizaciju i sa razumevanjem predstavili šta se zapravo događa unutar modela, Zeiler i Fergus su predložili višeslojnu dekonvoluciju. Oni u radu govore o višeslojnoj dekonvolucionoj mreži (eng. *deconvolutional network*), koja se može posmatrati kao konvolucionni model koji koristi iste komponente, ali unazad.

3) VGG Net i GoogLeNet (2014)

Dve arhitekture CNN koje su 2014. godine bile najzapaženije na ILSVRC-u, bile su *VGG Net* i *GoogLeNet*. Oba modela su odnela pobjedu, i to *GoogLeNet* postižući „*classification error rate*” od 6.7% (26.5% za grešku lokalizacije) i *VGG Net* postižući procenat greške lokalizacije (eng. *localization error rate*) od 25.3% (7.4% za grešku klasifikacije). Karen Simonyan i Andrew Zisserman, kreirali su *VGG Net*, koju su predstavili u svom radu [29]. Ono što karakteriše ovu arhitekturu jesu dubina i jednostavnost. Radi se o arhitekturi koju čini 19 slojeva i koja striktno primenjuje filtere dimenzija 3×3, pri čemu je S=P=1, dok je operacija maksimalnog sažimanja unutar regiona dimenzija 2×2 i S=2. Druga arhitektura predstavljena u radu [30], je posebna inkarnacija arhitekture kodnog naziva „*Inception architecture*” koju je kreirao Christian Szegedy sa koautorima. Reč je o arhitekturi dubine 22 sloja (27 slojeva, ako se računa sažimanje). Ono što karakteriše ovu arhitekturu jeste modul započinjanja (eng. *inception module*). Ukratko reč je o devet specijalnih modula, koji su integrisani u mrežu i ponašaju kao mreža u mreži (eng. *Network In Network - NIN*).

4) ResNet (2015)

Veoma duboka mreža koju čini 152 sloja, na ILSVRC-u 2015.

godine odnela je pobjedu, postižući svega 3.6% za grešku klasifikacije (9.1% za grešku lokalizacije). Kaiming He je sa kolegama u radu [31] predstavio 8 puta dublju, ali jednostavniju arhitekturu od *VGG Net*, gde je predložio radni okvir (eng. *framework*) učenja dubokih ostataka (eng. *deep residual learning*) kojim olakšava optimizaciju i konvergenciju ekstremno dubokih mreža. Naziv *ResNet*, arhitektura je dobila od skraćenice za „*Residual Network*”. U radu [31] autori su predstavili poređenja njihove arhitekture sa *state-of-the-art* rešenjima i diskutovali o nekoliko različitih modela pomenute arhitekture (razlikuju se po broju slojeva od kojih su sačinjeni).

5) Pobjednički ansambli na ILSVRC 2016

Pobjednički tim pod nazivom „*Trimps-Soushen*” (TS) iz Kine, kombinovanjem postojećih modela došao je do zapaženog rezultata i odneo pobjedu na ILSVRC-u 2016. godine. U zavisnosti od problema koji se rešavao, tim je sastavio nekoliko ansambala (A) (eng. *ensembles*) arhitektura. Konkretno, Jie Shao je sa koautorima, za rešavanje problema klasifikacije, sastavio ansambl koji je postigao procenat greške lokalizacije od 7.7% , dok je drugi pobjednički ansambl, istog tima, ostvario procenat greške klasifikacije od 2,99%.

C. Transfer učenja

U praksi, veoma mali broj korisnika trenira celu CNN od nule (eng. *from scratch*) [32]. Duboke neuronske mreže zahtevaju veliku količinu podataka, koju često nije moguće obezbediti, zbog čega se u kontekst neuronskih mreža, uvodi pojam transfer učenja (eng. *transfer learning*). Navedeni pojam upućuje na razmenu znanja između različitih modela neuronskih mreža. Kada se govori o razmeni znanja, misli se na razmenu naučenih karakteristika istreniranih mreža, sa mrežama koje se primenjuju na novi problem. Moderne duboke neuronske mreže pokazuju interesantan fenomen, da kada se treniraju slikama, imaju tendenciju da u prvom sloju uče karakteristike, koje liče na Gaborove filtere ili na mrlje u boji (eng. *color blobs*) [33]. U radu [33] diskutuje se o mogućnostima upotrebe razmenjenih karakteristika. Umesto obučavanja celih mreža, koristi se prethodno navedeno znanje,

TABELA I
RAZLIČITE ARHITEKTURE CNNs PREDSTAVLJENE NA ILSVRC-U

Naziv	ILSVRC			Broj slojeva	Vreme treniranja (GPU)	Specifičnost
	greška klasifikacije	greška lokalizacije	Godina predstavljanja			
AlexNet	15.4%	33.6%	2012	8	~6 dana (2 x GTX580)	Početak novog poglavlja u CNN
ZF Net	13.5%	-	2013	8	12 dana (GTX580)	Dekonvolucija
VGG Net	7.4%	25.3%	2014	19	2-3 nedelje (4 x NVIDIA Titan Black)	Jednostavna i duboka
GoogLeNet	6.7%	26.5%	2014	22 (27)	nedelju dana (“nekoliko high-end”)	Modul započinjanja
ResNet	3.6%	9.1%	2015	>150	2-3 nedelje (8 GPU-a)	Blok ostatka
TS – A2	2.99%	7.76%	2016			Kombinacija postojećih modela
TS – A3	2.99%	7.7%				

pri čemu se u zavisnosti od problema, vrše određena podešavanja. Cilj je prenošenje znanja iz izvornog domena (eng. *source domain* – D_S) u ciljni domen (eng. *target domain* – D_T), pri čemu se zadaci učenja (eng. *learning tasks* – T) mogu, ali ne moraju podudarati [34]. Transfer učenja može se klasifikovati na tri različita podešavanja[34]:

- induktivni,
- transduktivni i
- nenadgledani.

Specifičnost induktivnog transfera učenja (eng. *inductive transfer learning*) je da su zadaci, izvorni (eng. *source task* – T_S) i ciljni (eng. *target task* – T_T) različiti ($T_S \neq T_T$), bez obzira da li su domeni isti ili nisu. Kod ovog tipa transfera potrebno je malo označenih podataka u D_T . Za razliku od induktivnog, kod transduktivnog transfera učenja (eng. *transductive transfer learning*) zadaci su isti ($T_S = T_T$), dok su domeni različiti. U ovom slučaju, u D_S na raspolaganju su mnogi označeni podaci, dok ih u D_T nema. Slično induktivnom, nenadgledani transfer učenja (eng. *unsupervised transfer learning*) ima različite zadatke ($T_S \neq T_T$), ali je fokus na rešavanju zadataka nenadgledanog učenja u ciljnom domenu. Ni u jednom domenu, kod ovog tipa transfera, ne postoje označeni podaci prilikom treniranja.

D. Softverski alati

Povećanjem broja participanata u DL domenu, povećao se i broj softverskih alata koji se koriste za rešavanje mnogobrojnih problema iz domena mašinskog učenja. Alati više nisu rezervisani samo za mali broj programskih jezika, a veliki broj njih je optimizovan za kreiranje rešenja, izvršivih na veoma moćnim grafičkim adapterima. Paralelizacijom biblioteka, postiglo se veoma značajno ubrzanje postupka obučavanja veoma zahtevnih algoritama, što se može videti po vremenu obučavanja prikazanom u Tabeli 1. U okviru [35] se može naći spisak alata koji se koriste u *deep learning* domenu, kao i URL adrese do njihovih resursa.

V. ZAKLJUČAK

CNNs su danas u širokoj upotrebi i njihov značaj je veliki. Pokazano je da je preciznost klasifikacije kod novijih, veoma dubokih CNNs, veoma visoka. Jedna od posledica te činjenice je primena dubokih CNNs u medicini, ali na nivou istraživanja i pilot-studija za sada. Moć prepoznavanja tumorsnih tkiva pomoću CNNs je na približno istom nivou kao kod eksperta.

Problemi, poput detekcije laži sa lica ispitanika, su nešto čime se istraživači intenzivno bave. Mnoge velike kompanije ulažu velika sredstva u istraživanja u oblasti DL/CNNs. Duboke arhitekture pokazale su se korisnim čak i za rešavanje nekih specifičnih problema, kao što je autonomno igranje računarskih igara od strane mašine ili prepoznavanje aktivnosti ljudi iz videa. Ceo razvoj DL domena, praćen je razvojem veoma snažnih računara, čije performanse rastu iz dana u dan, što je od ključne važnosti za budući rad istraživača u ovoj oblasti.

ZAHVALNICA

Marko Dabović je stipendista Ministarstva prosvete, nauke i tehnološkog razvoja Republike Srbije, angažovan na projektu

TR32010. Igor Tartalja je angažovan na projektima TR32039 i TR32047 finansiranim od strane istog Ministarstva.

LITERATURA

- [1] Computer Science Department, "UFLDL Tutorial-Convolutional Neural Network", <http://ufldl.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/>, Feb, 2017.
- [2] C. Gulcehre, "Convolutional Neural Networks (LeNet)", <http://deeplearning.net/tutorial/lenet.html/>, Feb, 2017.
- [3] A. Mehrtash, A. Sedghi, M. Ghafoorian, M. Taghipour, C. M. Tempany, W. M. Wells, A. Fedorov, "Classification of clinical significance of MRI prostate findings using 3D convolutional neural networks", *Medical Imaging 2017: Computer-Aided Diagnosis*, 2017.
- [4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the Inception Architecture for Computer Vision.", *Computer Vision and Pattern Recognition*, 2015, <https://arxiv.org/abs/1512.00567/>.
- [5] R. B. Fisher, T. P. Breckon, K. Dawson-Howe, A. Fitzgibbon, C. Robertson, E. Trucco, C. K. I. Williams, "Dictionary of computer vision and image processing", Second Edition, Wiley, 2014.
- [6] D. R. Plata, R. Ramos-Pollán, F.A. González, "Effective training of convolutional neural networks with small, specialized datasets", *J. of Intelligent & Fuzzy Sys.*, vol. 32, no. 2, pp. 1333-1342, 2017
- [7] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, "Recent Advances in Convolutional Neural Networks", *Computer Vision and Pattern Recognition*, 2015, <https://arxiv.org/abs/1512.07108/>.
- [8] D. H. Hubel, T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", *J. Physiol.*, pp. 106-154, Great Britain, 1962.
- [9] D. Britz, "Understanding convolutional neural networks for NLP", <http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/>, Feb, 2017.
- [10] A. Saxena, "Convolutional Neural Networks (CNNs): An Illustrated Explanation", <http://xrds.acm.org/blog/2016/06/convolutional-neural-networks-cnns-illustrated-explanation/>, Feb, 2017.
- [11] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-Based Learning Applied to Document Recognition", *Proc. of the IEEE*, vol. 86, no. 11, pp. 2278 – 2324, Nov, 1998.
- [12] Y. LeCun, "LeNet-5, convolutional neural networks", <http://yann.lecun.com/exdb/lenet/>, Feb, 2017.
- [13] J. Johnson, A. Karpathy, "Convolutional Neural Networks", Stanford Computer Science, <http://cs231n.github.io/convolutional-networks/>, Feb, 2017.
- [14] U. Kam, "An Intuitive Explanation of Convolutional Neural Networks", <https://uijwalkam.me/2016/08/11/intuitive-explanation-convnets/>, Feb, 2017.
- [15] I. Goodfellow, Y. Bengio, A. Courville, "Deep Learning", MIT Press, 2016, <https://www.deeplearningbook.org/>.
- [16] T. Wang, D. J. Wu, A. Coates, "End-to-end text recognition with convolutional neural networks", *Pattern Recognition*, 2012, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.252.8930&rep=rep1&type=pdf/>.
- [17] Y. Boureau, J. Ponce, Y. LeCun, "A Theoretical Analysis of Feature Pooling in Visual Recognition", *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, Nov, 2010.
- [18] D. Yu, H. Wang, P. Chen, Z. Wei, "Mixed Pooling for Convolutional Neural Networks", *The 9th International Conference on Rough Sets and Knowledge Technology*, vol. 8818, pp. 364-375, Shanghai, China, Oct, 2014.
- [19] J. Bruna, A. Szlam, Y. LeCun, "Signal recovery from Pooling Representations", *Machine Learning*, 2013, <https://arxiv.org/abs/1311.4025/>.
- [20] K. He, X. Zhang, S. Ren, J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition", *Computer Vision and Pattern Recognition*, 2014, <https://arxiv.org/abs/1406.4729/>.
- [21] M. D. Zeiler, R. Fergus, "Stochastic Pooling for Regularization of Deep Convolutional Neural Networks", *Machine Learning*, 2013, <https://arxiv.org/abs/1301.3557/>.
- [22] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge", *Computer Vision and Pattern Recognition*, 2014, <https://arxiv.org/abs/1409.0575/>.

- [23] Stanford University, Princeton University, "About ImageNet", <http://www.image-net.org/about-overview/>, Mar, 2017.
- [24] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks", Advances in Neural Information Processing Systems, 2012.
- [25] Stanford Vision Lab, "ILSVRC Challenge", <http://image-net.org/challenges/LSVRC/>, Mar, 2017.
- [26] Google Scholar, "Citation indices of Alex Krizhevsky", <https://scholar.google.com/citations?user=xegzhJcAAAAJ&hl=en/>, Mar, 2017.
- [27] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors", Computer Vision and Pattern Recognition, 2012, <https://arxiv.org/abs/1207.0580/>.
- [28] M. D. Zeiler, R. Fergus, "Visualizing and Understanding Convolutional Networks", Computer Vision and Pattern Recognition, 2013, <https://arxiv.org/abs/1311.2901/>.
- [29] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", Computer Vision and Pattern Recognition, 2014, <https://arxiv.org/abs/1409.1556/>.
- [30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going Deeper with Convolutions", Computer Vision and Pattern Recognition, 2014, <https://arxiv.org/abs/1409.4842/>.
- [31] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", Computer Vision and Pattern Recognition, 2015, <https://arxiv.org/abs/1512.03385/>.
- [32] J. Johnson, A. Karpathy, "Transfer Learning", Stanford Computer Science, <http://cs231n.github.io/transfer-learning/>, Mar, 2017.
- [33] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, "How transferable are features in deep neural networks?", Machine Learning, 2014, <https://arxiv.org/abs/1411.1792/>.
- [34] S. J. Pan, Q. Yang, "A Survey on Transfer Learning," IEEE Trans. on Knowledge and Data Engineering, vol. 22, no. 10, pp. 1345 - 1359, Oct, 2010.
- [35] C. Gulcehre, "Software Links", [http://deeplearning.net/software_links /](http://deeplearning.net/software_links/), Mar, 2017.

ABSTRACT

In this paper the concepts and architecture of convolutional neural networks, whose application is widespread, are described. The importance of application of these types of neural networks is significant, and consequently, research in this area is of special interest. Considerably important results have already been achieved, and therefore efforts for systematization in the area are justified. In this paper important research results in the area are presented, consistent terminology is proposed, and it is pointed to challenges for further research.

Deep Convolutional Neural Networks – Concepts and Current Research

Marko M. Dabović, Igor Tartalja