

Prepoznavanje saobraćajnih znakova koristeći CNN

Seminarski rad u okviru kursa
Računarska inteligencija
Matematički fakultet

Jana Jovičić ???/2015, Jovana Pejkić 435/2016
jana.jovicic755@gmail.com, jov4ana@gmail.com

16. maj 2019.

Sažetak

Sadržaj

| | | |
|---|------------------------------|----|
| 1 | Uvod | 3 |
| 2 | Neuronske mreže | 3 |
| 3 | Konvolutivne neuronske mreže | 4 |
| 4 | Sloj agregacije | 8 |
| 5 | Zaključak | 10 |
| | Literatura | 10 |
| A | Dodatak | 10 |

1 Uvod

2 Neuronske mreže

U masinskom ucenju, neuronske mreze (eng. neural networks) predstavljaju jednu od najprimenjenijih metoda. Neuronske mreže predstavljaju skup statističkih modela učenja inspirisana biološkim neuronima, za rešavanje klasifikacijskih i regresijskih problema. Njihove primene su mnogobrojne, a neke od njih su kategorizacija teksta, medicinska dijagnostika, prepoznavanje objekata na slikama, autonomna voznja, igranje igara poput igara na tabli (tavla i go) ili video igara, masinsko prevodenje prirodnih jezika, modelovanje semantike reci prirodnog jezika i slicno. Problemi zahvaljujuci kojima su se neuronske mreze proslavile su velika kolicina podataka i učenje na osnovu sirove reprezentacije podataka. Male kolicine podataka u slucaju neuronskih mreza lako vode preprilagodavanju, a učenje nad do sada diskutovanim vektorskim reprezentacijama podataka ne koristi u dovoljnoj meri kljucnu prednost neuronskih mreza – da same konstruisu nove atribute nad sirovom reprezentacijom podataka.

¹

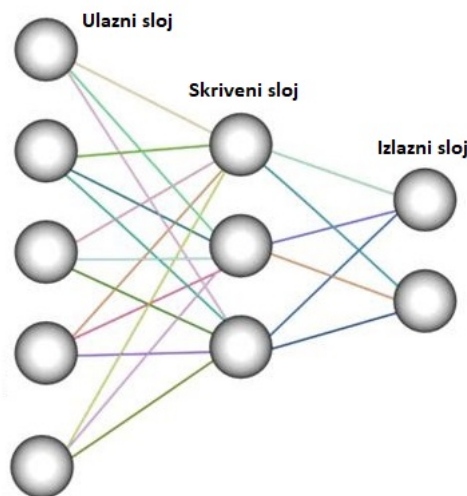
2.1 Arhitektura

Osnovnu jedinicu gradje neuronske mreze predstavljaju neuroni, koji su medjusobno povezani vezama sa tezinama koje se podesavaju tokom učenja mreže. Povezani neuroni prosledjuju signale jedni drugima, a organizovani su u slojeve. Najjednostavniji oblik neuronske mreze je perceptron, koji sadrzi samo jedan ulazni i jedan izlazni sloj. Medjutim, kako on služi samo za učenje linearnih modela, a u praksi se javlja potreba da i slozeniji modeli mogu da se nauce, osim perceptrona, koriste se viseslojne neuronske mreze. Viseslojna neuronska mreza osim ulaznog i izlaznog sloja, dodatno ima jedan ili vise skrivenih slojeva. Broj skrivenih slojeva je proizvoljan, dok sama struktura skrivenog sloja utiče na performanse mreže. Kako se neuronske mreze uglavnom koriste za klasifikaciju uzorka u razlicite kategorije, ulazni sloj se sastoji od onoliko neurona kolika je dimenzionalnost ulaznog prostora, a broj neurona na izlazu jednak je broju klasa. Samo učenje neuronske mreže je zapravo podešavanje težina sve dok se ne dobije zadovoljavajuća aproksimacija između ulaznih i izlaznih veličina.

2.2 Prepoznavanje

Jednostavni zadaci prepoznavanja mogu se dosta dobro rešiti skupovima podataka malih veličina, na primer desetine hiljada slika. Međutim, objekti u realističnim postavkama pokazuju značajnu varijabilnost, stoga, da bi bilo moguće naučiti prepoznati ih, potrebno je koristiti mnogo veće skupove za treniranje. Da bismo naučili o hiljadama objekata iz miliona slika, potreban nam je model sa velikim kapacitetom učenja. Konvolutivne neuronske mreže (CNN), koje ce biti detaljno obradjene u ostatku rada, čine jednu takvu klasu modela. Njihov kapacitet se moze kontrolisati variranjem njihove dubine i širine, a oni takođe daju jake i uglav-

¹Iako domenski eksperti nekad mogu pretpostaviti koji su atributi najinformativniji za predviđanje ciljne promenljive, njihovi izbori nekada mogu biti i pogresni, a neretko losiji od onoga sto bi algoritam ucenja mogao da detektuje u sirovoj informaciji ako bi bio primenljiv na nju.



Slika 1: Višeslojna neuronska mreža sa jednim skrivenim slojem

nom ispravne pretpostavke o prirodi slika (naime, stacionarnost statistike i lokalitet zavisnosti piksela). Tako, u poređenju sa standardnim (feed-forward) neuronskim mrežama sa slojevima sličnih veličina, CNN imaju mnogo manje veza i parametara, tako da ih je lakše trenirati, dok je njihov teoretski-najbolji učinak verovatno samo nešto lošiji.

3 Konvolutivne neuronske mreže

Konvolutivne neuronske mreže predstavljaju podklasu neuronskih mreža koja ima najmanje jedan konvolutivni sloj (a može ih imati i više). Ova vrsta neuronskih mreža je inspirisana vizuelnim korteksom. Svaki put kada nešto vidimo, aktivira se niz slojeva neurona, i svaki sloj otkriva skup karakteristika kao što su linije, ivice. Visi nivoi slojeva otkrivaju složenije karakteristike kako bi prepoznali ono što smo videli. Konvolutivne mreže rade po istom principu i praktično su uvek duboke neuronske mreže, upravo zbog toga što je potrebno od sitnijih detalja, poput uspravnih, kosih i horizontalnih linija, koji obično bivaju detektovani u nižim slojevima mreže, iskonstruisati složenije oblike poput delova lica. Konvolutivne neuronske mreže se koriste u obradi signala (slike, zvuka), ali i teksta. U odnosu na ostale vrste neuronskih mreža, ističu se u prikupljanju lokalnih informacija (na primer o susednim pikselima na slici ili surrounding recima u tekstu) i smanjenju složenosti modela (faster training, needs fewer samples, reduces the chance of overfitting). Konvolutivne neuronske mreže se zasnivaju na sposobnosti mreža da iz sirovog signala konstruisu attribute. Nazivaju se konvolutivnim zato što uče filtere, čijom konvolutivnom primenom detektuju određena svojstva signala.

Unutrasnja struktura konvolutivne mreže se sastoji od nekoliko naimeničkih konvolutivnih slojeva (eng. convolution layer) i slojeva agregacije (eng. pooling layer), pri čemu je dozvoljeno pojavljivanje iste vrste sloja više puta. U dosad opisanoj strukturi neurona (neuronskih mreža) izlaz iz svakog od njih je bio skalarna veličina. Izlazi konvolutivnog sloja

su dvodimenzionalni i nazivaju se mapama karakteristika (engl. feature maps). Izlazi konvolutivnog sloja se transformisu nelinearnom aktivacionom funkcijom. Na izlaze poslednjeg od tih slojeva se obicno nadovezuje potpuno povezana mreza, koja uci nad atributima koje prethodni slojevi konstruisu (nekoliko potpuno povezanih slojeva koji su jednodimenzionalni).



Slika 2: Arhitektura konvolutivne neuronske mreže

3.1 Konvolucija slika preko CNN

Konvolutivne mreze primenjuju mnogo filtera nad jednom (jedinom) slikom, gde svaki prikuplja različite informacije. Na pocetnom sloju, mozete ih zamisliti kako prolazi filter za horizontalnu liniju, filter za vertikalnu liniju, i filter za dijagonalnu liniju da bi kreirali mapu ivica slike.

Konvolutivne mreze uzimaju ove filtere, isecke prostora karakteristika slika, i mapiraju ih jednog po jednog; tj. oni kreiraju mapu gde beleže svako mesto gde se ta karakteristika javlja. Učeći razlicite delove prostora karakteristika, konvolutivne mreze dopustaju robusni inženjering karakteristika. Konvolutivne mreze uče slike iz delova koje zovemo mape karakteristika.

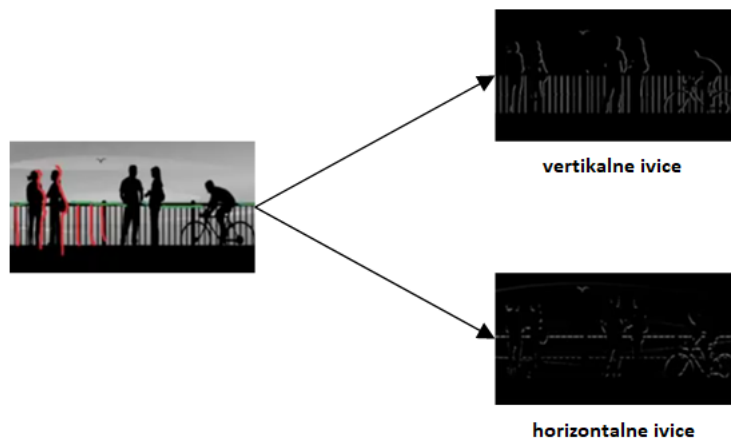
Konvolutivne mreze obavljaju neku vrstu pretrage. Zamislite malo uvelicavajuće staklo koje klizi s leva na desno preko vece slike, i nastavlja s leve strane kada dodje do kraja jednog prelaza (kao kod pisace masine). Taj pokretni (klizni) prozor moze da prepozna samo jednu stvar, recimo kratku vertikalnu liniju. Tri tamna piksela naslagana jedan na drugi?! Pomeri taj filter za prepoznavanje (koji prepoznaje) vertikalnu liniju preko stvarnih piksela slike, trazeci podudaranja.

Svaki put kada se pronadje poklapanje (dodje do poklapanja), ono se mapira u (na) prostor sa karakteristikama koji je specifican za taj vizuelni element. U tom prostoru se cuva (belezi) lokacija svakog poklapanja sa vertikalnom linijom. Konvolutivna mreza pokrece mnogo, mnogo pretraga nad jednom slikom - horizontalne linije, dijagonalne linije, nad onoliko elemenata koliko se trazi (onoliko koliko ima vizuelnih elemenata koji trebaju da budu pretrazeni).

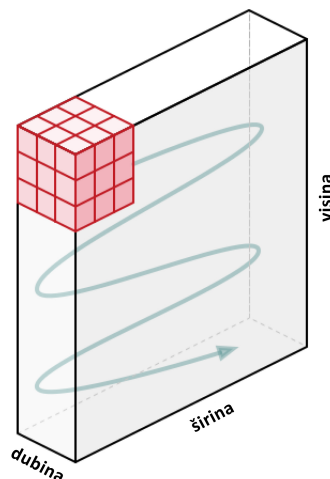
Nakon konvolutivnog sloja, ulaz se salje kroz nelinearnu transformaciju kao sto je tanh ili preciscena linearna jedinica?!, koja ce prevesti ulaz u opseg izmedju -1 i 1 (which will squash input values into a range between -1 and 1).

3.2 Konvolutivni sloj

Konvolutivni sloj je glavni deo konvolutivne neuronske mreze. On radi najvise izracunavanja u mrezi, a ima ulogu konstrukcije novih atributa. To



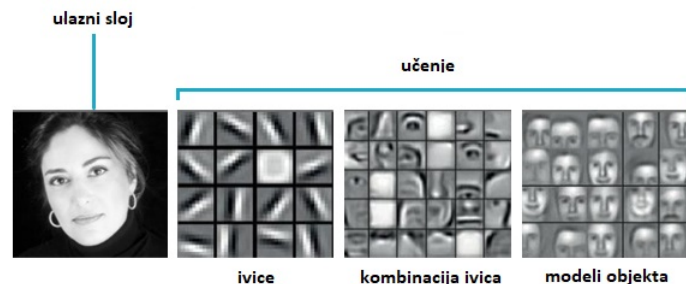
Slika 3: Detektovanje ivica



Slika 4: Kretanje filtera

je prvi sloj koji ekstrahuje karakteristike iz ulazne slike. Konvolucija je matematička operacija koja uzima dva ulaza - dve matrice f i g dimenzija $m \times n$ i $p \times q$. Definisana je na sledeći način: $(f * g)_{ij} = \sum_{k=0}^{Xp-1} \sum_{l=0}^{Yq-1} f_{i-k, j-l} g_{k, l}$

Matrica f je obično ulaz, poput slike, dok je matrica g filter (ili jezgro) – pomoćna matrica koja izdvaja neku vrstu informacije iz ulaza. Određena ulazna reprezentacija podatka (originalna slika) konvolvira se s konvolucijskom jezgrom (filterom) s određenim parametrima. Procesom učenja ovi parametri (težine koje je potrebno naučiti kako bi mreža davala dobre rezultate) se podešavaju. Filteri su najčešće manjih prostornih dimenzija od ulaza, ali uvek su jednake dubine kao i ulaz. Konvolucijska jezgra (filteri) filtriraju sliku, tj. mapu karakteristika kako bi izlučila neku korisnu informaciju kao što je recimo određeni oblik, boja ili ivica. U pr-



Slika 5: Učenje mreže

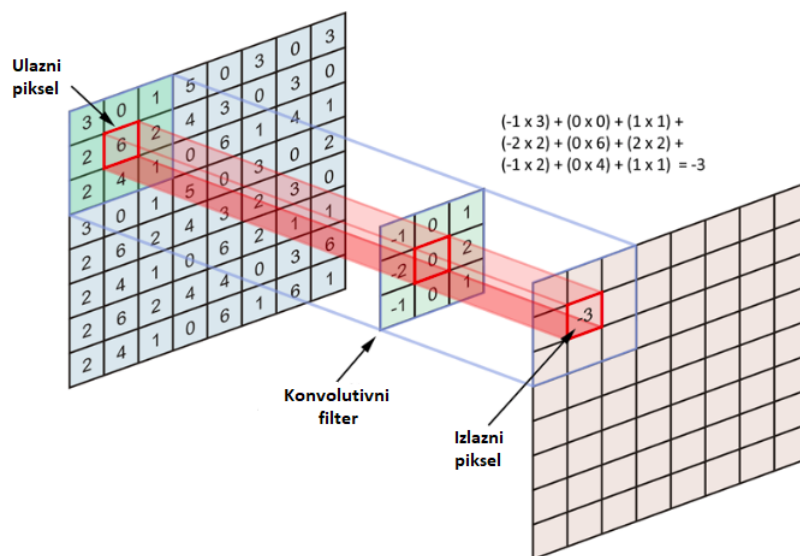
vih nekoliko slojeva izlučuju se osnovni oblici (npr. ivice). Jednostavan primer je filter za detekciju uspravnih ivica, horizontalnih ivica, dijagonalnih (kosih) ivica. Postojanje nekih objekata na slici se može utvrditi na osnovu specifičnog rasporeda ovih (uspravnih, kosih i horizontalnih) linija (ivica), koje detektuje prethodni konvolutivni sloj. Dalje, posmatra se kombinacija svih ovih ivica i tako prepoznaju i drugi oblici. Svaki konvolutivni sloj raspolaže nizom parametrizovanih filtera koji vrše ove poslove. Upravo, parametri filtera predstavljaju parametre konvolutivne mreže i bivaју nauci u procesu obucavanja mreže.

Tokom prvog prolaza svaki filter se pomera po širini i visini i računa se skalarni proizvod ulaza i vrednosti filtera. Izlaz jednog filtera će biti dvodimenzioni niz. Ako imamo više filtera, izlaz iz konvolucionog sloja će biti rezultati svakog filtera poredjanih po dubini. Obično se koristi više konvolutivnih slojeva koji svojom kombinacijom mogu izučiti apstraktnije nivoe karakteristika karakteristične za pojedini objekt (npr. ljudsko lice). U zavisnosti od tipa filtera (jezgra) moguće je detektovati razne karakteristike (ivice, tačke). Vrednosti parametara jezgra se podešavaju procesom učenja tj. konvolucijske neuronske mreže na taj način „nauče“ prepoznavati određene karakteristike. Operacijom konvolucije dobija se transformirana slika $I * K$. Dobijena transformacija je lokalna tj. pikseli izlaza zavise o lokalnom susedstvu piksela ulaza.

3.3 Prosirivanje

Formula konvolucije koja je data iznad nije definisana za sve indekse $i=0, m \text{ minus } 1$ i $j=0, n \text{ minus } 1$. Na primer, ako je $i, j = 0$ i $k, l \notin 0$, vrednost f minus k, i minus l nije definisana. Ukoliko bi se u obzir uzele samo definisane vrednosti, dimenzija konvolucije bi bila manja od dimenzije matrice f .

Medjutim, to nije uvek poželjno, i može se izbesci tako sto se vrši prosirivanje (eng. padding) matrice f , na primer nulama ili vrednostima koje su vec na obodu, tako da velicina rezultujuce matrice bude jednaka velicini matrice f pre prosirivanja. Ovo je prikazano na slici ??????. Takode, prilikom racunanja konvolucije, filter se duz slike ne mora pomerati za jedan piksel, vec za neki veci korak (eng. stride).



Slika 6: Konvolutivni sloj, pomeranje filtera

4 Sloj agregacije

Uloga sloja za agregaciju je smanjenje broja parametara kada su slike prevelike, kao i smanjenje broja racunskih operacija u visim slojevima smanjuje. Agregacija smanjuje dimenzionalnost svake mape, ali zadržava važne informacije. Sve to rezultuje smanjenjem racunske zahtevnosti pri optimizaciji i pomaze u kontroli overfitinga. Zato zelimo da slojevi za agregaciju prate konvolucione slojeve kako bismo postepeno smanjili prostornu veličinu (širinu i visinu) prikaza podataka.

Cesto nas konačni zadatak postavlja neko globalno pitanje o slici, npr., Da li sadrži mačku? Tako da čvorovi našeg zadnjeg sloja moraju biti osetljivi na celi ulaz. Postepenom agregacijom informacija, stvarajući grublje i grublje mape karakteristika, postize se taj cilj da se na kraju nauči globalna reprezentacija, zadržavajući sve prednosti konvolucijskih slojeva na srednjim slojevima obrade.

Sloj agregacije ukupnjuje informacije, tako sto racuna neku jednostavnu funkciju agregacije susednih jedinica prethodnog sloja, poput maksimuma (eng. Max pooling) [koji vraca maksimalnu vrednost dela slike pokrivene filterom] ili proseka (eng. Average pooling) [koji vraca prosečnu vrednost dela slike pokrivene filterom]. Ukoliko agregira, na primer, 3 x 3 piskela, onda je broj izlaza ovog sloja 9 puta manji od broja izlaza prethodnog. Kada se racuna maksimum, dolazi do zanemarivanja informacije o tome gde je precizno neko svojstvo (poput uspravne linije) pronadeno, ali se ne gubi informacija da je pronadeno. Ovakva vrsta zanemarivanja informacije cesto ne steti cilju koji treba postici. Na primer, ako su na slici pronadeni kljun i krila, informacija o tacnoj poziciji najverovatnije nije bitna za odlucivanje da li se na slici nalazi ptica. Ipak, ukoliko je potrebno napraviti mrežu koja igra igru u kojoj su pozicije objekata na ekranu bitne, nije poželjno koristiti agregaciju.

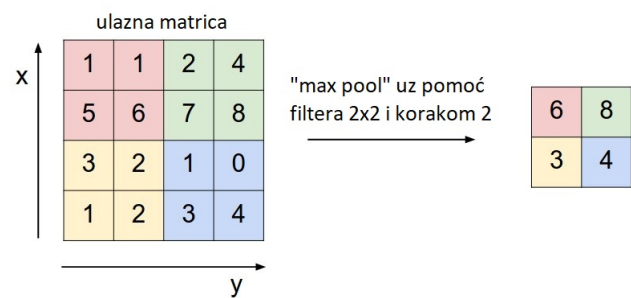
Average pooling was often used historically but has recently fallen out of favor compared to the max pooling operation, which has been shown

to work better in practice.

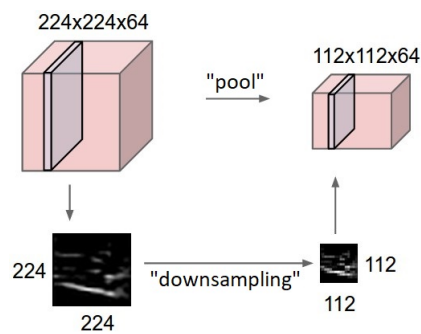
Pooling layer downsamples the volume spatially, independently in each depth slice of the input volume.

Prva slika (pool): In this example, the input volume of size $[224 \times 224 \times 64]$ is pooled with filter size 2, stride 2 into output volume of size $[112 \times 112 \times 64]$. Notice that the volume depth is preserved.

Druga slika (maxpool): The most common downsampling operation is max, giving rise to max pooling, here shown with a stride of 2. That is, each max is taken over 4 numbers (little 2×2 square).



Slika 7: Maxpool



Slika 8: Pool

Tabela 1: Primer tabele

| | | |
|-------|---------|-------------|
| tekst | tekst | tekst tekst |
| tekst | ”tekst” | tekst |

4.1 Podnaslov 6

Kod 1 demonstrira...

```
table[key] = value
```

Kod 1: Primer koda

Podsekcija 1

PodPodsekcija 1

4.2 Podnaslov 7

4.3 Podnaslov 8

5 Zaključak

A Dodatak

A.1 Podnaslov 1

Dodatna objasnjenja

By visualizing the output from different convolution layers in this manner, the most crucial thing that you will notice is that the layers that are deeper in the network visualize more training data specific features, while the earlier layers tend to visualize general patterns like edges, texture, background etc.

This knowledge is very important when you use Transfer Learning whereby you train some part of a pre-trained network (pre-trained on a different dataset, like ImageNet in this case) on a completely different dataset. The general idea is to freeze the weights of earlier layers, because they will anyways learn the general features, and to only train the weights of deeper layers because these are the layers which are actually recognizing your objects.

Convolved Feature, Activation Map or Feature Map is the output volume formed by sliding the filter over the image and computing the dot product. Perceptrons come first in 1950s, and it uses a brittle activation function to do classification, so if $w \cdot x$ is greater than some value it predicts positive, otherwise negative.

Neurons uses a softer activation function by introducing a sigmoid function, a tanh function or other activation functions to pass on values to other neurons in the network.

So perceptrons do not use in a network setting, they do classification on their own, hence they can't classify XOR, however neurons can because they all contribute forward to the final output, using more complicated structure(i.e. multiple layers in network), they are able to classify XOR and other complicated problems.

A CNN, in specific, has one or more layers of convolution units. A convolution unit receives its input from multiple units from the previous layer which together create a proximity. Therefore, the input units (that form a small neighborhood) share their weights.

The convolution units (as well as pooling units) are especially beneficial as:

They reduce the number of units in the network (since they are many-to-one mappings). This means, there are fewer parameters to learn which reduces the chance of overfitting as the model would be less complex than a fully connected network. They consider the context/shared information in the small neighborhoods. This feature is very important in many applications such as image, video, text, and speech processing/mining as the neighboring inputs (eg pixels, frames, words, etc) usually carry related information.