

# Prepoznavanje saobraćajnih znakova koristeći CNN

Seminarski rad u okviru kursa  
Računarska inteligencija  
Matematički fakultet

Jana Jovičić 215/2015, Jovana Pejkić 435/2016  
jana.jovicic755@gmail.com, jov4ana@gmail.com

16. maj 2019.

**Sažetak**

# Sadržaj

1	Uvod	3
2	Neuronske mreže	4
3	Konvolutivne neuronske mreže	5
4	Implementacija i eksperimentalni rezultati	11
5	Zaključak	12
	Literatura	13
A	Dodatak	13

# 1 Uvod

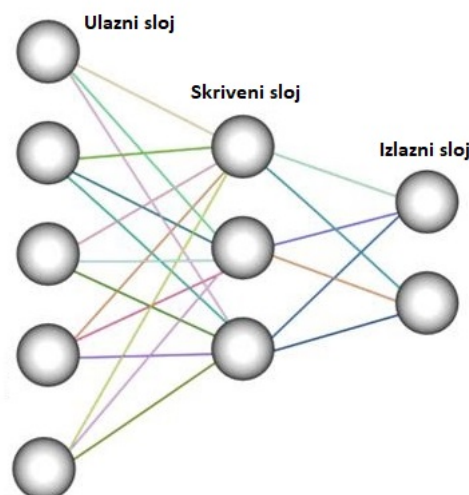
[illegible]

## 2 Neuronske mreže

Neuronske mreže predstavljaju skup statističkih modela učenja inspirisana biološkim neuronima, za rešavanje klasifikacionih<sup>1</sup> i regresijskih problema<sup>2</sup>. Njihove primene su mnogobrojne, a neke od njih su: kategorizacija teksta, raspoznavanje i sinteza govora, autonomna voznja, igranje video igara, masinsko prevodenje prirodnih jezika i slicno. Ključna prednost neuronskih mreža je da same mogu da konstruisu nove attribute nad sirovom reprezentacijom podataka<sup>3</sup> i odlično baratanje sa velikim količinama podataka.

### 2.1 Arhitektura

Osnovnu jedinicu gradje neuronske mreže 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 mreže je perceptron, koji sadrži samo jedan ulazni i jedan izlazni sloj. Medjutim, kako on služi samo za učenje linearnih modela, a u praksi se javlja potreba da i složeniji modeli mogu da se nauče, osim perceptrona, koriste se višeslojne neuronske mreže. Višeslojna neuronska mreža osim ulaznog i izlaznog sloja, ima jedan ili više skrivenih slojeva (slika 1). Kako se neuronske mreže uglavnom koriste za klasifikaciju uzorka u različite 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 podesavanje težina sve dok se ne dobije zadovoljavajuća aproksimacija između ulaznih i izlaznih veličina.



Slika 1: Višeslojna neuronska mreža sa jednim skrivenim slojem

<sup>1</sup>Klasifikacioni problem - ako je izlazna promenljiva kategorickog tipa, npr. „zdrav” i „bolestan”.

<sup>2</sup>Regresioni problem - ako je izlazna promenljiva neprekidnog tipa, npr. „plata” ili „težina”.

<sup>3</sup>Iako se nekad mogu pretpostaviti koji su atributi najinformativniji za predviđanje ciljne promenljive, izbori tih atributa su neretko losiji od onoga sto bi algoritam učenja mogao da detektuje u sirovoj informaciji.

## 2.2 Razlog uvođenja CNN

Jednostavni zadaci prepoznavanja mogu se dosta dobro resiti skupovima podataka malih velicina, 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 će biti detaljno obrađene u ostatku rada, čine jednu takvu klasu modela. Njihov kapacitet se može kontrolisati variranjem njihove dubine i širine, a oni takođe daju jake i uglavnom ispravne pretpostavke o prirodi slika. Tako, u poredjenju sa standardnim (feedforward) neuronskim mrežama sa slojevima sličnih velicina, 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 (eng. Convolutional neural network, CNN) 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 itd. 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, konstruisati 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 „okružujućim“ (eng. surrounding) rečima u tekstu) i smanjenju složenosti modela (brže treniranje, potrebno je manje izoraka, manja šansa da dođe do preprilagođavanja (eng. overfitting)). Konvolutivne neuronske mreže se zasnivaju na sposobnosti mreža da iz sirovog signala konstruišu atribut. Nazivaju se konvolutivnim zato što uče **filtere** (pojam objašnjen u tabeli 1), čijom konvolutivnom primenom detektuju određena svojstva signala. U narednoj tabeli su opisani neki od bitnijih parametara koje treba podesiti pri učenju mreže.

tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst  
tekst tekst tekst tekst tekst tekst

### 3.1 Parametri

U ovoj sekciji je dat tabelarni prikaz parametara koji su najznacajniji za implementaciju konvolutivne mreze. U tabeli 1 je dat samo kratak opis svakog od njih radi boljeg razumevanja teksta koji sledi. Ipak, u nastavku je svaki detaljnije opisan.

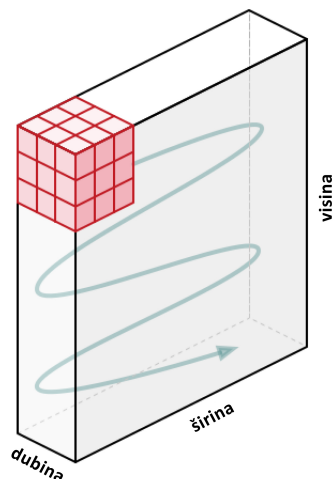
Tabela 1: Primer tabele

Filter (jezgro, kernel)	<ul style="list-style-type: none"><li>- matrica sa tezinama za konvoluciju ulaza</li><li>- daje meru koliko deo ulaza liči na karakteristiku</li><li>- tezine u matricama filtera su izvedene za vreme treniranja podataka</li></ul>
Padding	<ul style="list-style-type: none"><li>- koristi se za dodavanje kolona i redova nula da bi se održala konstantna velicina matrice (mape) nakon konvolucije</li><li>- ovaj parametar može da unapredi performanse tako što zadržava informacije u okvirima</li></ul>
Stride	<ul style="list-style-type: none"><li>- broj piksela koji želite preskočiti, dok prelazite ulaz vodoravno i uspravno tokom konvolucije, nakon množenja svakog elementa iz ulazne matrice težina s onima u filtru</li></ul>
Number of Channels	<ul style="list-style-type: none"><li>- It is the equal to the number of color channels for the input but in later stages is equal to the number of filters we use for the convolution operation.</li><li>- The more the number of channels, more the number of filters used, more are the features learnt, and more is the chances to over-fit and vice-versa.</li></ul>
Pooling-layer Parameters	<ul style="list-style-type: none"><li>- ima iste parametre kao i konvolutivni sloj</li><li>- uglavnom se koristi Max-Pooling opcija</li><li>- The objective is to down-sample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality by keeping the max value(activated features) in the sub-regions binned.</li></ul>

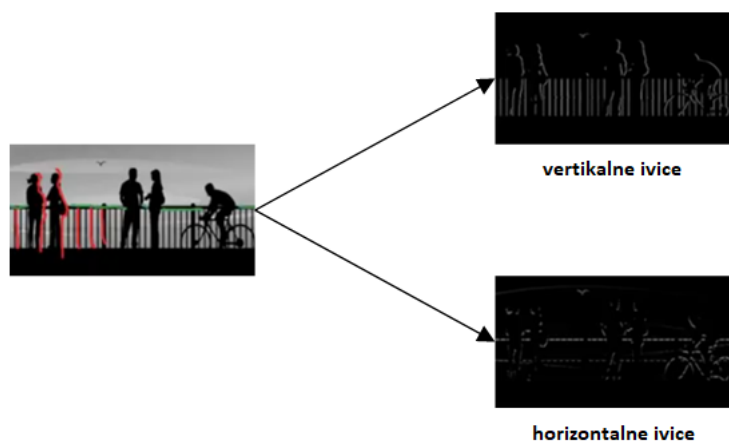
### 3.2 Konvolucija slika preko CNN

Da bi izvršile klasifikaciju slika, konvolutivne mreze (preciznije, konvolutivni sloj, detaljnije u poglavlju ????) obavljaju neku vrstu pretrage. Ovo se može zamisliti kao mali pokretni (ili klizni) prozor (prikazano na

slici 2) koji 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 - koji ustvari predstavlja filter, moze da prepozna samo jednu stvar, recimo kratku vertikalnu liniju (tri tamna piksela naslagana jedan na drugi). Slicno, neki drugi filter moze da služi za prepoznavanje horizontalnih linija, i on se takodje pomera preko piksela slike, trazeci podudaranja. Rezultat koji se postize filterima koji prepoznaju vertikalne i horizontalne linije prikazan je na slici 3.



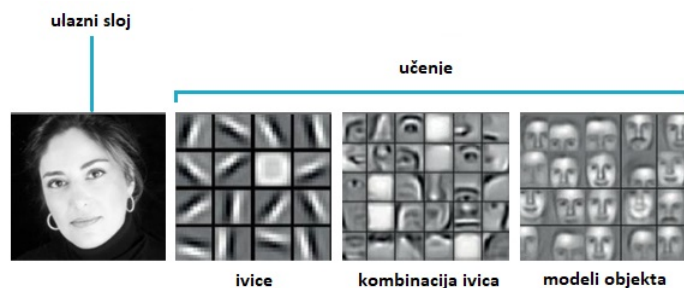
Slika 2: Kretanje filtera



Slika 3: Detektovanje ivica

Svaki put kada dodje do poklapanja (filtera sa ulazom), ono se mapira u prostor sa karakteristikama - koji se zove **mapa karakteristika** (eng. feature maps), koji je specifičan za taj vizuelni element. U tom prostoru (tj. mapi) se cuva (odnosno beleži) lokacija svakog poklapanja

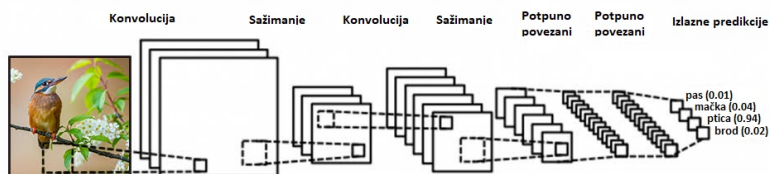
sa vertikalnom linijom. Konvolutivna mreža pokreće mnogo pretraga nad jednom istom slikom. Na početnom sloju mreže koriste se filteri koji prepoznaju horizontalnu liniju, vertikalnu liniju i dijagonalnu liniju, da bi kreirali mapu ivica slike. U narednim koracima (odnosno slojevima) posmatraju se kombinacije ovih ivica i tako prepoznaju složeniji oblici. Ovo je demonstrirano na slici 4.



Slika 4: Učenje mreže

### 3.3 Unutrasnja struktura CNN

Unutrasnja struktura konvolutivne mreže se sastoji od nekoliko naizmeničnih konvolutivnih slojeva (eng. convolution layer) i slojeva agregacije (eng. pooling layer), pri čemu je dozvoljeno pojavljivanje iste vrste sloja više puta (prikazano na slici 5). 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 (eng. feature maps). Oni se transformišu nelinearnom aktivacionom funkcijom (na primer tanh), koja će prevesti ulazne vrednosti u opseg između -1 i 1.



Slika 5: Arhitektura konvolutivne neuronske mreže

### 3.4 Konvolucija

Konvolutivni sloj je glavni deo konvolutivne neuronske mreže koji radi najviše izračunavanja u mreži. Njegova uloga je konstrukcija novih atributa. To je prvi sloj koji ekstrahuje karakteristike iz ulazne slike. Konvolucija je matematička operacija koja uzima dva ulaza - dve matrice  $f$  i



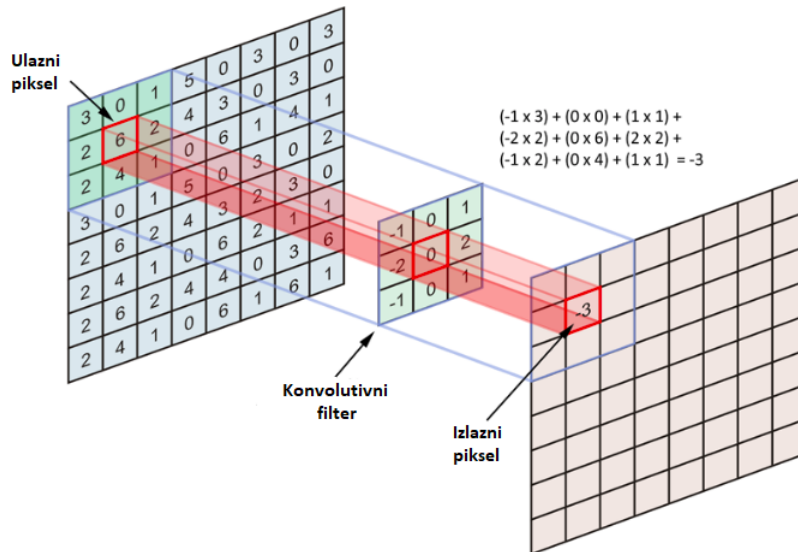
$g$ , dimenzija  $m \times n$  i  $p \times q$ , a definisana je na sledeci nacin:

$$(f * g)_{ij} = \sum_{k=0}^{p-1} \sum_{l=0}^{q-1} f_{i-k, j-l} g_{k,l}$$

Matrica  $f$  je obicno ulaz, poput slike, dok je matrica  $g$  filter. Određena ulazna reprezentacija podatka (originalna slika) konvolvira se sa filterom sa određenim parametrima (ti parametri predstavljaju i parametre konvolutivne mreže). Procesom učenja ovi parametri (težine koje je potrebno nauciti kako bi mreža davala dobre rezultate) se podešavaju i bivaju nauceni. Filteri su najčešće manjih prostornih dimenzija od ulaza, ali uvek su jednake dubine kao i ulaz. Kao što je već receno, konvolucijska jezgra (filteri) filtriraju sliku, tj. mapu karakteristika kako bi izlučila neku korisnu informaciju kao što je recimo određeni oblik, boja ili ivica.

Tokom prvog prolaza svaki filter se pomera po širini i visini i računa se skalarni proizvod ulaza i vrednosti filtera (prikazano u formuli iznad). Izlaz iz konvolutionog sloja će biti dvodimenzioni niz. Ako imamo više filtera, izlaz iz konvolutionog sloja će biti rezultati svakog filtera poredjanih po dubini. Operacijom konvolucije dobija se transformirana slika dimenzije  $I \times K$ , gde je  $I$  dimenzija ulazne matrice slike, a  $K$  je dimenzija filtera koji je primenjen nad tom slikom. Takodje, dobijena transformacija je lokalna tj. pikseli izlaza zavise od lokalnih, susednih piksela ulaza. Ceo ovaj proces je prikazan na slici 6.

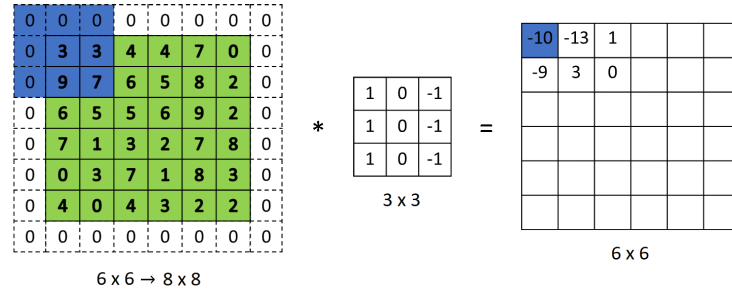
Slika 6: Konvolucija primenom filtera dimenzije  $3 \times 3$  na matricu dimenzije  $8 \times 8$



### 3.4.1 Prosirivanje

Formula konvolucije koja je data u poglavlju 3.4 nije definisana za sve indekse  $i=0, m-1$  i  $j=0, n-1$ . Na primer, ako je  $i, j = 0$  i  $k, l \notin 0$ , vrednost  $f_{i-k, j-l}$  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 izbeći tako što se vrši

**prosirivanje** (eng. padding) matrice  $f$ , na primer nulama ili vrednostima koje su već na obodu, tako da velicina rezultujuće matrice bude jednaka veličini matrice  $f$  pre prosirivanja. Ovo je prikazano na slici 7. Takođe, prilikom računanja konvolucije, filter se duž slike ne mora pomerati za jedan piksel, već za neki veći **korak** (eng. stride).



Slika 7: Primer prosirivanja

### 3.5 Agregacija

Uloga sloja za agregaciju je smanjenje broja parametara kada su slike prevelike, kao i smanjenje broja računskih operacija u visim slojevima smanjuje. Agregacija smanjuje dimenzionalnost svake mape, ali zadržava važne informacije. Sve to rezultuje smanjenjem računске zahtevnosti pri optimizaciji i pomaže u kontroli overfittinga. Zato želimo 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.

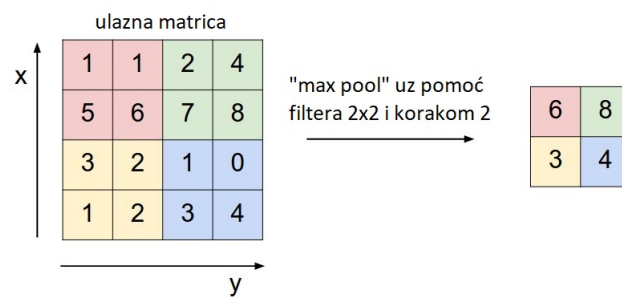
Sloj agregacije ukupnjuje informacije, tako što računa neku jednostavnu funkciju agregacije susednih jedinica prethodnog sloja, poput maksimuma (eng. Max pooling) [koji vraća maksimalnu vrednost dela slike pokriven filterom] ili proseka (eng. Average pooling) [koji vraća prosečnu vrednost dela slike pokriven filterom]. Ukoliko agregira, na primer, 3 x 3 piskela, onda je broj izlaza ovog sloja 9 puta manji od broja izlaza prethodnog. Kada se računa maksimum, dolazi do zanemarivanja informacije o tome gde je precizno neko svojstvo (poput uspravne linije) pronađeno, ali se ne gubi informacija da je pronađeno. Ovakva vrsta zanemarivanja informacije često ne šteti cilju koji treba postići. Na primer, ako su na slici pronađeni kljun i krila, informacija o tačnoj poziciji najverovatnije nije bitna za odlučivanje 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.

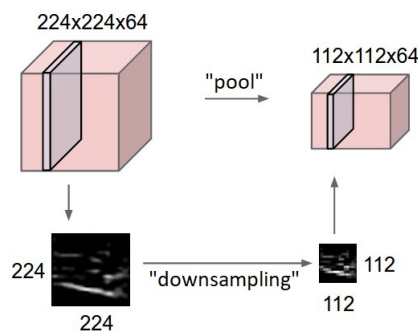
Pooling layer downsamples the volume spatially, independently in each depth slice of the input volume.

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  $2 \times 2$  square).



Slika 8: Maxpool



Slika 9: Pool

## 4 Implementacija i eksperimentalni rezultati

Kod 1 demonstrira...

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

Kod 1: Primer koda

## 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 use 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.