

Neuronske mreže

Konvolutivne neuronske mreže

MASTER AKADEMSKE STUDIJE, ZIMSKI SEMESTAR 2017/2018

FAKULTET TEHNIČKIH NAUKA, NOVI SAD

Konvolutivne neuronske mreže

Viewpoint variation



Scale variation



Deformation



Occlusion



Illumination conditions



Background clutter



Intra-class variation

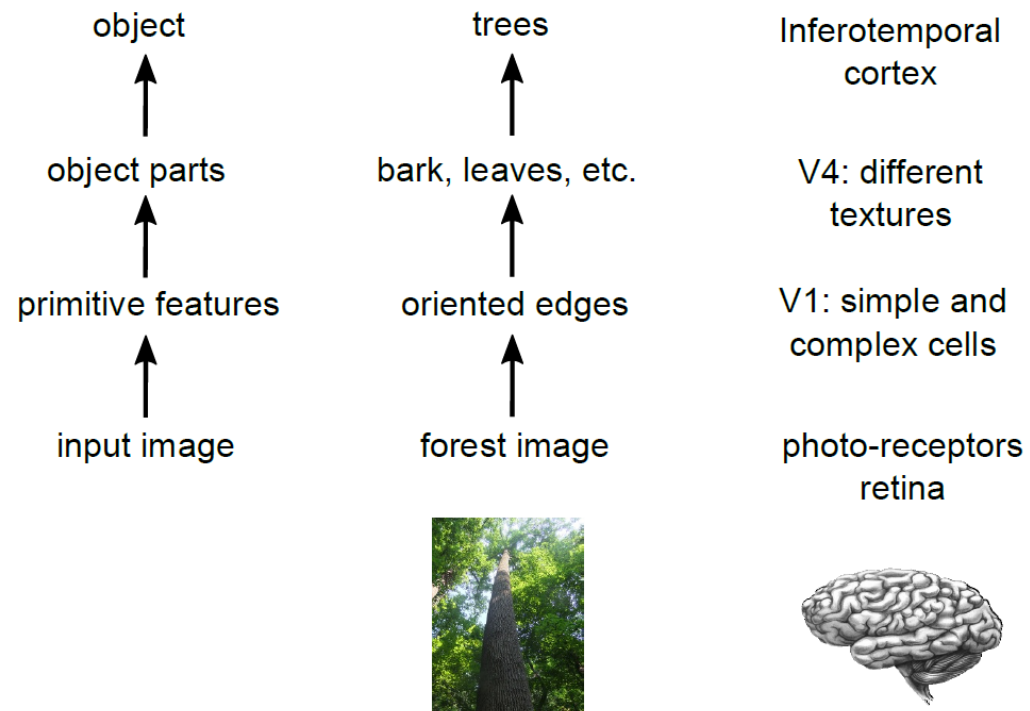


Konvolutivne neuronske mreže

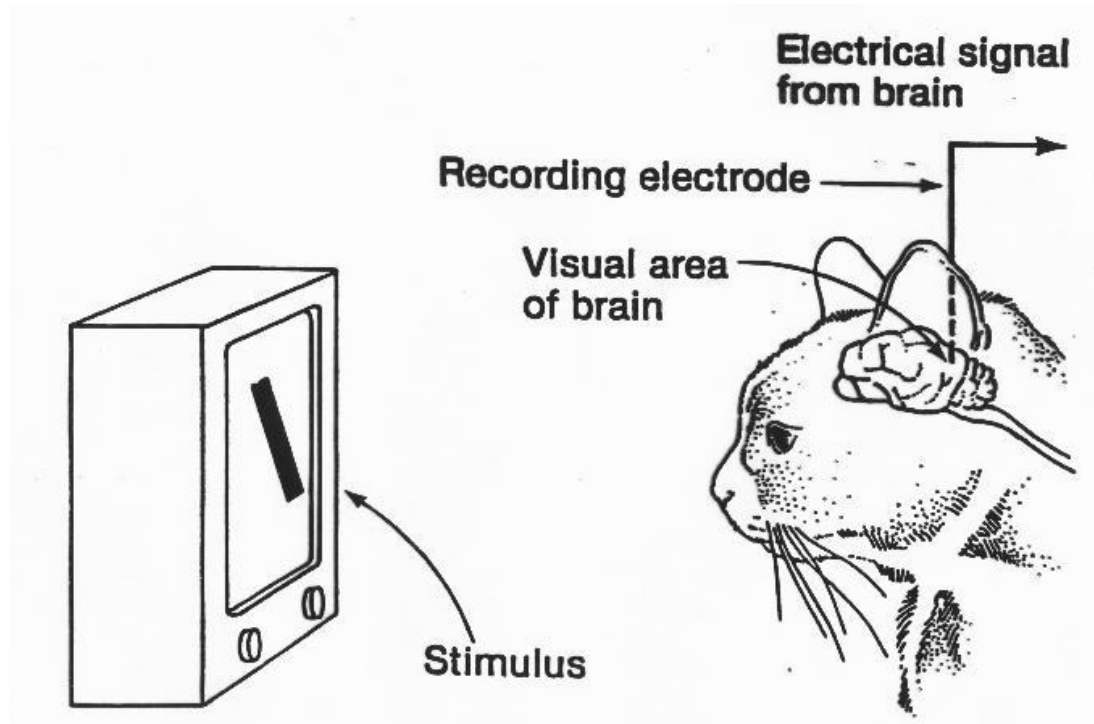
- **Convolutional neural network – CNN, convnet**
- Napraviti dobru internu reprezentaciju vizuelnog sveta u cilju prepoznavanja
 - Detekcija i klasifikacija objekata u kategorije, nezavisno od poze, veličine, rotacije, osvetljenja, zaklonjenosti...
- Da li veštački sistem može da nauči odgovarajuću internu reprezentaciju automatski, na sličan način kao što ljudi uče jednostavnim posmatranjem sveta?
- **PRE:** „ručno“ nameštanje i pravljenje internih reprezentacija
- **SADA:** teži se da mašina sama nauči reprezentaciju iz samih podataka

CNN – inspiracija u biologiji

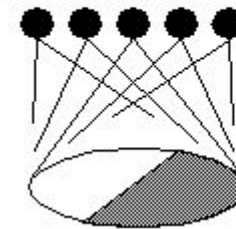
Biološka vizija je hijerarhijski organizovana (duboka hijerahija)



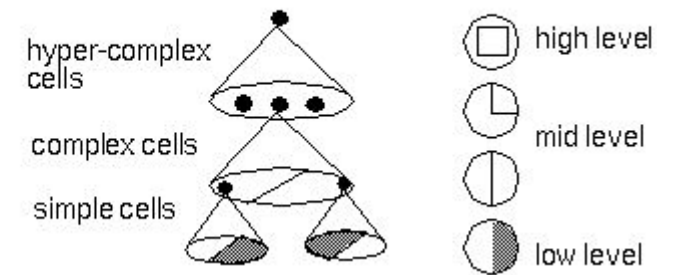
CNN – Hubel & Weisel (1962)



Hubel & Weisel
topographical mapping



featural hierarchy



CNN – going deep...

- Plitke arhitekture su neefikasne u reprezentovanju „dubokih“ funkcija
- Dvoslojne neuronske mreže sa dovoljno velikim brojem neurona u skrivenom sloju mogu implementirati bilo koju funkciju - „**univerzalni aproksimator**“
- Ali, ako je funkcija „duboka“, potreban je izuzetno velik skriveni sloj – **ogroman** broj parametara (težina) koji se moraju obučiti – računski neefikasno
- Ok, dakle koristićemo duboke neuronske mreže...

Ali zašto ne obične duboke neuronske mreže?

Koliko parametara ima ima NM?

$$|\theta| = 3D^2 + D$$

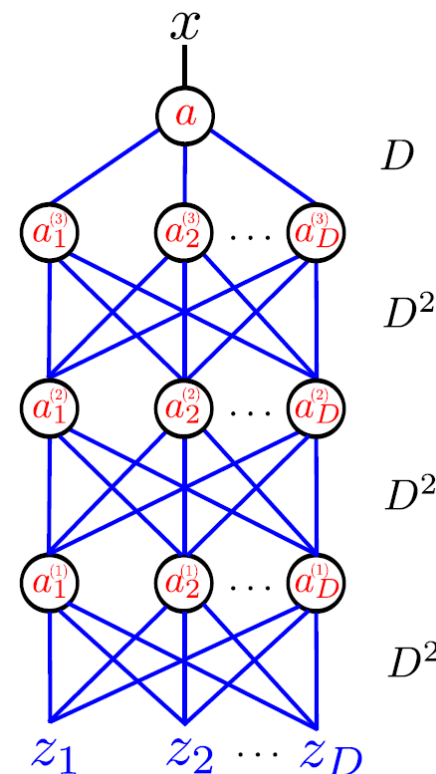
Što je za malu sliku 32x32:

$$|\theta| = 3 \times 32^4 + 32^2 \approx 3 \times 10^6$$

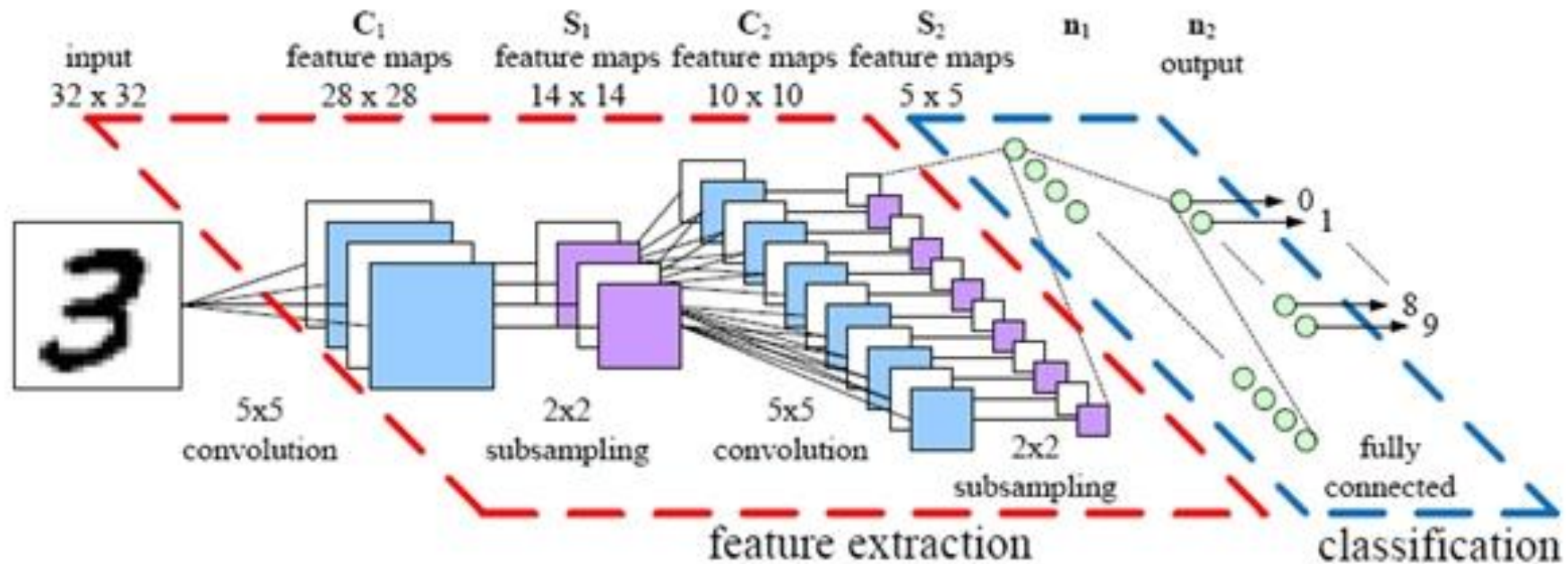
Teško obučavanje

- Over-fitting, under-fitting, lokalni optimumi

Konvolutivne mreže smanjuju broj parametara

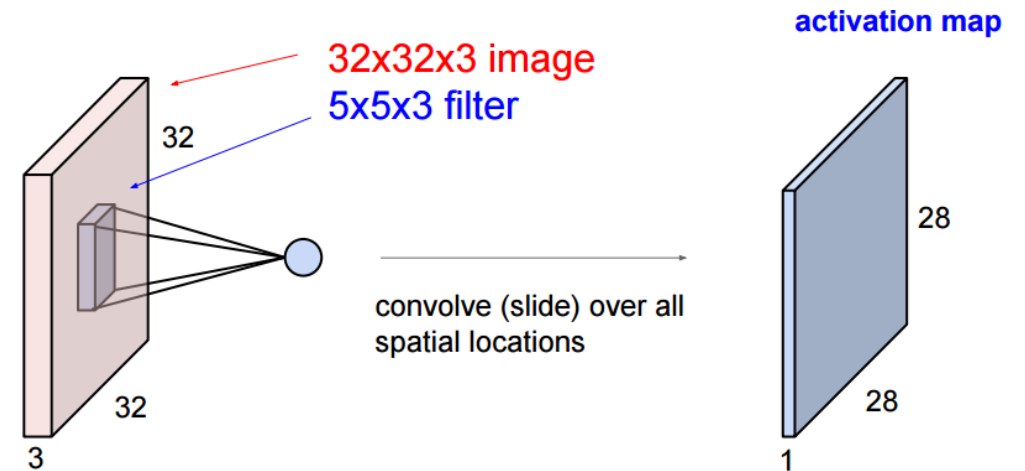
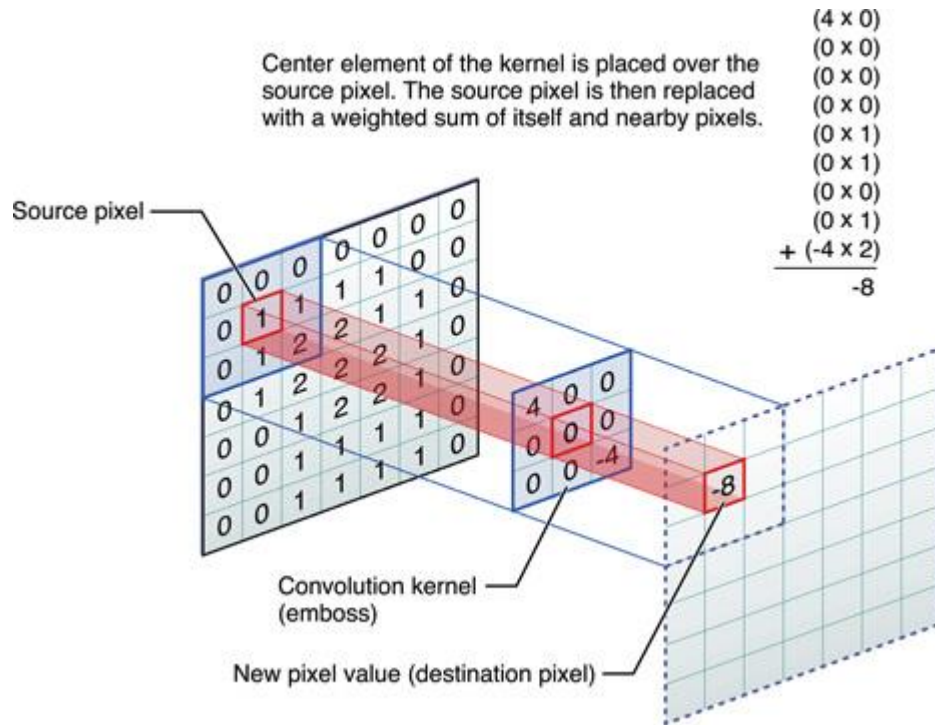


CNN – the big picture



CNN – konvolutivni sloj

- Operator konvolucije



CNN – konvolutivni sloj

- **Padding**

In practice: Common to zero pad the border

0	0	0	0	0	0				
0									
0									
0									
0									

(in each channel)

e.g. input 7x7
neuron with receptive field 3x3, stride 1
pad with 1 pixel border => what is the output?

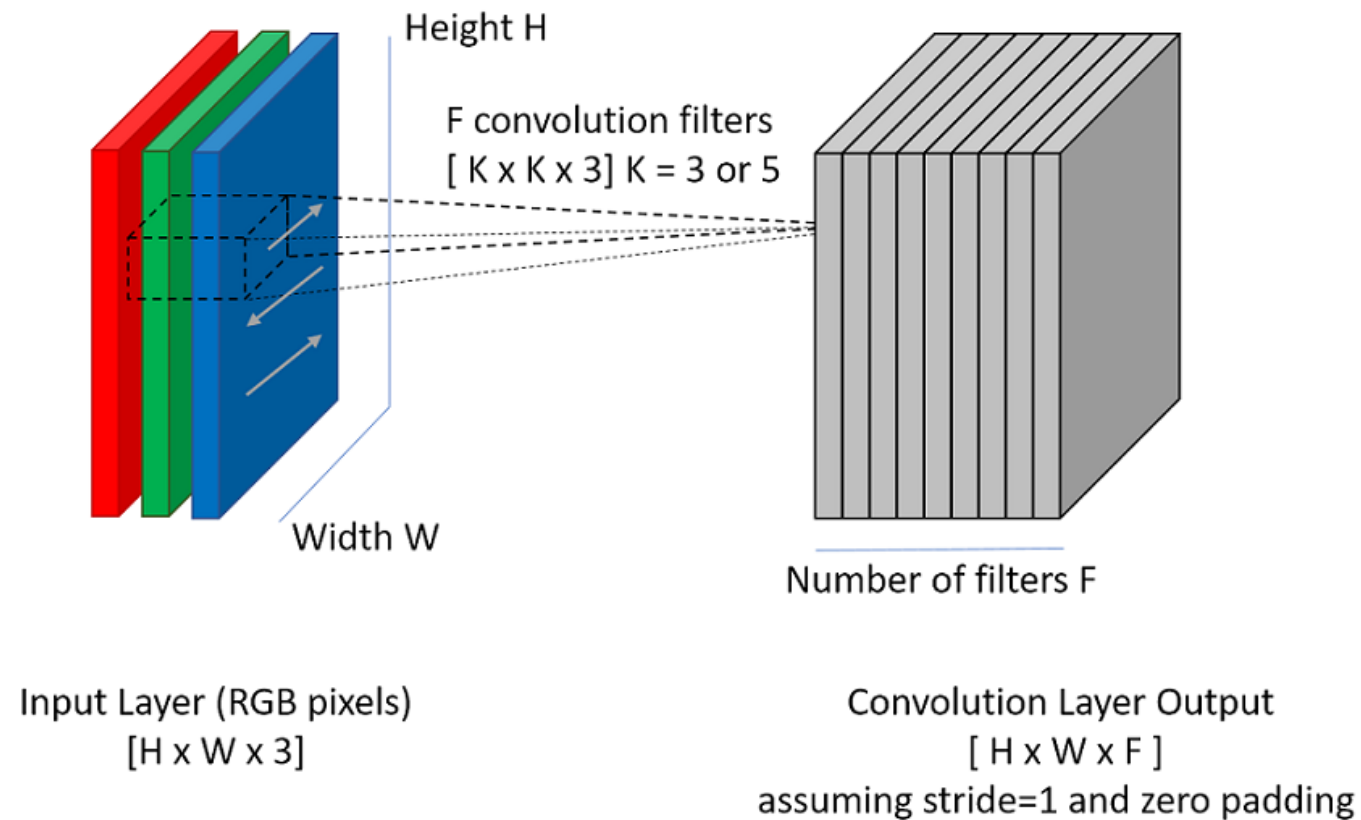
7x7 => preserved size!

in general, common to see stride 1, size F, and
zero-padding with $(F-1)/2$.
(Will preserve input size spatially)

- **Stride (korak)**

Predstavlja korak kojim pomeramo kernel
prilikom računanja skalarnog proizvoda
(operatora konvolucije). Najčešće se
pretpostavlja da je korak = 1

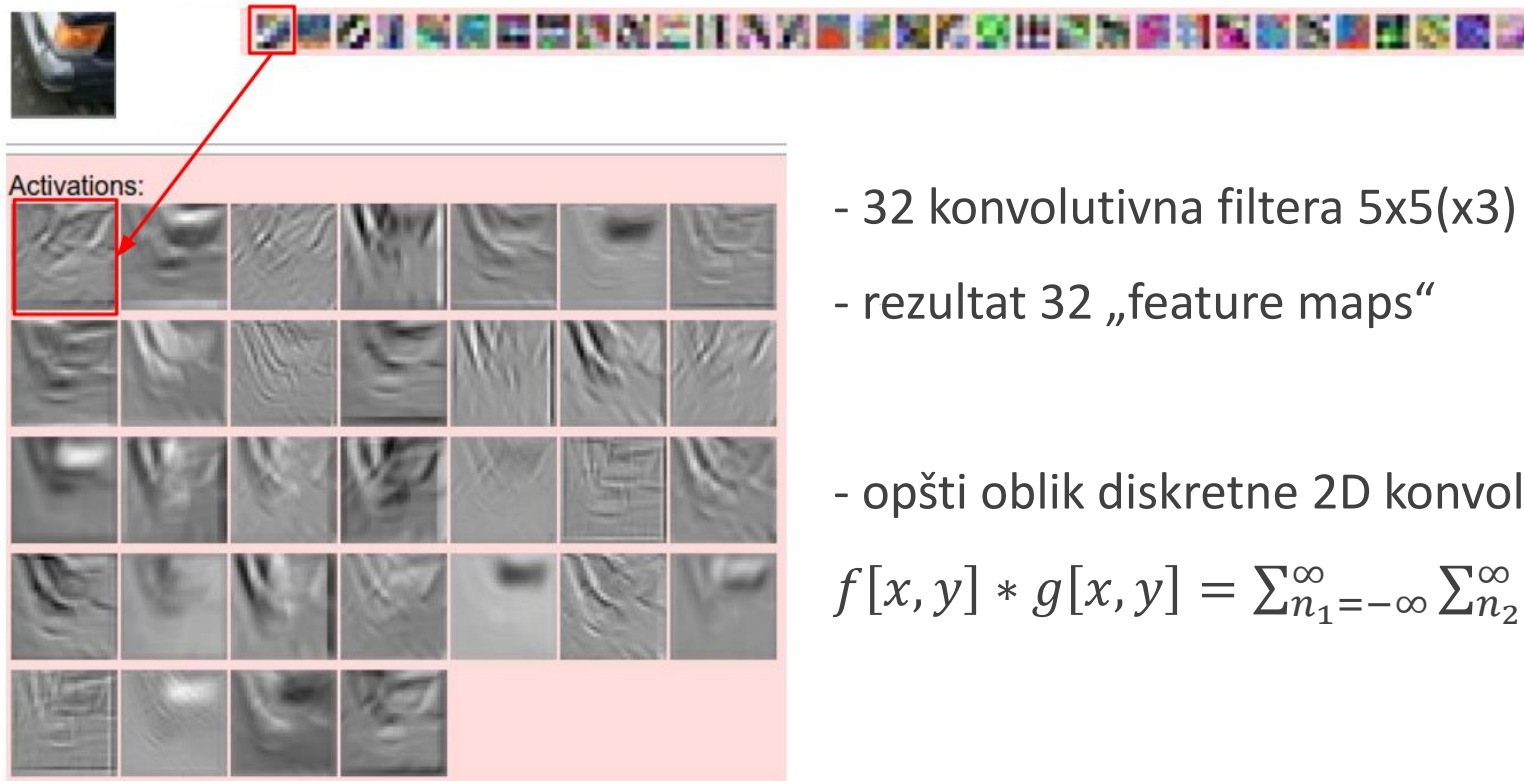
CNN – konvolutivni sloj



CNN – osnovne komponente

- **Ulaz** – slika (npr. $32 \times 32 \times 3$, x3 za RGB kanale)
- **Konvolutivni sloj** – N konvolutivnih filtera (npr. dimenzija $3 \times 3 \times 3$, $5 \times 5 \times 3$) koji kao proizvod daju ulaznu sliku konvuliranu ovim filterima
 - Konvolutivni filteri su zapravo **neuroni**
 - **Težine neurona** su zapravo elementi matrice konvolutivnog filtera
 - Konvolutivni sloj je veličine $32 \times 32 \times (3 \times N)$ i rezultat ovo sloja su N tzv. „feature maps“ (mape osobina)
 - „Feature maps“ mogu biti i manje veličine, u zavisnost od toga kako je definisana konvolucija graničnih piksela (slika 32×32 kada je konvulirana rezultat može biti npr. 28×28)
- **Aktivaciona funkcija** – primena neke nelinearne aktivacione funkcije nad „feature maps“
- **Pooling sloj** – radi tzv. „downsampling“ nad „feature maps“, tj. smanjuje dimenzije na npr. 16×16
- **MLP** (fully connected) – najobičniji višeslojni perceptron na kraju CNN, radi klasifikaciju

CNN – konvolutivni sloj



- 32 konvolutivna filtera 5x5(x3)

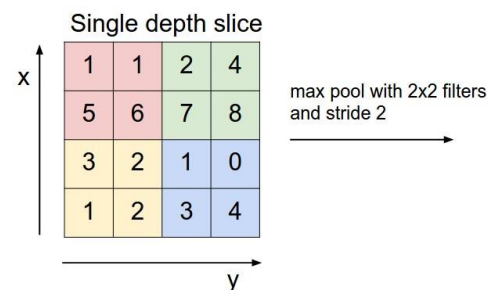
- rezultat 32 „feature maps“

- opšti oblik diskretne 2D konvolucije (f je filter, g je slika):

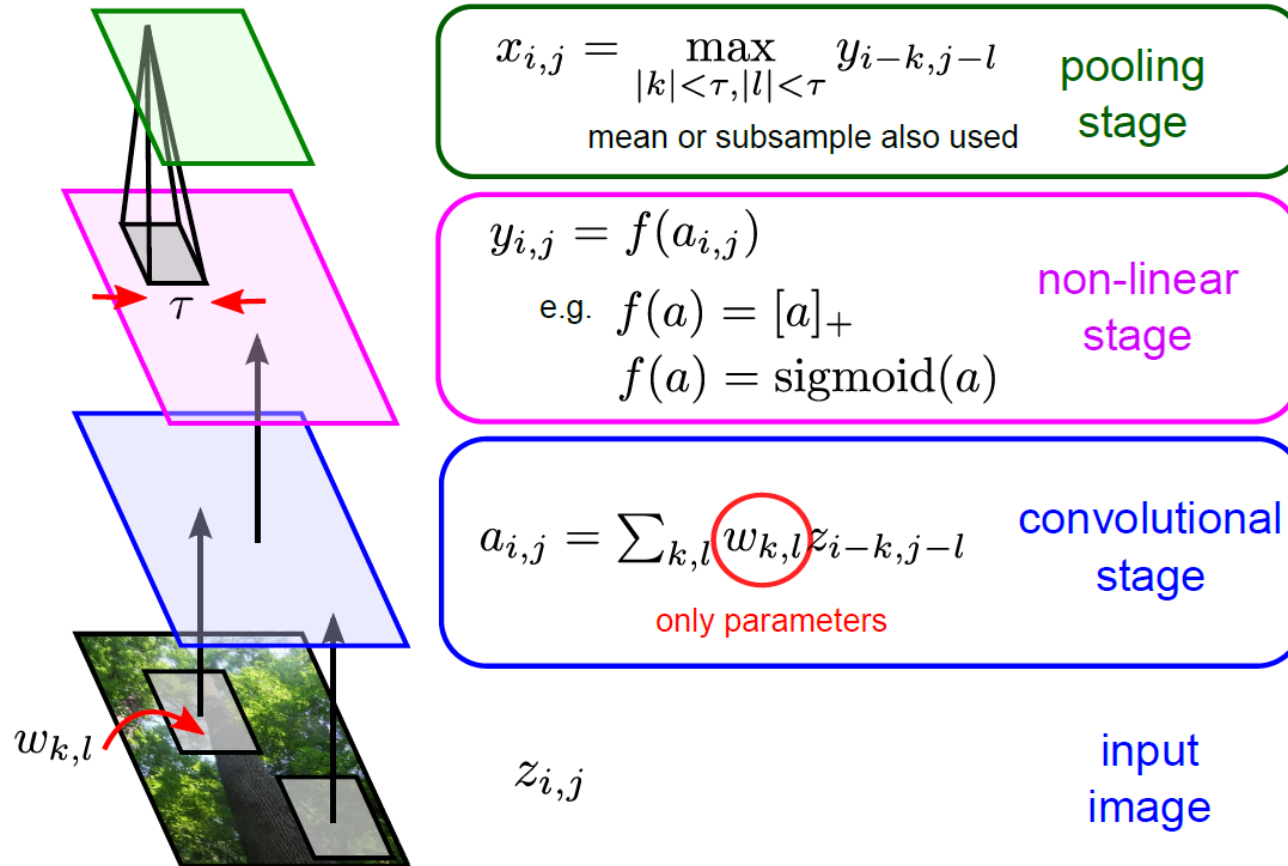
$$f[x, y] * g[x, y] = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} f[n_1, n_2] * g[x - n_1, y - n_2]$$

CNN – pooling sloj

- Dodatno smanjuje broj parametara smanjivanjem veličine izlaza iz konvolutivnog sloja
- Downsampling
- Npr. 32x32 -> 16x16 (4 puta manje parametara!)
- Uglavnom se uzima filter 2x2 koji se pomera po slici (npr. 32x32) sa korakom 2 (eng. *stride*) i za svaki 2x2 region se računa vrednost koja će biti upisana u rezultujuću sliku (16x16)
- Koja funkcija računa koja će vrednost biti upisana u rezultujuću sliku?
 - AVERAGE pooling – prosek vrednosti
 - **MAX pooling** – najveća vrednost



CNN – conv + activation + pool



Obučavanje CNN

- **Back-propagation sa SGD**
 - Jedino što je sad malo kompleksnije izračunavanje izlaza
- Poboljšanje podataka (eng. *data augmentation*)
 - Translacija, rotacija, preslikavanje ulaznih podataka (slika) i dodavanje šuma/distorzija
 - Značajno poboljšava performanse obučavanja CNN
 - Sprečava over-fitting
- Najviše računanja/memorije je u prvim konvolutivnim slojevima
- Najviše parametara je u poslednjim FC (fully connected) slojevima

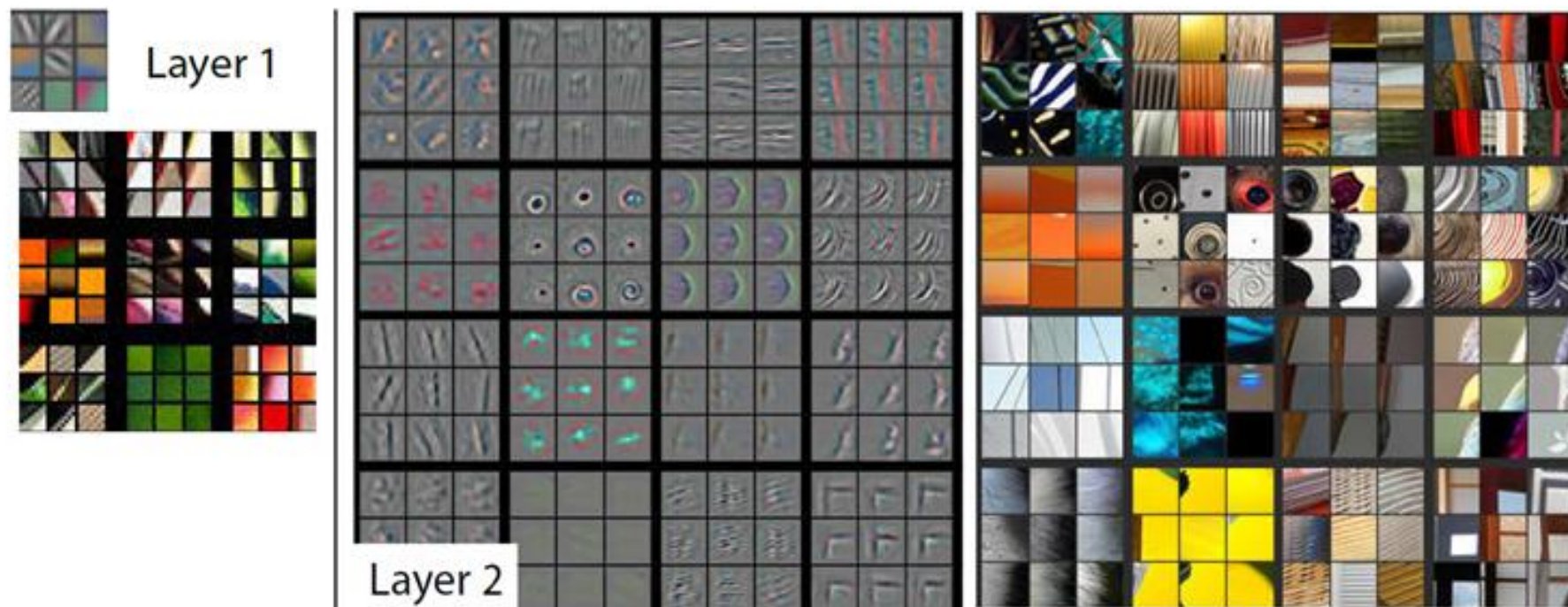
Tipične CNN

- 5 conv+activation+pooling slojeva
- 3 sloja u MLP na izlazu
- 500k neurona
- 50M parametara
- Vreme obučavanja – nedelju dana (GPU)

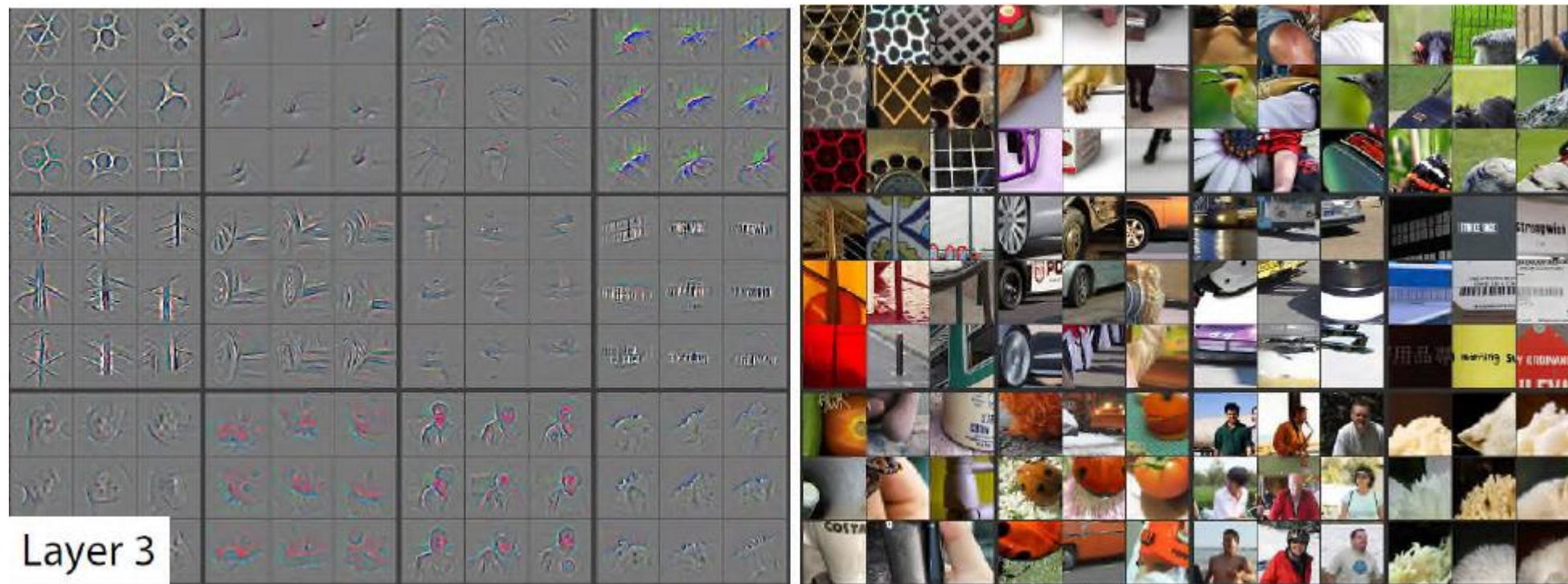
Moderne CNN

- Konvolutivni filteri 3x3 (neke čak i 2x2 i **1x1!?**)
 - Pooling filter 2x2 (neke čak i manje – „fractional pooling“)
 - Pomeranje pooling filtera za korak 1
 - **Veoma duboke** (preko 10 slojeva)
-
- ... za ovake mreže nemamo resurse ... još ;)

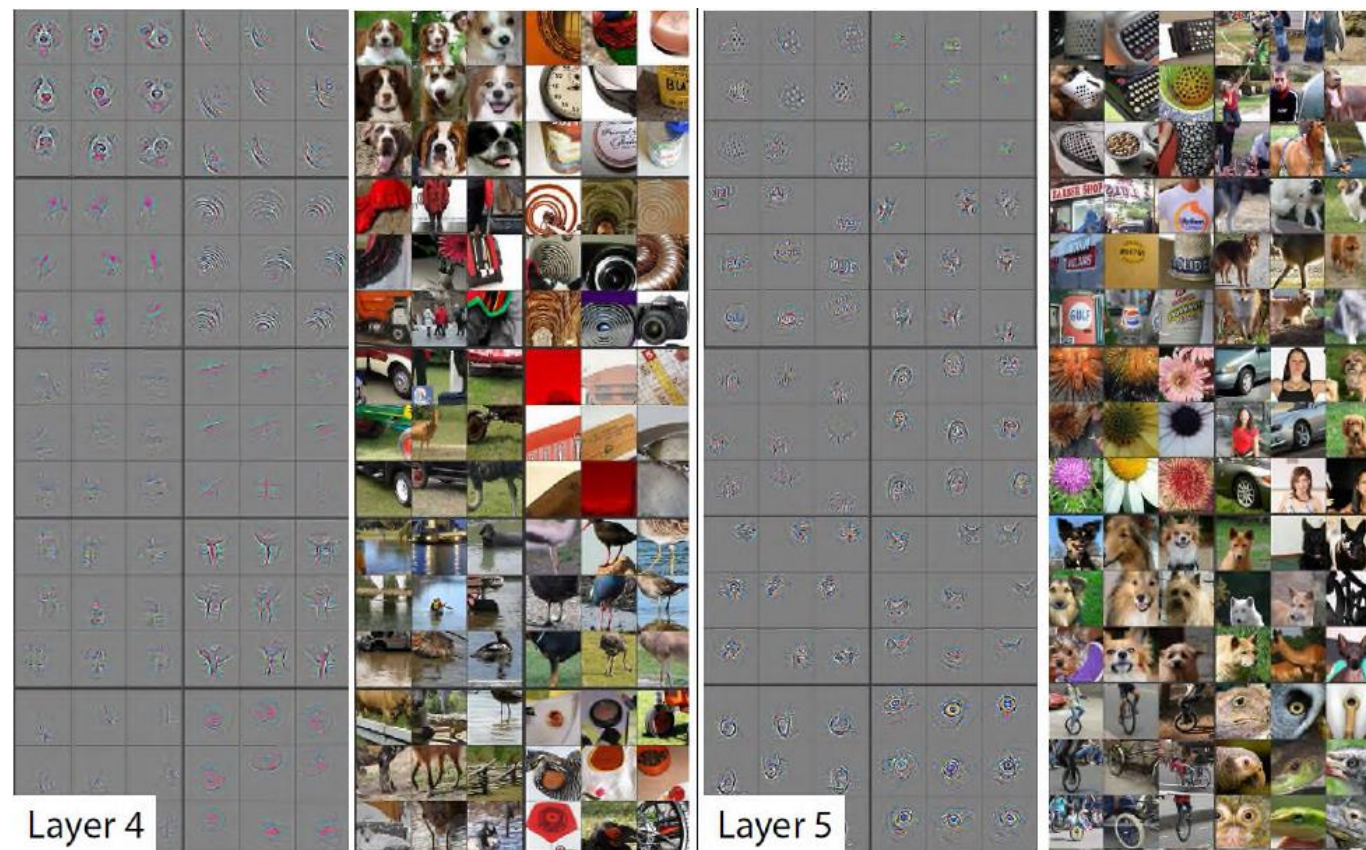
Obučeni slojevi CNN



Obučeni slojevi CNN



Obučeni slojevi CNN



Kratak pregled svega o CNN

- Viši (dublji) nivoi reprezentuju više apstraktne osobine
- Viši nivoi su nezavisni od:
 - Translacije
 - Rotacije
 - Osvetljenja
- Odličan metod učenja i detekcije osobina (eng. *feature detectors/extraction*)
 - Prvi sloj nauči npr detekciju ivica
 - Dublji slojevi kompleksnije stvari
 - Integracija obučavanja klasifikatora (MLP) sa učenjem interne reprezentacije osobina

Ali...

- Modeli sa dubokim hijerarhijama su **zapravo veoma stara ideja, ne nova**
- Čitava revolucija „dubokog učenja“ (eng. *deep learning*) je proizvod pre svega nekih novih metoda za inicijalizaciju i obučavanje dubokih mreža, ali i dostupnog hardvera (GPU)
- Trenutne metode kao cilj imaju nezavisnost od translacije, rotacije, itd... Ali ovo je još uvek prilično daleko od biološke, prirodne vizije
- Klasifikacija slika je sve što (bar trenutno) možemo – kažemo CNN „Reci mi šta se nalazi na ovoj slici“

Za čitanje

<http://cs231n.github.io/convolutional-networks/>

CNN – DEMO (CIFAR 10)

- CIFAR 10 data set: 50k slika za obučavanje, 10k slika za testiranje u **10** kategorija
- <http://cs.stanford.edu/people/karpathy/convnetjs/demo/cifar10.html>

