

A mesterséges intelligencia alkalmazása a videoanalitikai rendszerekben

HTE előadás 2019 február 28.
Imre Gábor

AI - azaz mesterséges intelligencia robbanásszerű növekedése

- ▶ A pénzügyi szektorban tőzsdei robotok, kriptó valuták
- ▶ Telekom, marketing Google, Amazon
- ▶ Logisztika, szállitmányozás
- ▶ Okosváros projektek

Mindezek egyik alapja a **gépi tanulás** vagy **Machine Learning**

- Robotika
- Prediktív analitikák
- Ügyfélszolgálati chatbotok
- Önvezető autók
- Gyártás automatizálás,
- Gyógyszerkutató, orvosi diagnosztika

Most elsősorban a gépi látással fogunk foglalkozni, de kell néhány alapfogalom

- ▶ Neurális hálózatok: az idegsejtek felépítésének illetve működésének megismerése indította el az ilyen irányú kutatásokat
- ▶ A ,természetes' neurális hálózatok nagyszámú hasonló vagy azonos felépítésű, egymással összeköttetésben lévő építőelemekből (neuronokból) kialakult hálózatok.

A mesterséges neurális hálózatok számítási feladatok megoldására létrejött párhuzamos feldolgozást végző, adaptív rendszerek, amelyek eredete a biológiai rendszerekből származtathatók. (Adaptív= a változó környezethez való alkalmazkodás képessége)

A hagyományos algoritmus megoldásokkal (procedurális és deklaratív programozás) szemben hatékonyabbak, ennek 2 oka van:

1- párhuzamos felépítés

2 - tanulási képesség

A problémamegoldás nehézségei:

- ▶ Jelenleg nem ismert az algoritmus ami a megoldáshoz elvezet
- ▶ Nem ismerjük a folyamat mögött meghúzódó törvényszerűséget
- ▶ Nem lehet a feladat megoldását reálisan elfogadható idő alatt megtalálni

A megoldás az adatokban rejlik, a tudás az adatokban rendelkezésünkre áll!

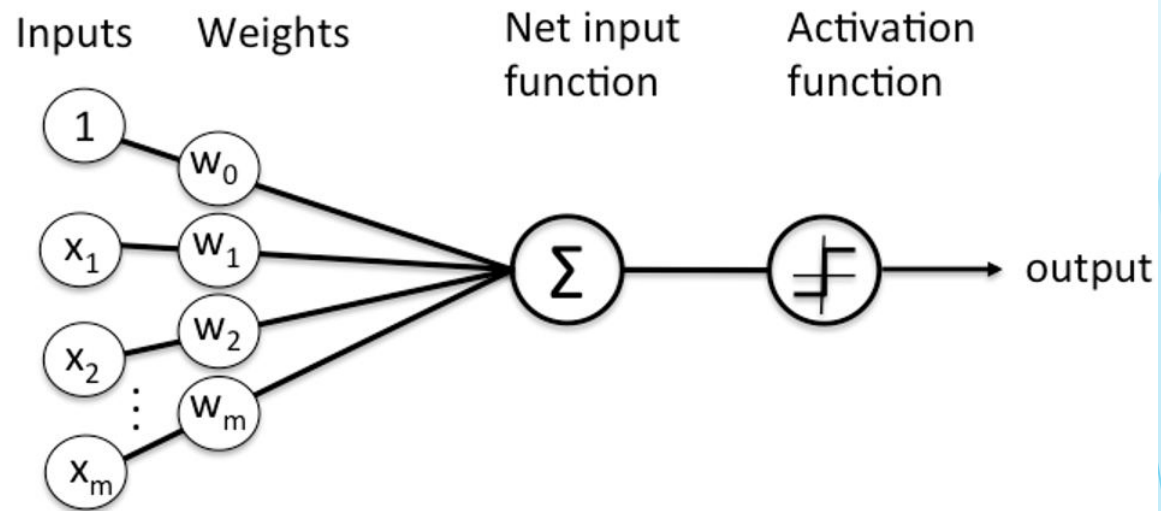
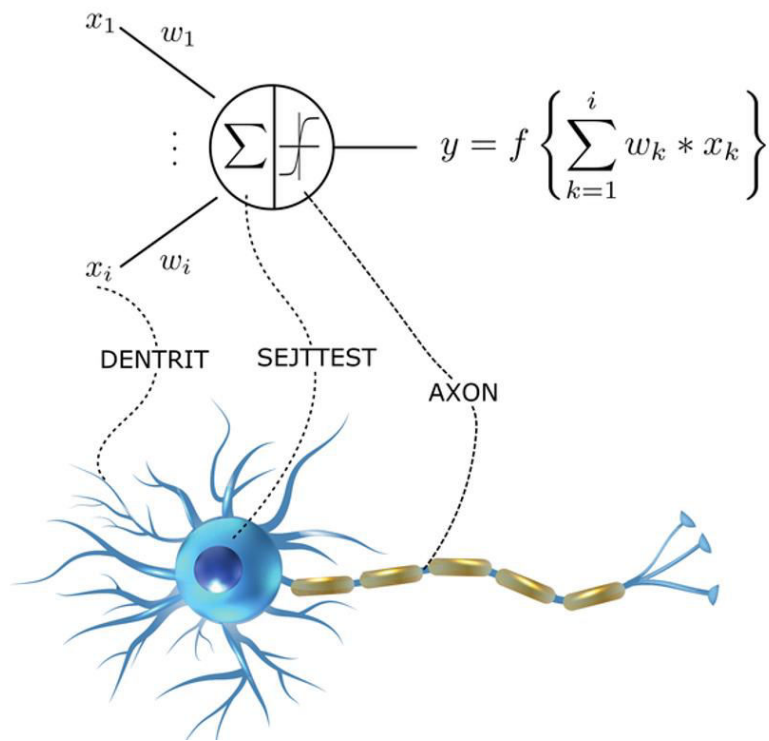
A neurális hálózatok képesek az adatokból nyert ismeretek általánosítására, azok kategorizálására

Univerzális approximátor

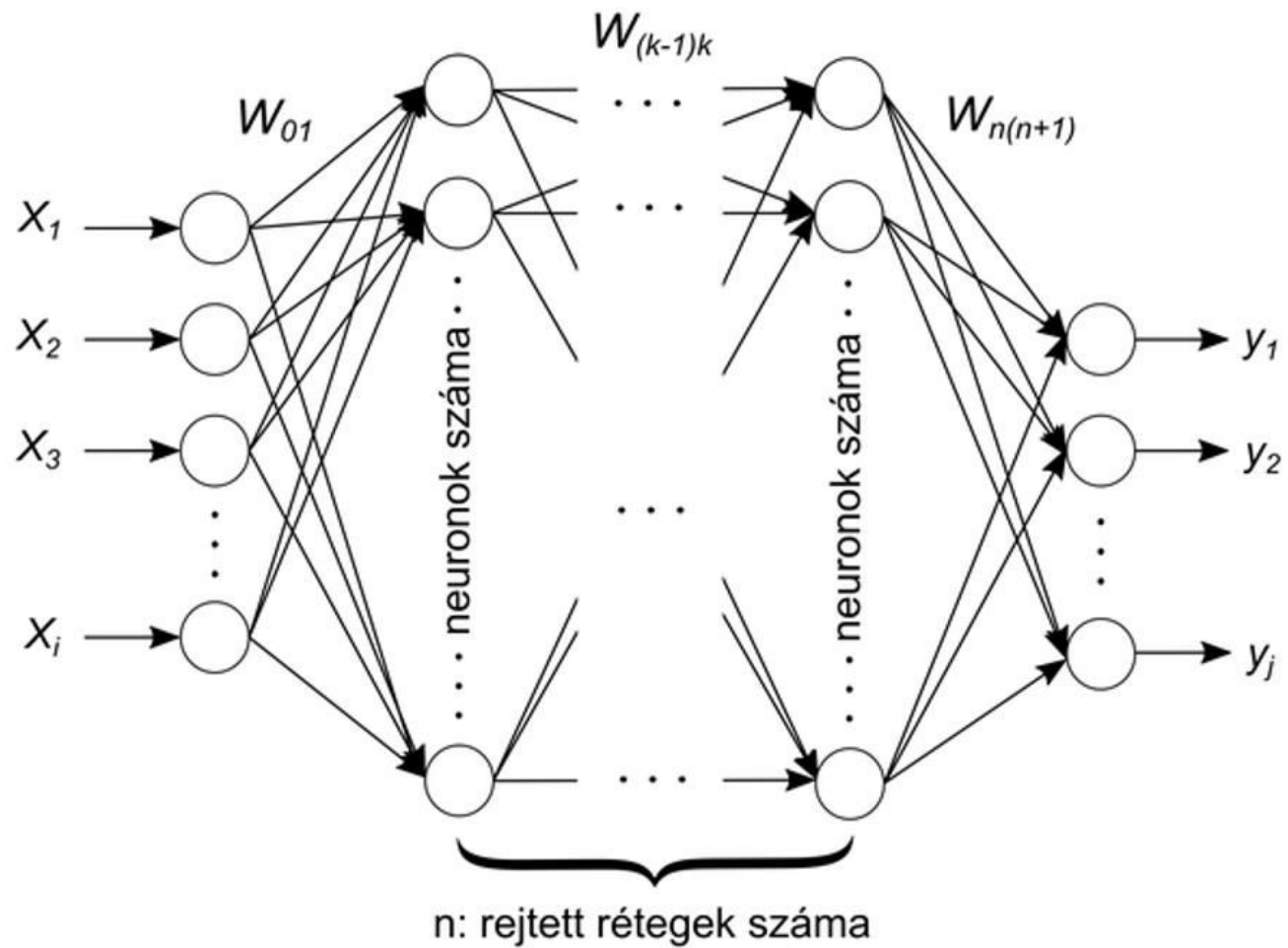
Tetszőleges be és kimeneti leképezéseket tudnak megvalósítani

Neurális háló alapelemei

- ▶ Neuron vagy műveleti elem (processing element)
- ▶ Aktivációs függvény vagy transfer függvény (csomópont vagy node)



Többrétegű struktúráról beszélünk



Tanulás adatokból

Automatikus teljesítőkéesség javítás

Az adaptív tanuló rendszerek alapvető jellemzője, hogy nem rögzített képességekkel rendelkeznek, hanem a képességeiket fejleszteni tudják, továbbá alkalmazkodni tudnak a változó körülményekhez, környezethez.

- ▶ Ellenőrzött vagy felügyelt tanulás (Tanító mintapárok)
- ▶ Nem ellenőrzött vagy felügyelet nélküli tanulás (nincs kivánt válasz a bemenethez)
- ▶ Analitikus tanulás (Növeljük a rendszer „tudását”)
- ▶ Tanulási fázis (learning phase) > *hosszú folyamat*
- ▶ Előhívási fázis (recall phase) > *rövid reakcióidő*

Konvolúciós Neurális Hálózatok CNN

Konvolúció???

Ez itt súlyozott, mozgó átlagot jelent

Jelfeldolgozásban:

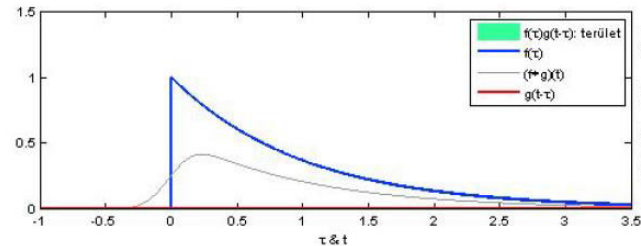
$$f * g : x \mapsto \int_{-\infty}^{\infty} f(t)g(x - t)dt$$

Diszkrét jelek esetén diszkrét konvolúció:

$$(f * g)(n) = \sum_{k \in D} f(k)g(n - k)$$

Képfeldolgozásban 2D konvolúció (szűrők):

$$C(i, j) = \sum_{m=0}^{(Ma-1)} \sum_{n=0}^{(Na-1)} A(m, n) * B(i - m, j - n)$$



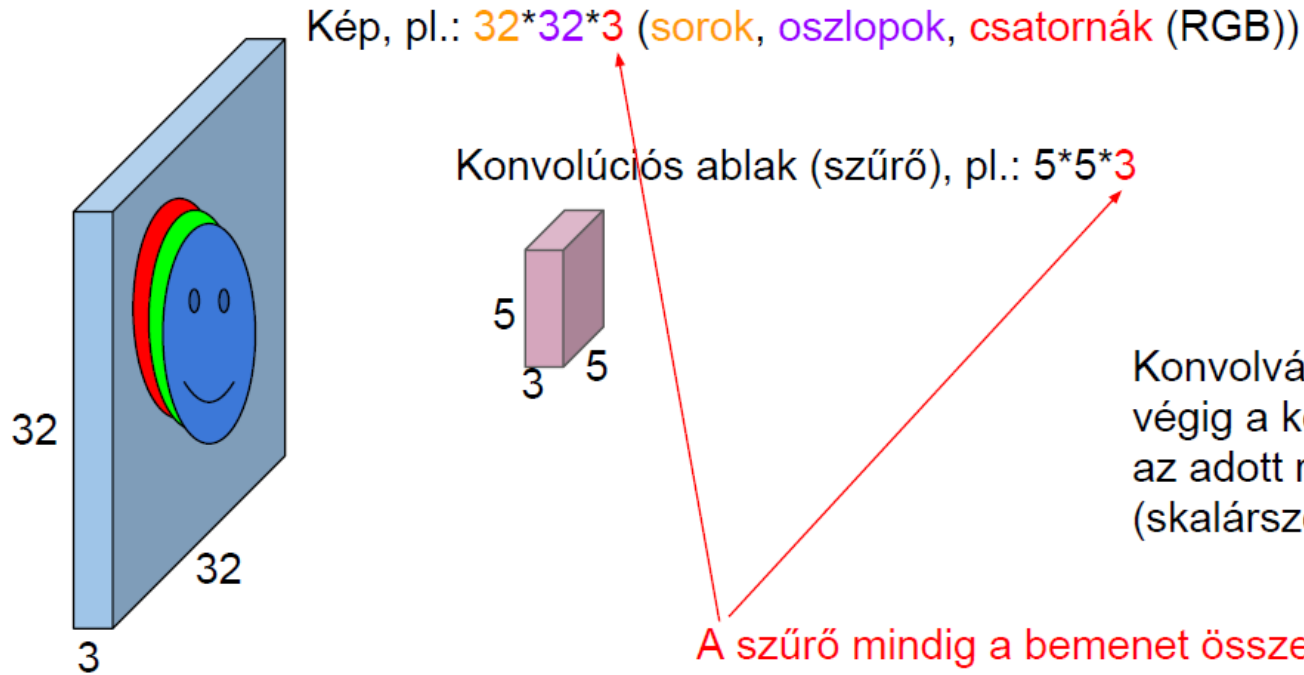
Gombási Gábor & Nagy Sándor

0	-1	0
-1	4	-1
0	-1	0

-1	-1	-1
-1	8	-1
-1	-1	-1



Konvolúciós réteg



Konvolváljuk a képet a szűrővel: húzzuk végig a kép minden részén, és számoljuk ki az adott résszel a konvolúciót (skalárszorzatot).

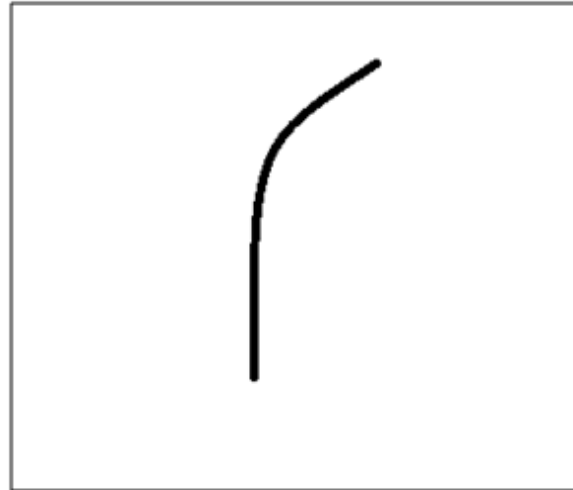
A szűrő mindig a bemenet összes csatornájára értelmezett, azonos a "mélységük"

A konvolúció eredménye egy skalár: a skaláris szorzata a kép részének a szűrőnek

Jellemző azonosítók keresése

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

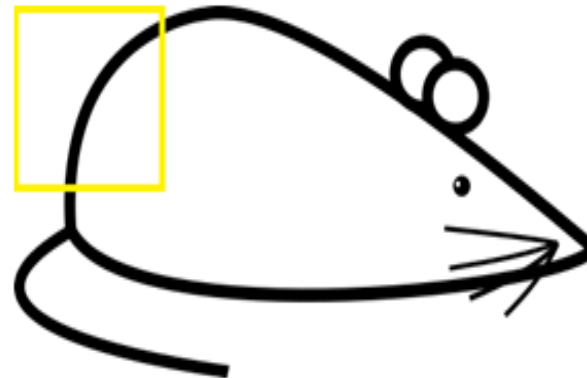
Pixel representation of filter



Visualization of a curve detector filter



Original image



Visualization of the filter on the image

Jellemző azonosítók keresése

- ▶ Vizuális kép pixel (elemi képpontok) Szűrő (súlyozási értékek)



Visualization of the receptive field

0	0	0	0	0	0	30
0	0	0	0	50	50	50
0	0	0	20	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0
0	0	0	50	50	0	0

Pixel representation of the receptive field

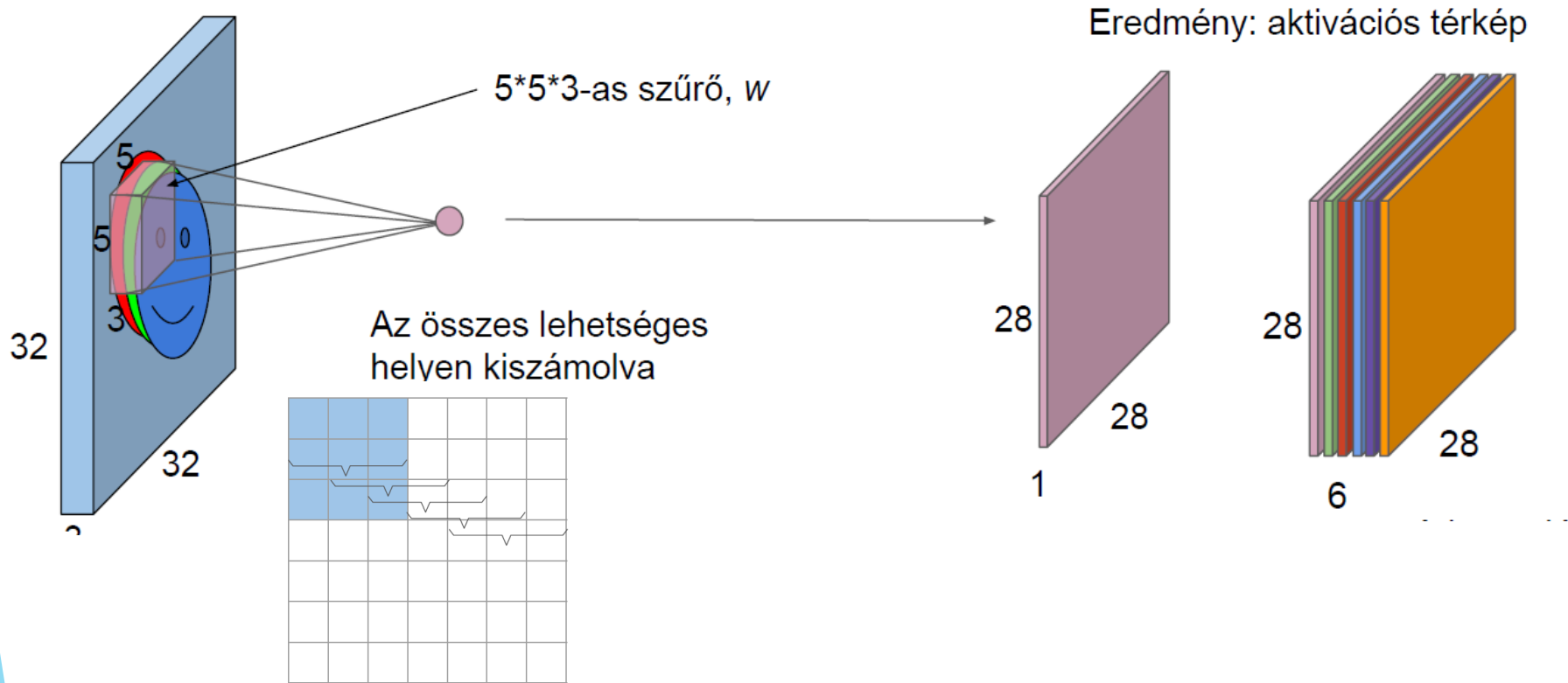
*

0	0	0	0	0	30	0
0	0	0	0	30	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	30	0	0	0
0	0	0	0	0	0	0

Pixel representation of filter

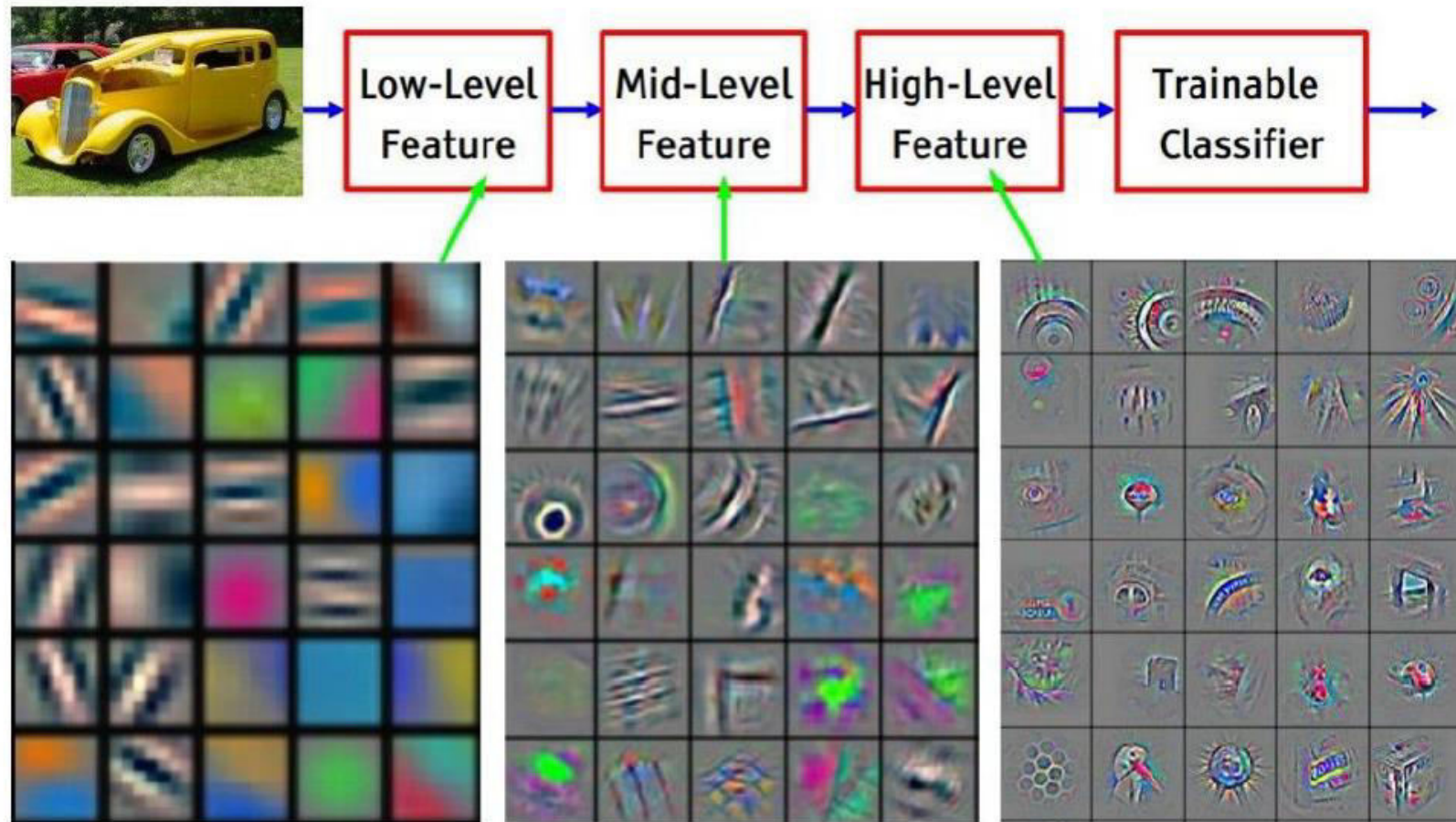
Multiplication and Summation = $(50*30)+(50*30)+(50*30)+(20*30)+(50*30) = 6600$ (A large number!)

Konvolúciós réteg



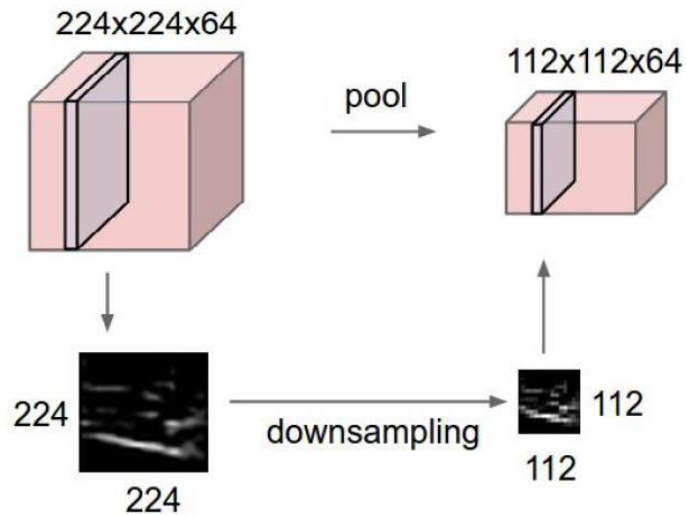
Több szűrőt is alkalmazhatunk W_1, W_2, \dots . Az aktivációs térképeket egymásra rakjuk, mint az RGB rétegeket.

A szűrők mint tulajdosság érzékelők



Összevonó (Pooling) réteg

az aktivációs térképek segítségével tovább csökkenti a reprezentáció méretét



Maximumot és átlagot számol

1	2	2	4
6	3	0	2
1	4	5	4
2	1	2	3

2 * 2 max pooling,
2-es lépésközzel

6	4
4	5

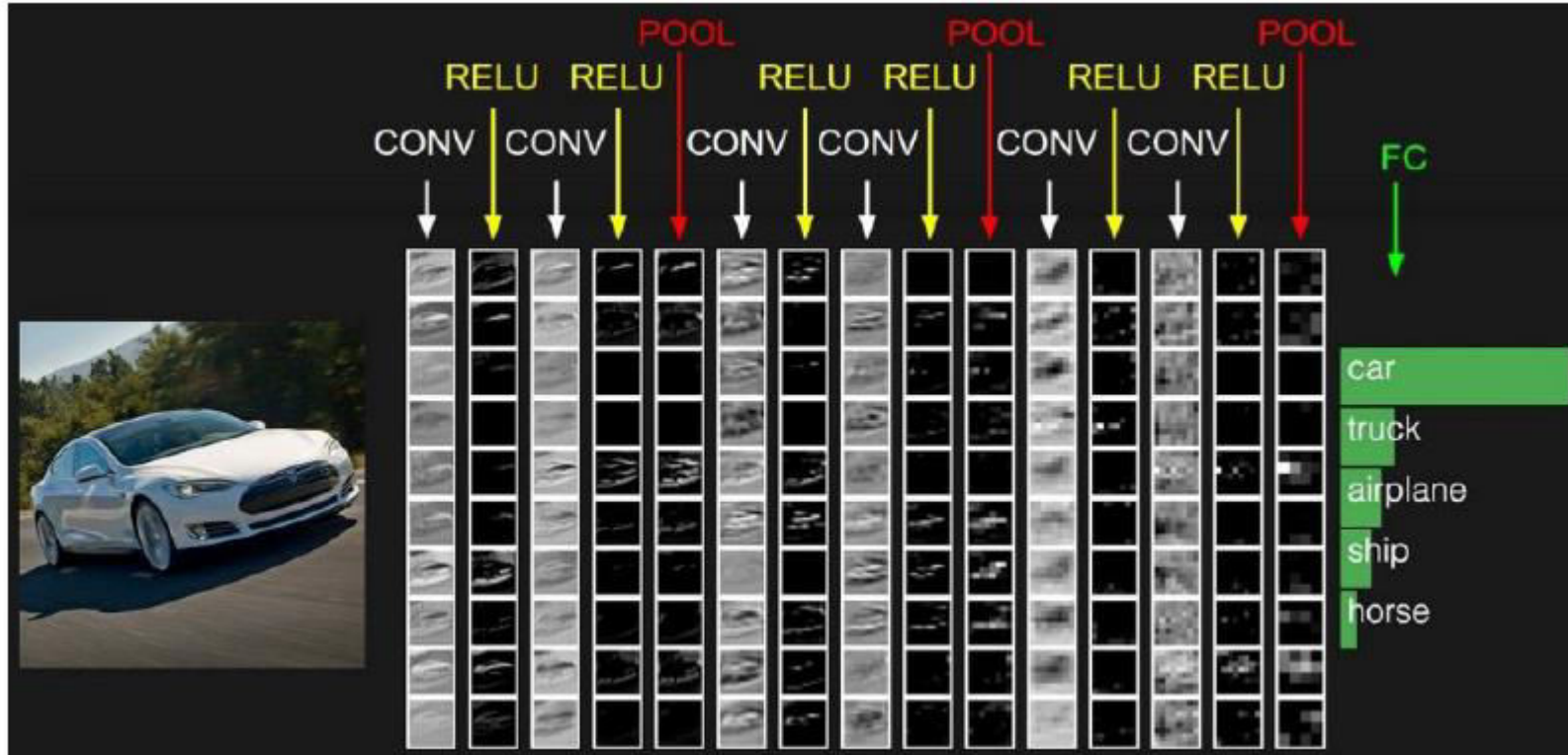
Értelmezés: van -e valahol az ablakon belül az adott jellemző? Pontos helye nem számít.

2 * 2 average pooling,
2-es lépésközzel

3	2
2	3.5

Értelmezés: az ablakon belül mennyire van jelen átlagosan az adott jellemző?

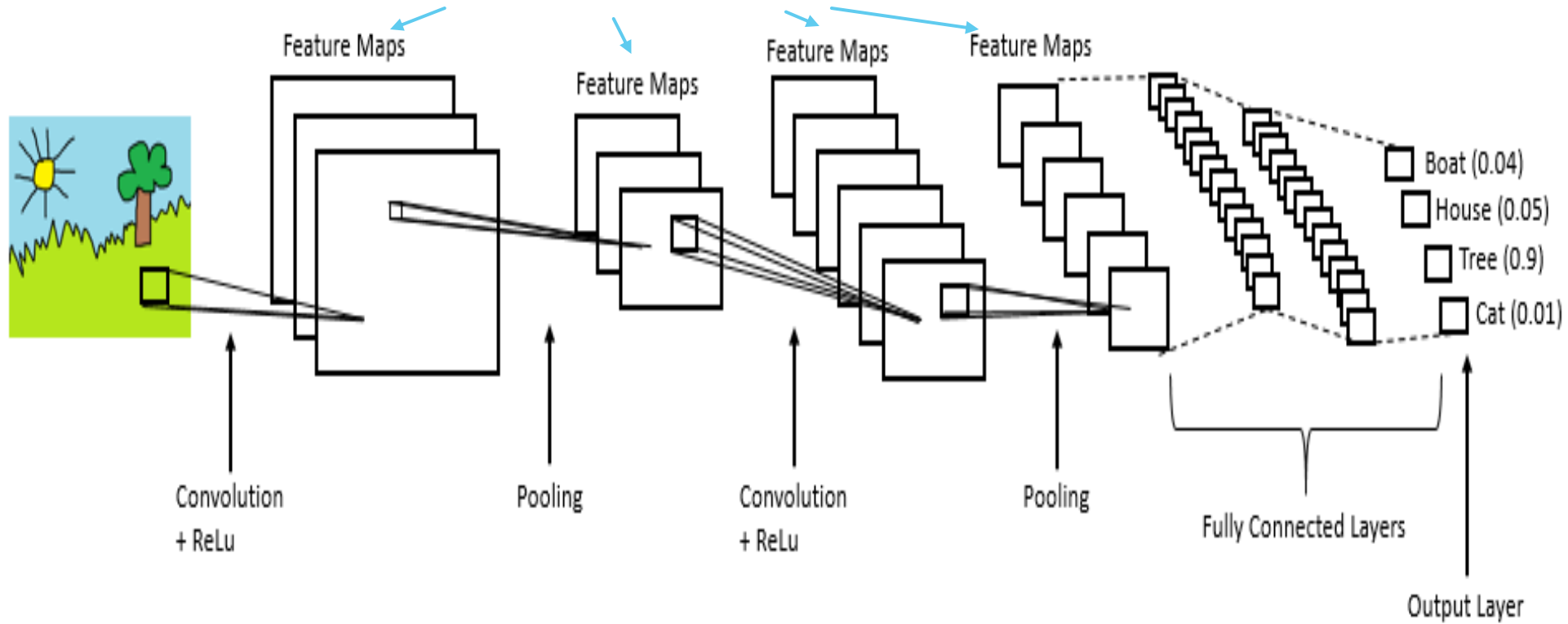
Teljesen összekötött réteg (Fully Connected Layer)



- ▶ A bemeneti térfogat minden eleme összevan kötve ezen réteg összes „neuronjával”
- ▶ A CNN utolsó rétege, a konvolúciós rétegek a jellemzők kiemelését végzik az FC rétegek pedig az osztályozást

CNN áttekintése

Aktivációs térképek



Konvolúciós réteg

összevonás

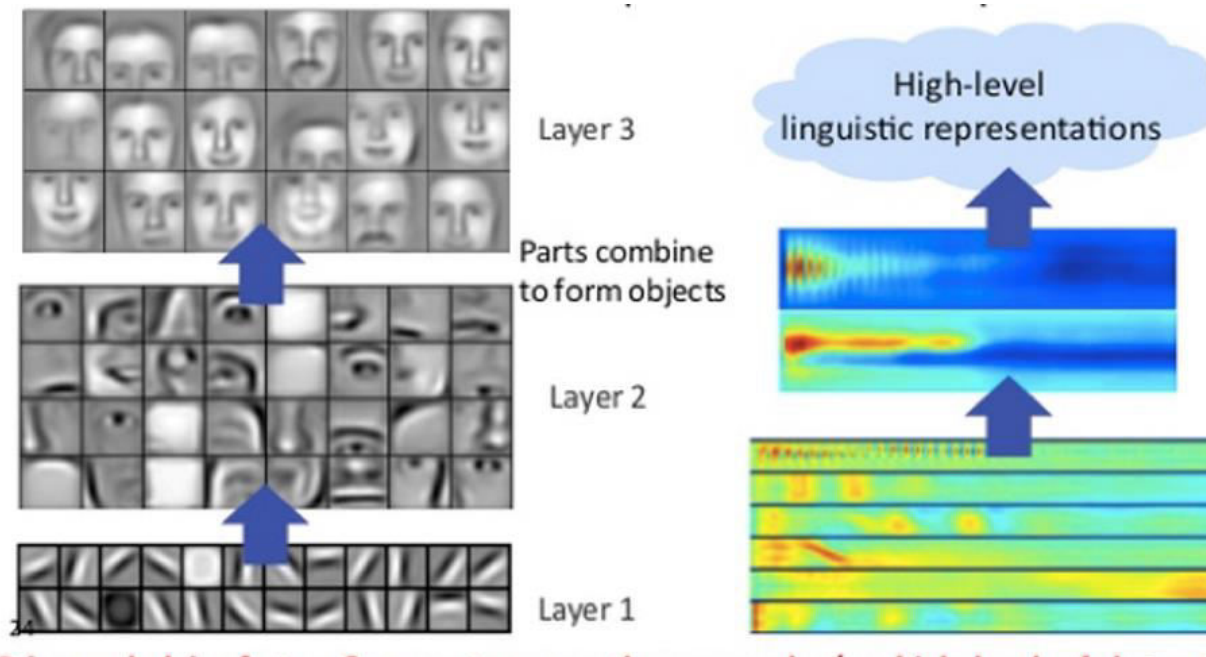
konvolúciós réteg

összevonás

teljesen összekötött réteg

kimeneti réteg

Mélytanulás egy gyakorlati alkalmazása Arcfelismerés



- ▶ A tanulási folyamat során a magasabb konvolúciós rétegek arcszerű jellemzőket tartalmaznak
- ▶ A mélytanuló rendszerek képesek hatalmas, többdimenziós adatállományokat kezelni

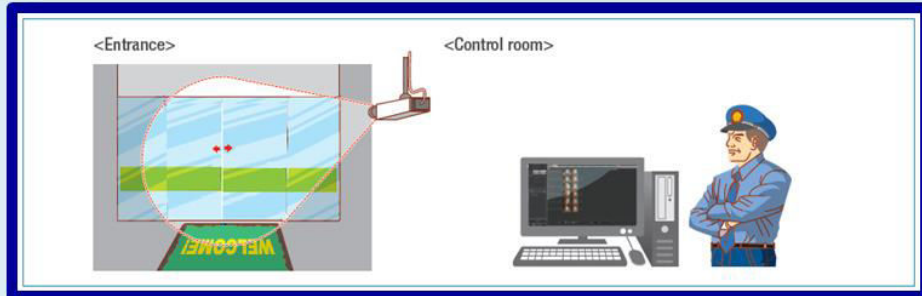
Párhuzamos feldolgozás hardver támogatással GPU (Graphical Processing Unit)



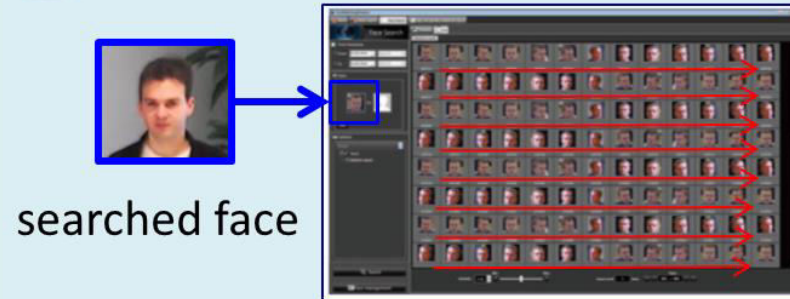
- Nagyszámú párhuzamos műveletek gyors elvégzése
- Hatékony kerső motorok kialakítása

Gyakorlati példa: arcfelismerés

Face matching (Real-time)



Face search (Post-search)



- Search people from each camera DB
- Specify Camera, Date & Time
- Display result (where/when he was)

People count, Age & Gender Statistics

- Count people of detected face and evaluate their age and gender
- Display result with statistical graph



Panasonic megoldás: arcfelismerés a kamerában

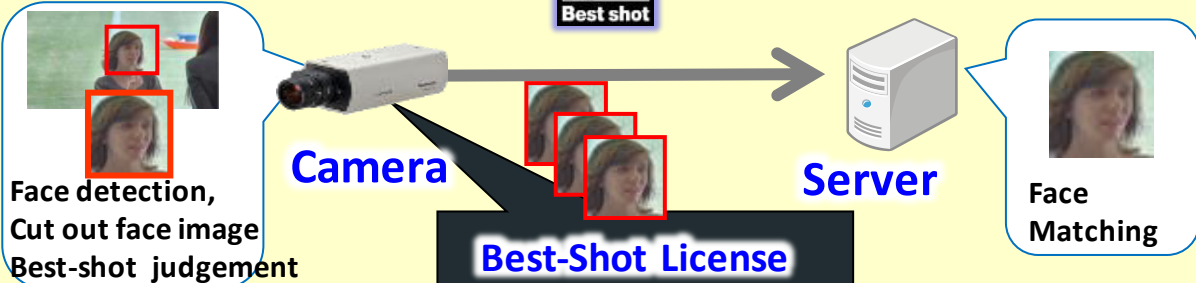
■ Face Best-Shot Technology

General streaming method



Send the entire image taken with the camera to the server, perform image analysis on the server side and check (server load: large)

Best-Shot Technology



Perform image analysis on our i-PRO series network camera with best shot license installed, and cut out the face image.

With the best shot function, only images suitable for authentication are sent to the server. Collect the sent face image on the server

(Server load: low)

■ Cost merit

Server workload reduction

"Image analysis" is performed by the camera side,

so it is **reduced to about 1/5 (*)**

With server load reduction, up to 20 cameras can be connected with one server.

Network load reduction

Since only the best shot image is sent from the camera,

System cost reduction*)

Approximately 40 to 50% reduction

(*) A system configuration of 10 or more cameras, compared with a general stream method

Best-Shot function Patent pending

Best shot determination example

- Face size image is larger than specified size
- Face angle is within specified angle
- Eyes are open



Köszönöm a figyelmet!