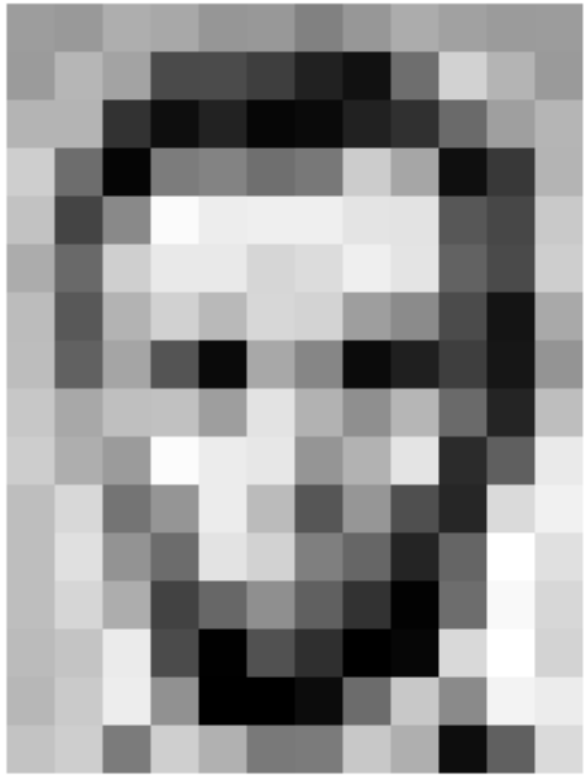
Deep Learning Szeminárium
III. előadás:
Konvolúciós hálózatok

Csiszárik Adrián (Rényi Intézet)

Mit lát a gép?



157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

Számok egy mátrixban: pixel intenzitások

Adjuk meg a képi jellemzőket

- Szem
- Orr
- Száj
- Haj

- Kerék
- Rendszám
- Lámpa

- Ablak
- Ajtó
- Tető

~~Adjuk meg a képi jellemzőket~~

Tanuljuk meg **automatikusan!**

- Szem
- Orr
- Száj
- Haj

- Kerék
- Rendszám
- Lámpa

- Ablak
- Ajtó
- Tető

Miért nehéz ez a gépnek?

Viewpoint variation

Scale variation

Deformation

Occlusion

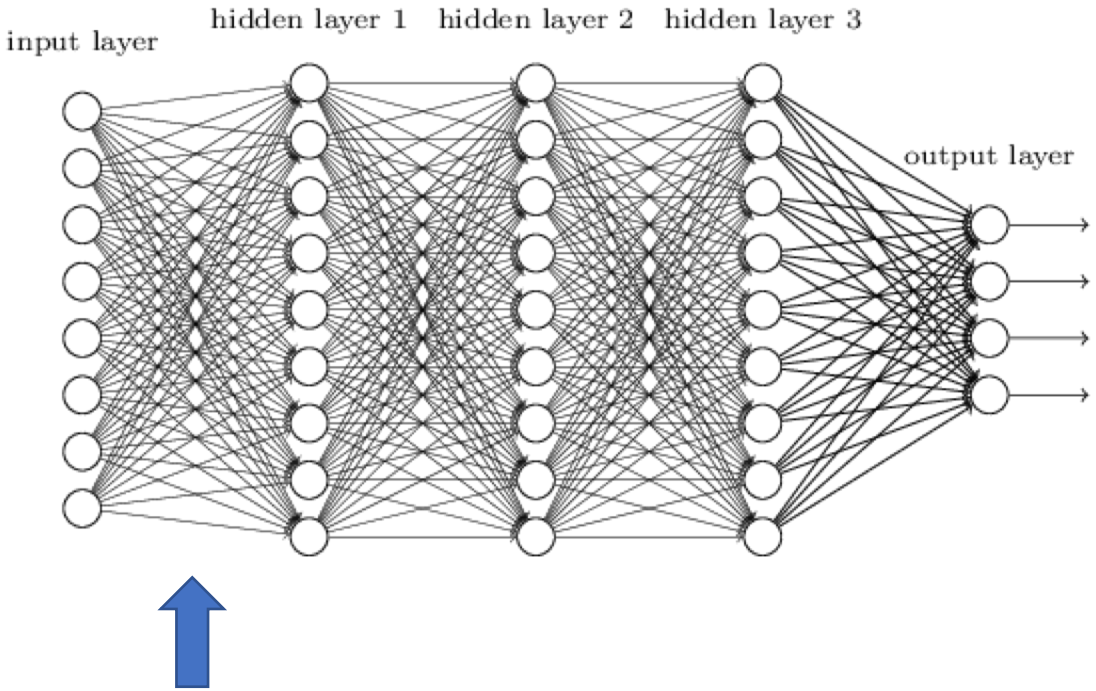
Illumination conditions

Background clutter

Intra-class variation

Ha egy sűrűn kapcsolt hálózattal tanulnánk…

Input: 2D kép
Pixel intenzitások vektora



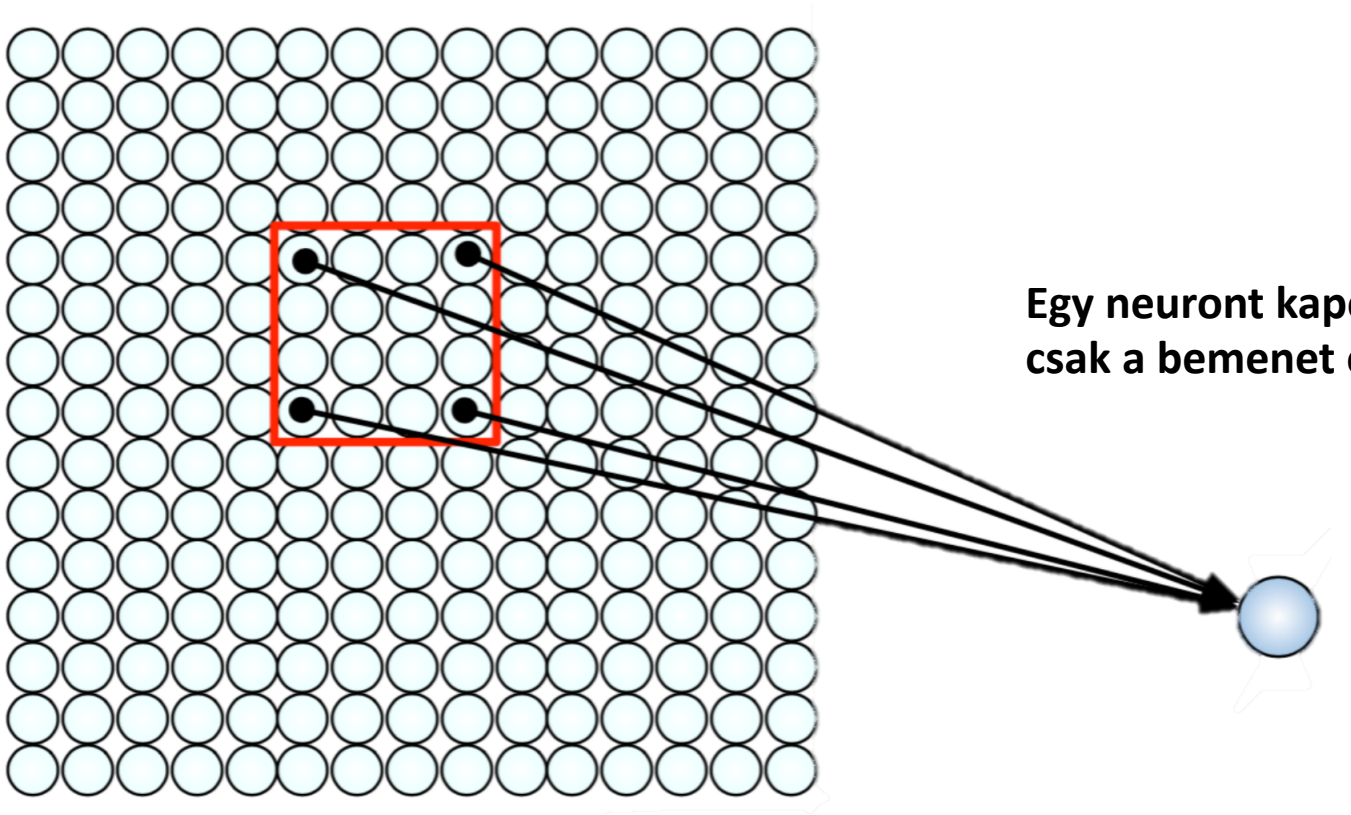
1.000x1.000 = 1.000.000 x 512 = 512.000.000 paraméter
Nincs térbeli lokalitás sem!

Lehetséges 1000x1000-es szürkeárnyalatos képek száma:

$$2^{1000000}$$

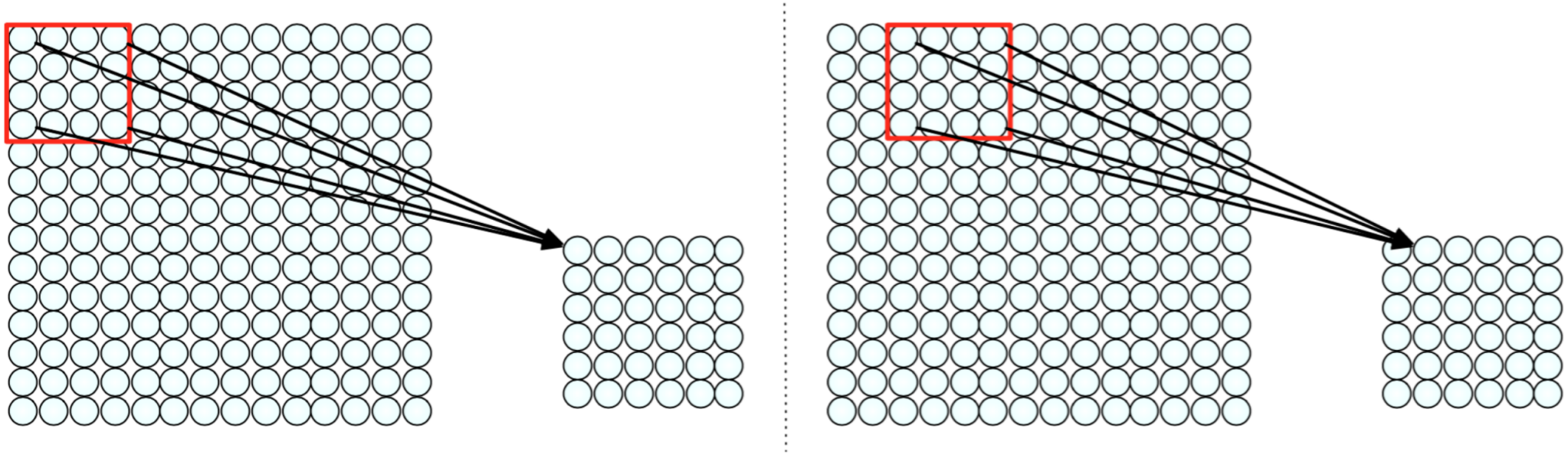
Használjuk ki a 2D struktúrát!

Input: 2D kép
Pixel intenzitások vektora

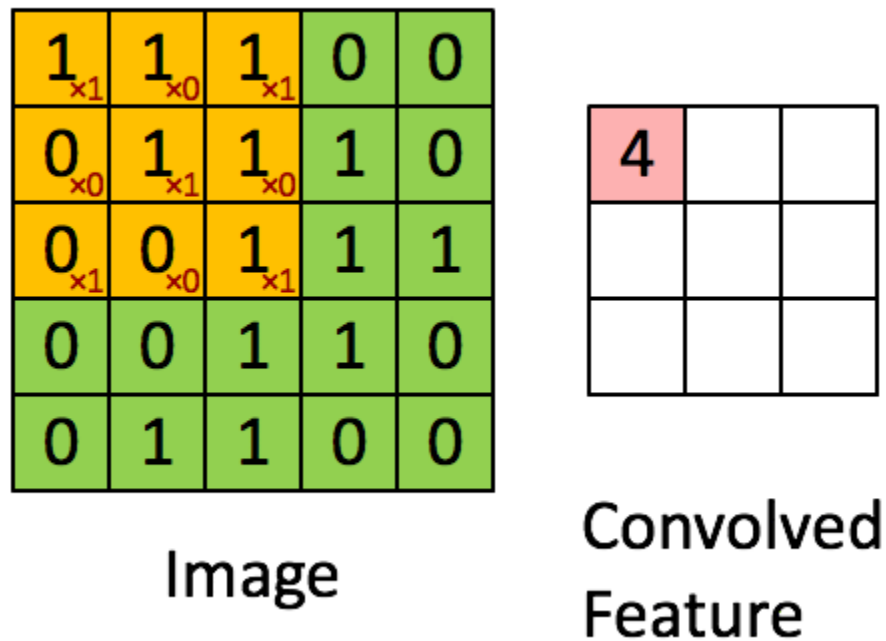


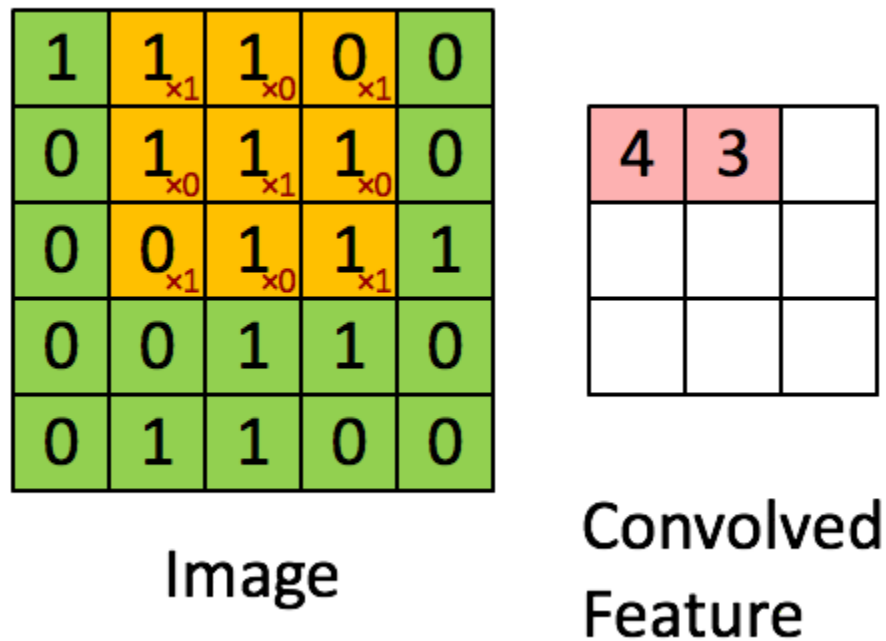
Egy neuront kapcsoljunk
csak a bemenet egy részéhez

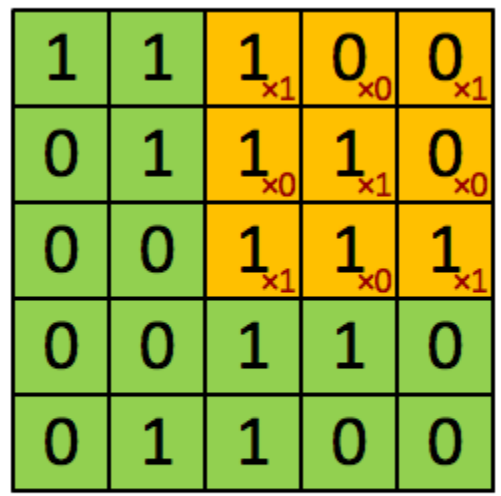
Használjuk ki a 2D struktúrát!



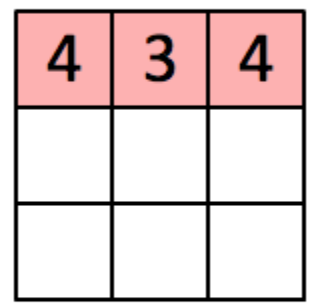
1. A neurális kapcsolatok meghatározásához csúsztassunk egy ablakot a bemeneten.
2. Alkalmazzunk súlymegosztást!


1×1 1×0 1×1 0 0
0×0 1×1 1×0 1 0
0×1 0×0 1×1 1 1
0 0 1 1 0
0 1 1 0 0
Image
4
Convolved Feature

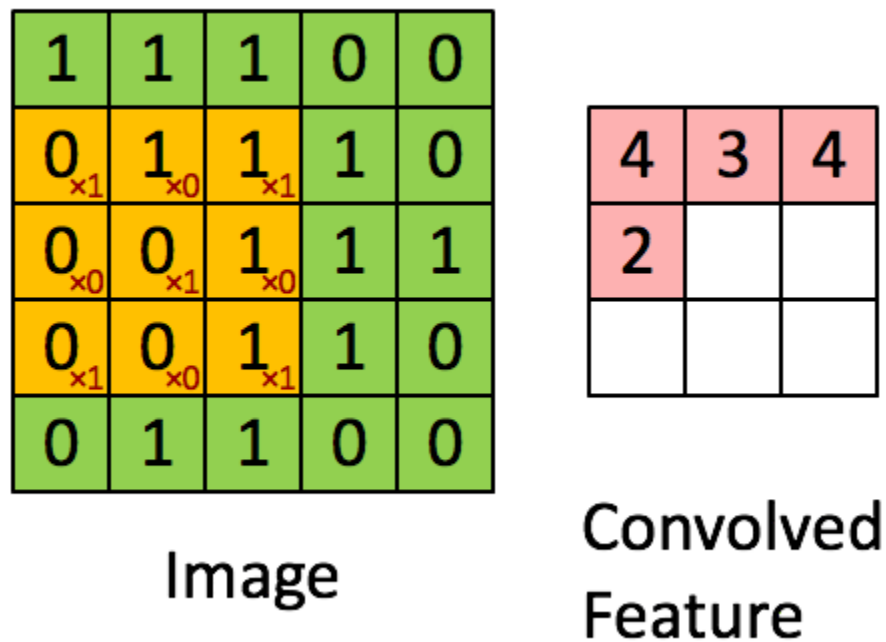

1 1×1 1×0 0×1 0
0 1×0 1×1 1×0 0
0 0×1 1×0 1×1 1
0 0 1 1 0
0 1 1 0 0
4 3
Image
Convolved Feature

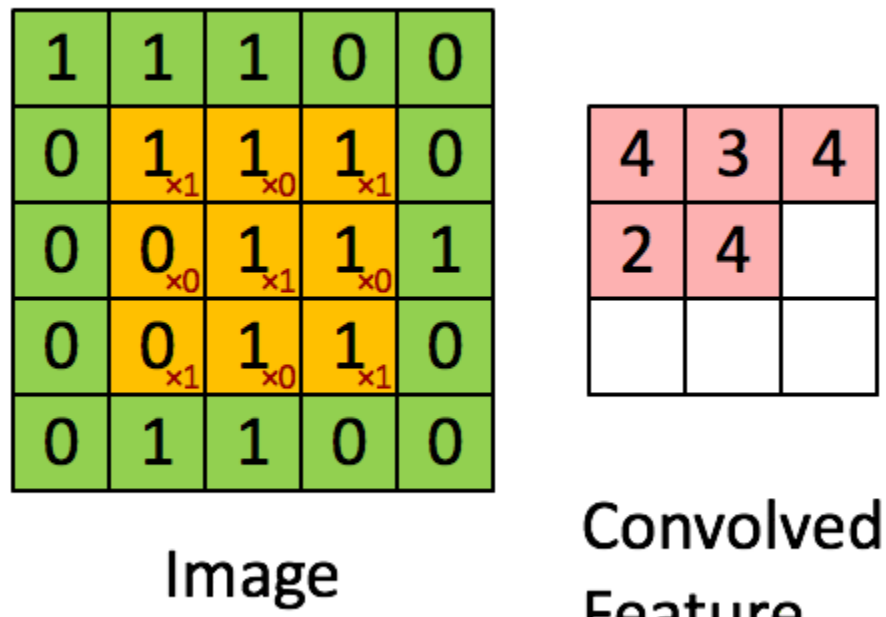


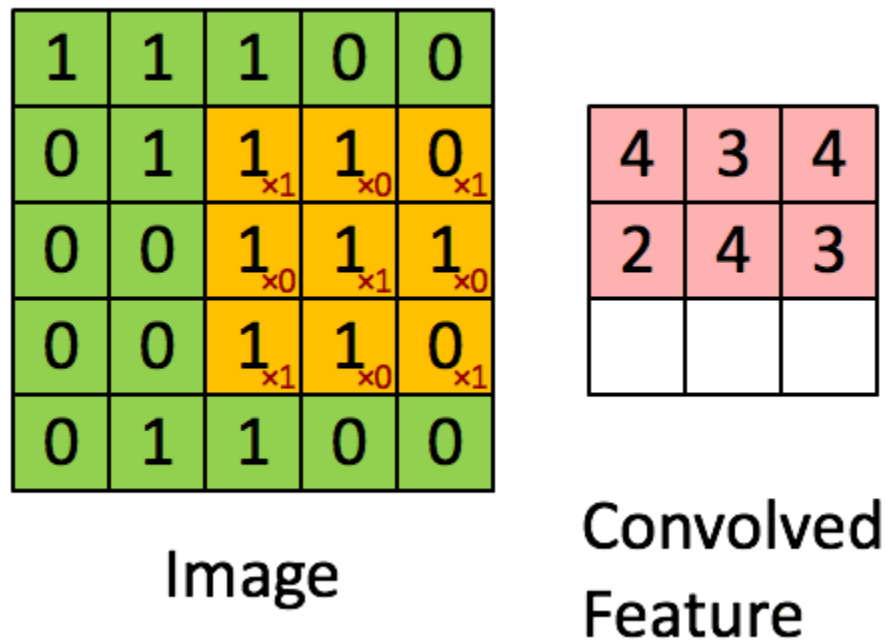
Image

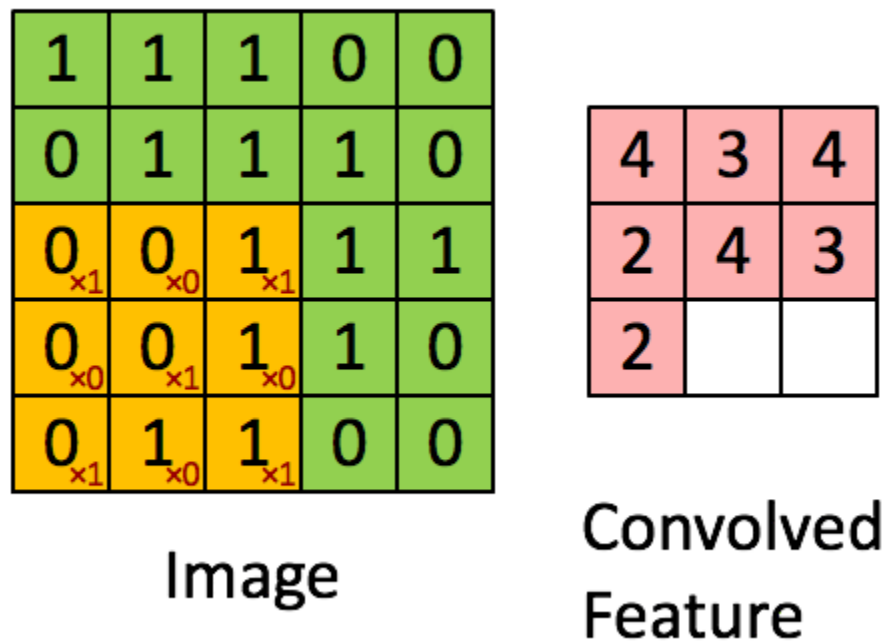


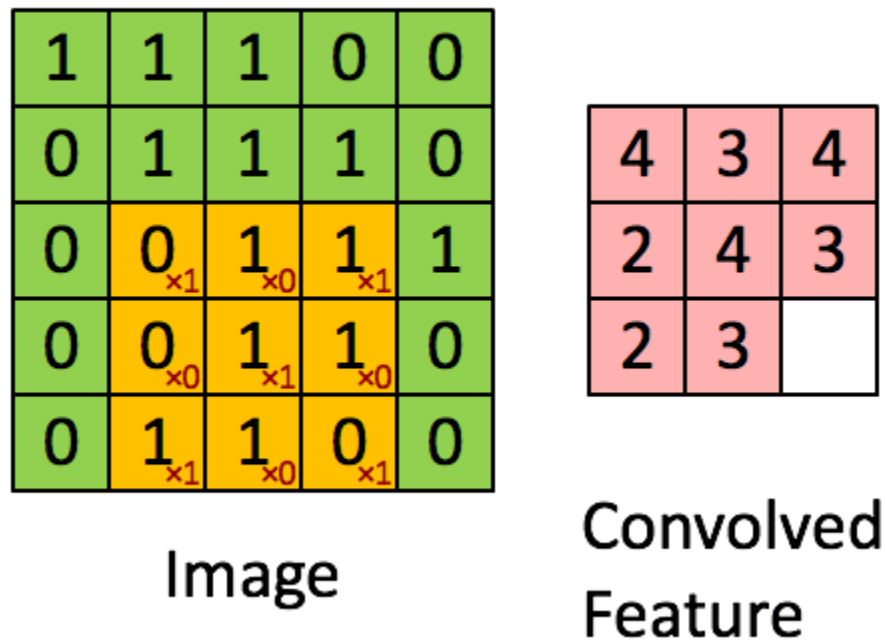
Convolved Feature

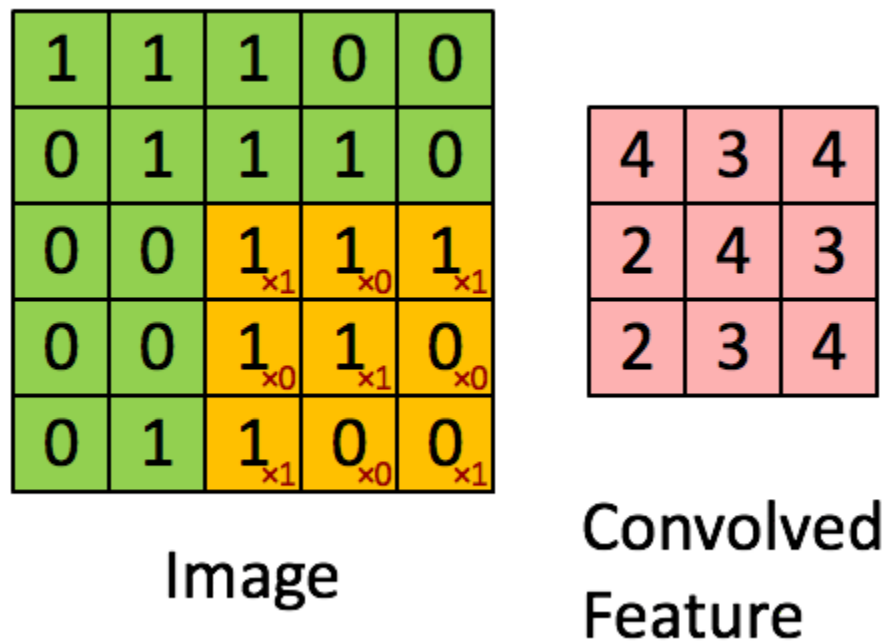

1 1 1 0 0
0×1 1×0 1×1 1 0
0×0 0×1 1×0 1 1
0×1 0×0 1×1 1 0
0 1 1 0 0
4 3 4
2
Image
Convolved Feature

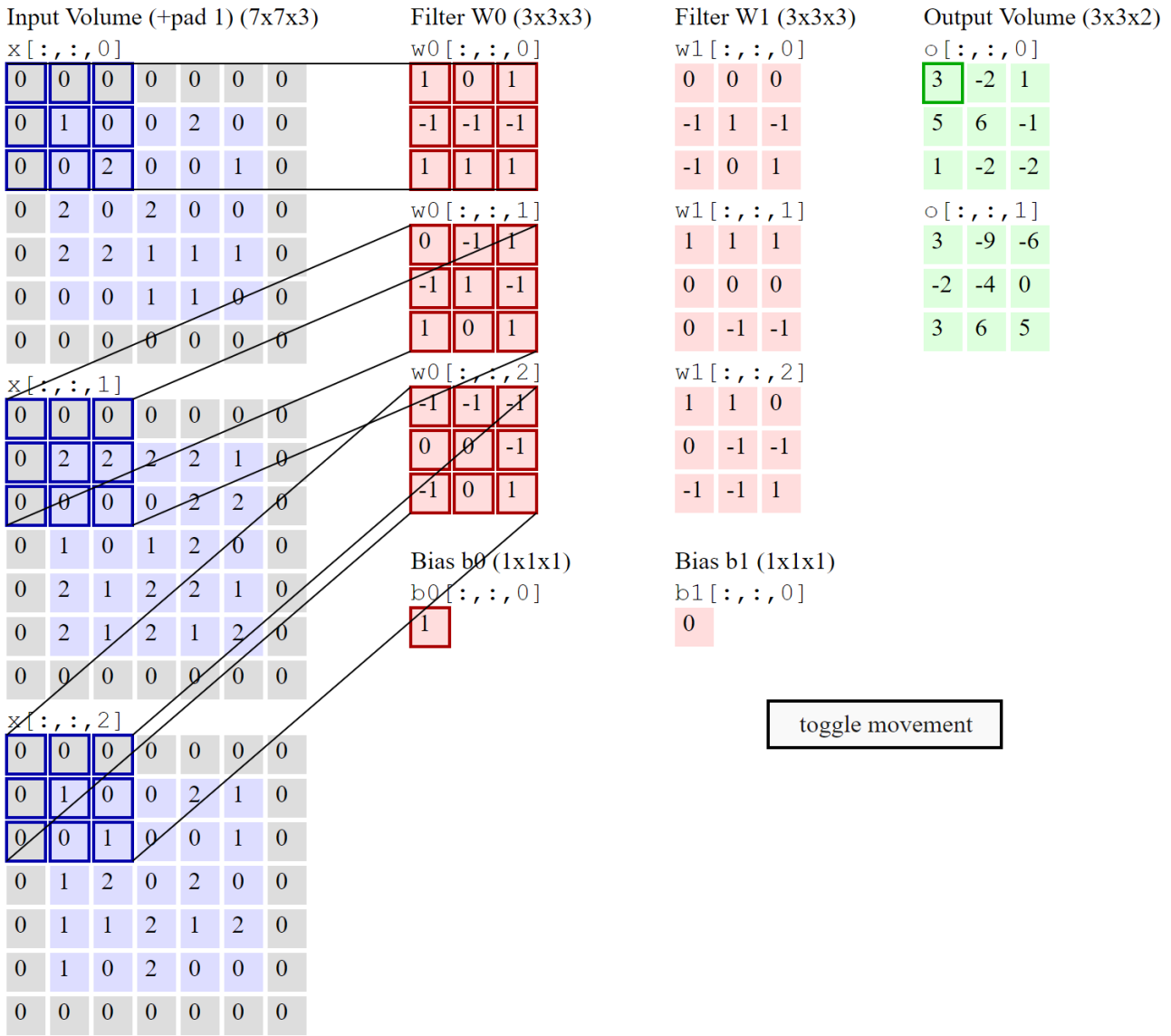

1 1 1 0 0
0 1 1 1 0
0 0 1 1 1
0 0 1 1 0
0 1 1 0 0
1 0 1
0 1 0
1 0 1
4 3 4
2 4
Image
Convolved Feature


1 1 1 0 0
0 1 1×1 1×0 0×1
0 0 1×0 1×1 1×0
0 0 1×1 1×0 0×1
0 1 1 0 0
4 3 4
2 4 3
Image
Convolved Feature


1 1 1 0 0
0 1 1 1 0
0×1 0×0 1×1 1 1
0×0 0×1 1×0 1 0
0×1 1×0 1×1 0 0
4 3 4
2 4 3
2
Image
Convolved Feature


1 1 1 0 0
0 1 1 1 0
0 0×1 1×0 1×1 1
0 0×0 1×1 1×0 0
0 1×1 1×0 0×1 0
4 3 4
2 4 3
2 3
Image
Convolved Feature


1 1 1 0 0
0 1 1 1 0
0 0 1 1 1
0 0 1 1 0
0 1 1 0 0
×1 ×0 ×1
×0 ×1 ×0
×1 ×0 ×1
4 3 4
2 4 3
2 3 4
Image
Convolved Feature



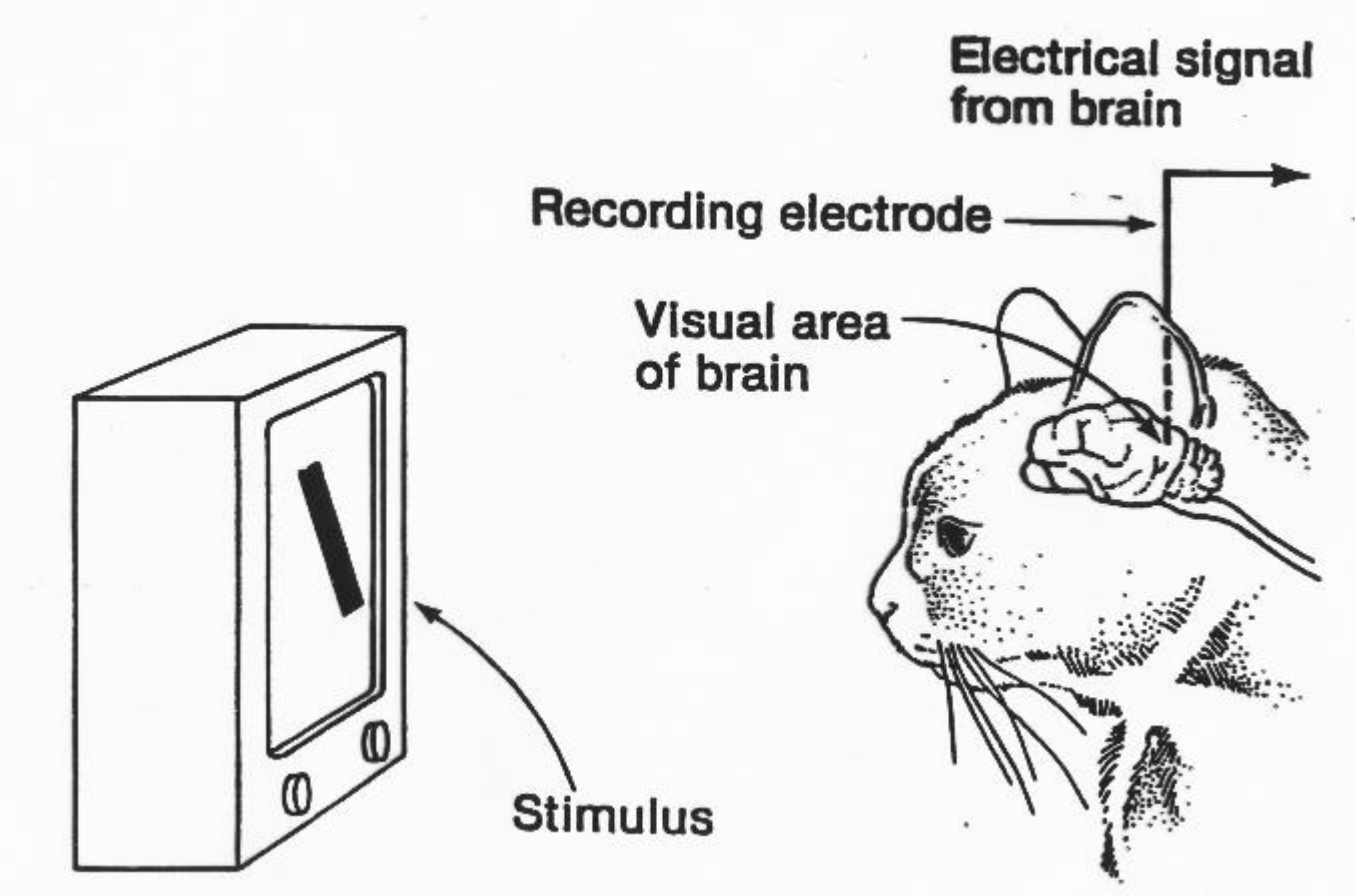
https://cs231n.github.io/convolutional-networks/

Input image

Convolution Kernel

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Feature map

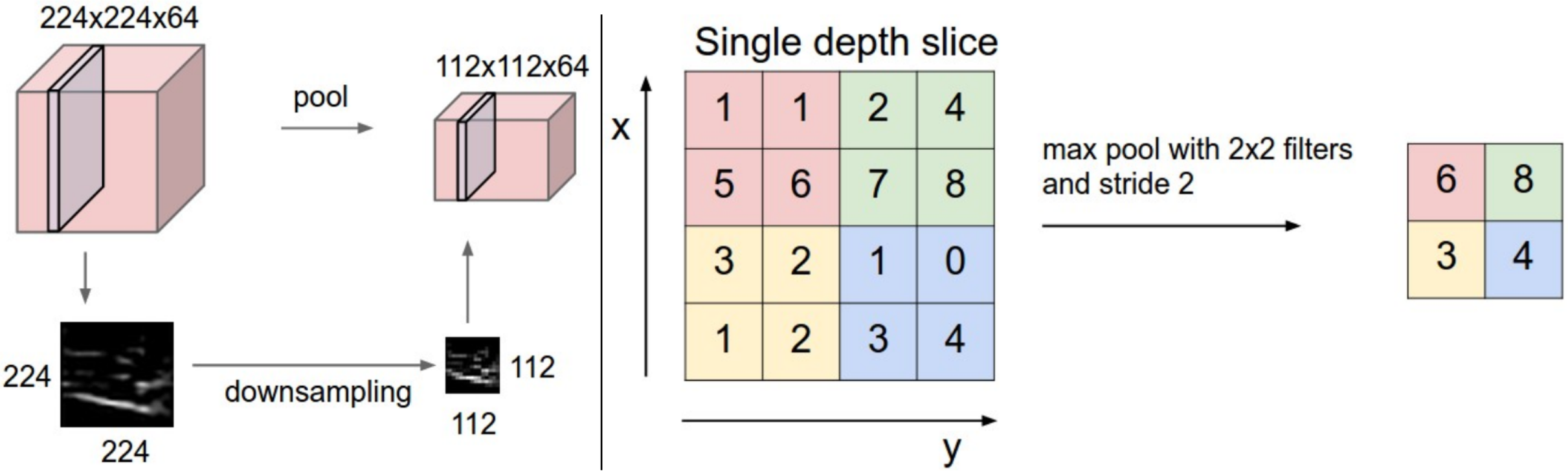


1. Spatial invariance
2. Receptive field
3. Hierarchy

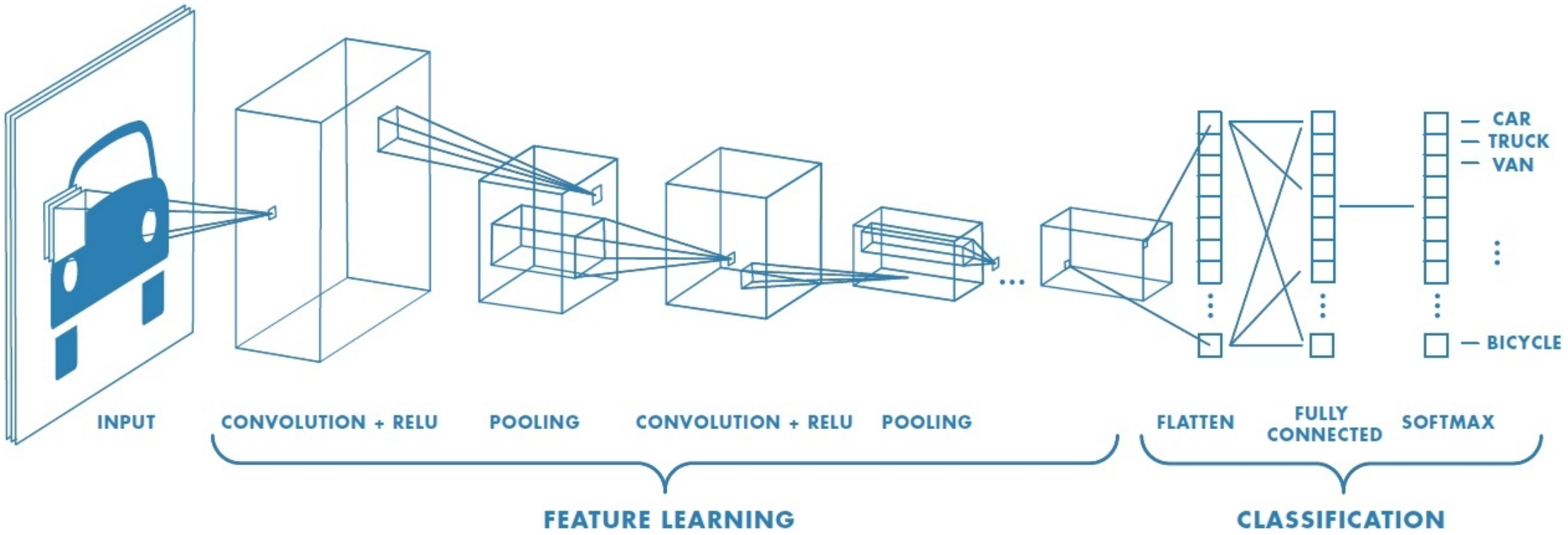
Visual cortex cell recording
Huebel & Wiesel
https://youtu.be/Cw5PKV9Rj3o?t=75

„Nothing seemed to work”
Huebel & Wiesel
https://youtu.be/IOHayh06LJ4?t=20

Pooling a dimenziócsökkentéshez

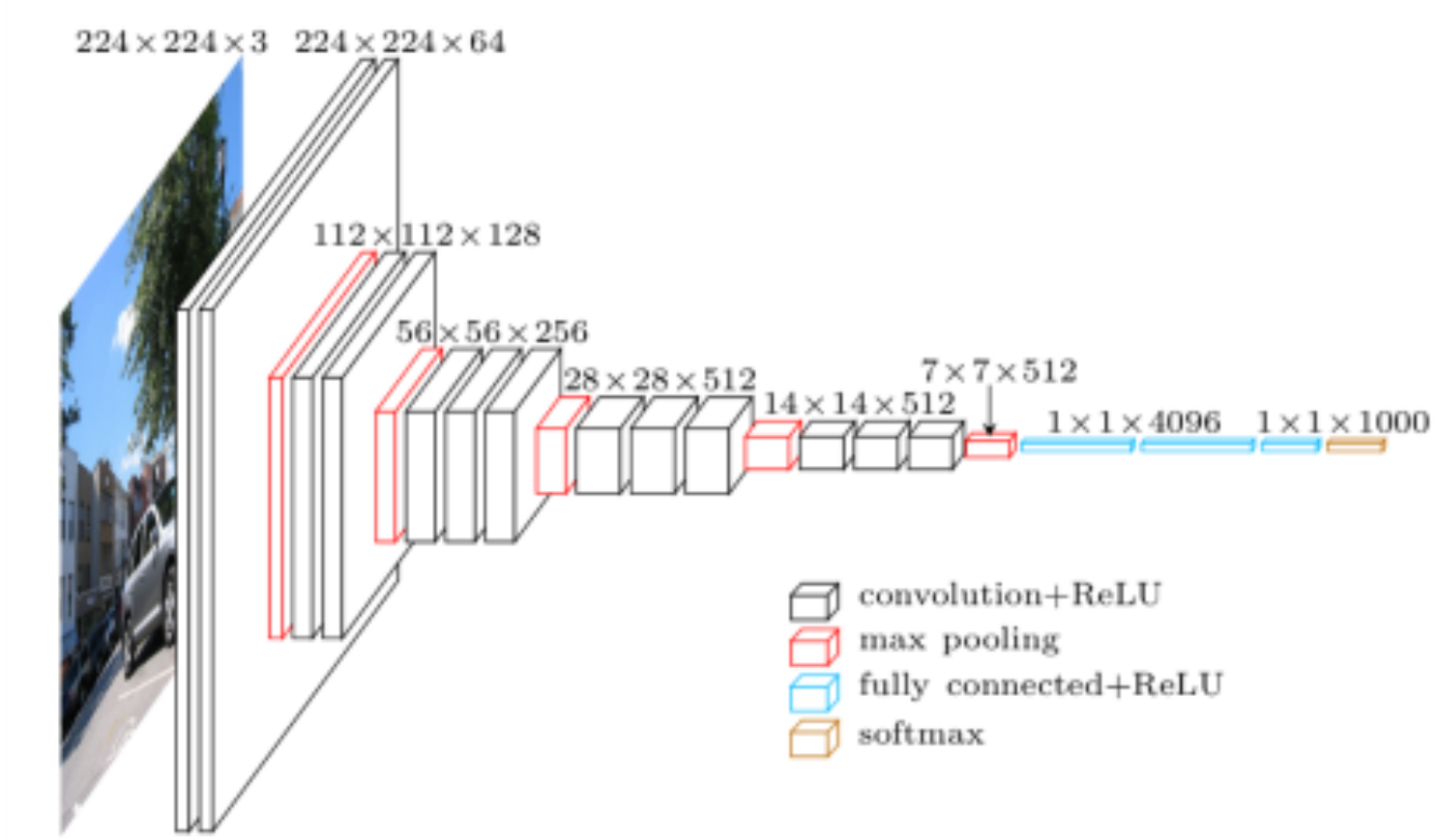


CNN általános séma

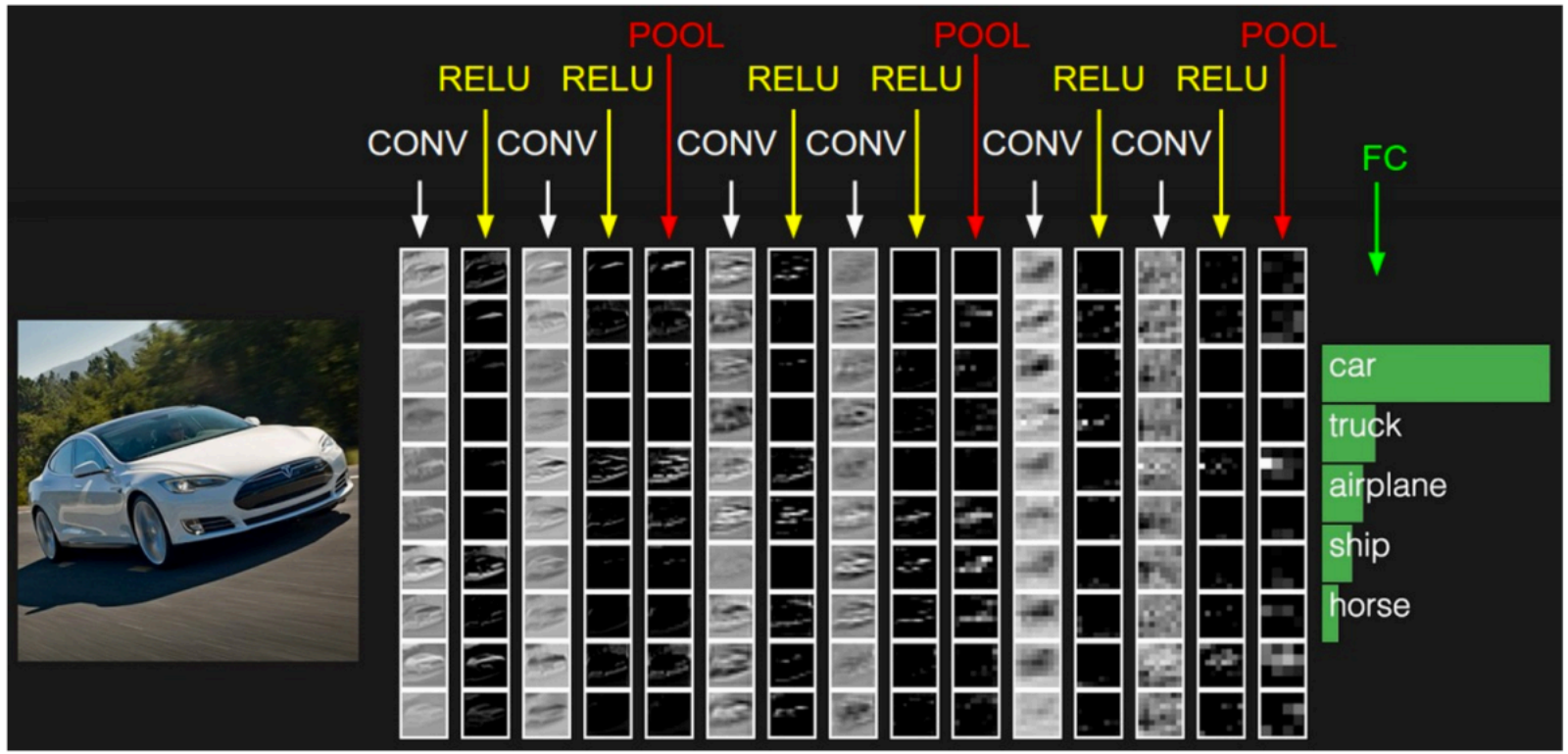


Sok-sok architektúra erre az általános sémára

- LeNet
- AlexNet
- GoogleNet
- VGGNet
- ResNet



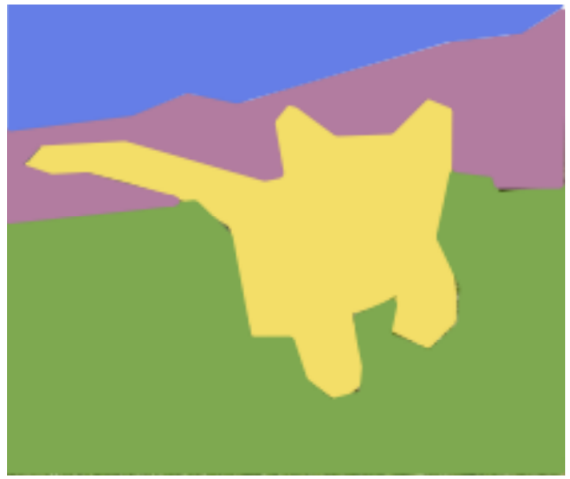
VGG-16



http://scs.ryerson.ca/~aharley/vis/conv/flat.html

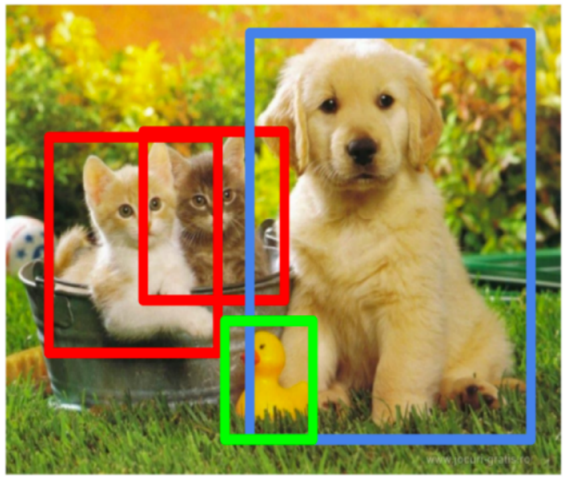
Ugyan az a konstrukció, sok-sok alkalmazás

Semantic Segmentation



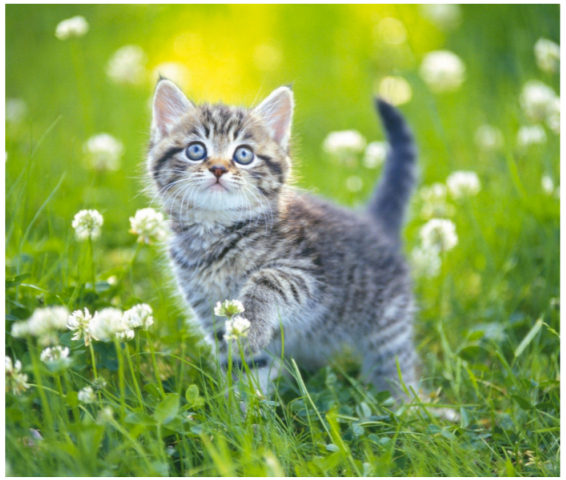
CAT

Object Detection



CAT, DOG, DUCK

Image Captioning

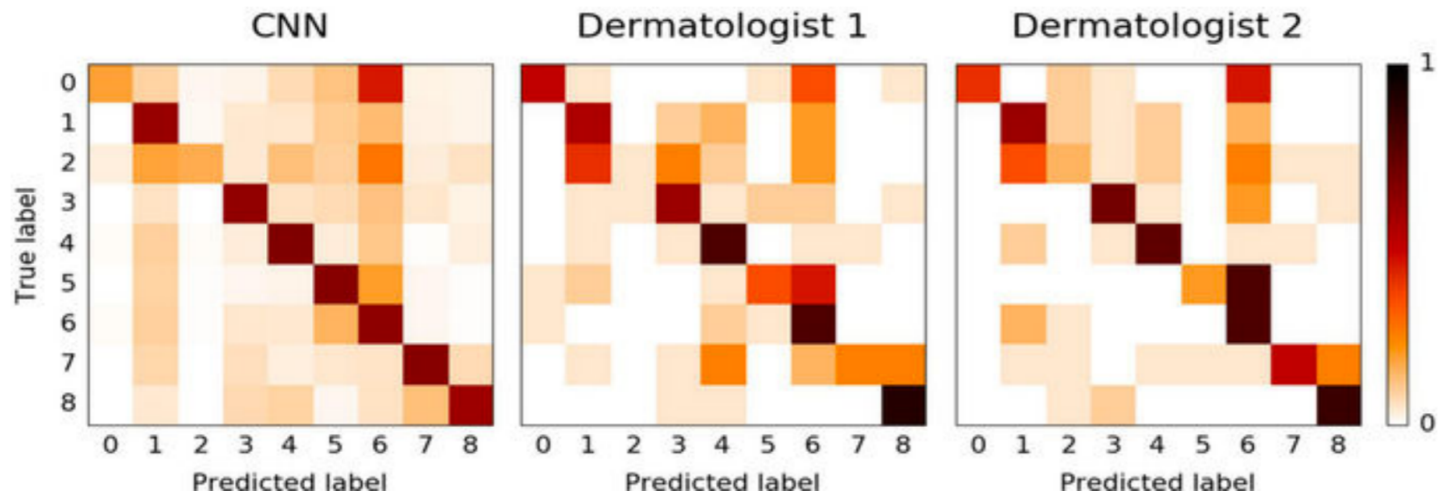


The cat is in the grass.

...

- https://www.youtube.com/watch?v=CxanE_W46ts

Dermatologist-level classification of skin cancer with deep neural networks



Transfer learning

Nem csak az architektúra, a súlyok is felhasználhatók más, a domain közelében lévő feladatokhoz

http://cs231n.github.io/transfer-learning/

Deep Image Prior

Dmitry Ulyanov, Andrea Vedaldi, Victor Lempitsky

CVPR, 2018

https://arxiv.org/abs/1711.10925

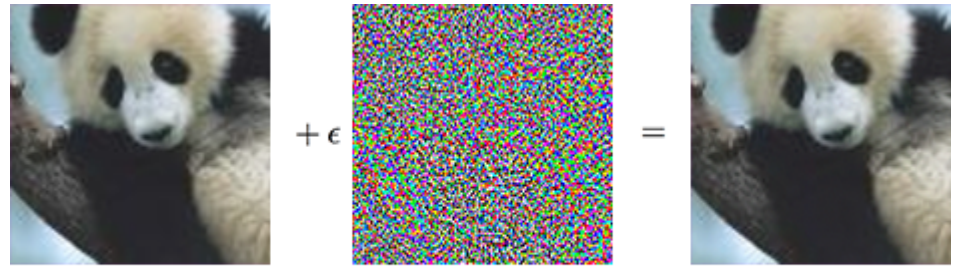
https://dmitryulyanov.github.io/deep_image_prior

Intriguing properties of neural networks

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. and Fergus, R.

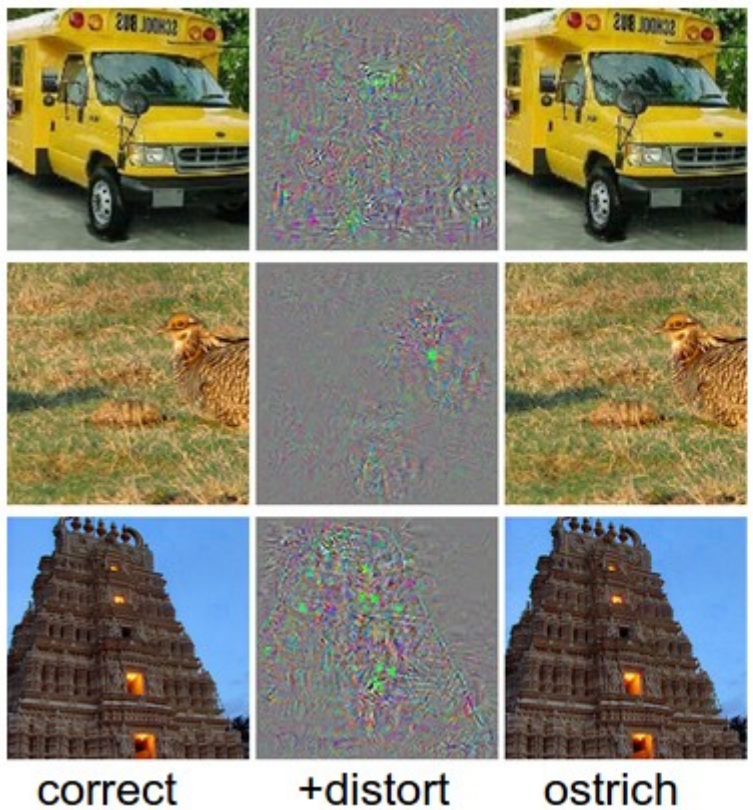
NIPS, 2014

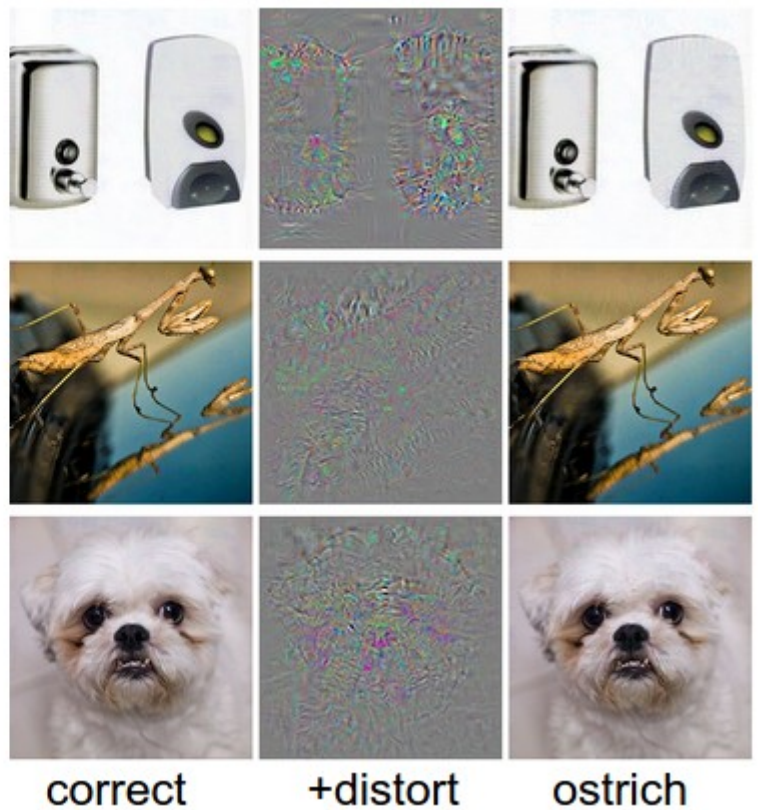
https://arxiv.org/abs/1312.6199



"panda"
57.7% confidence

"gibbon"
99.3% confidence





Measuring the tendency of CNNs to Learn Surface Statistical Regularities

Jason Jo, Yoshua Bengio

2017

https://arxiv.org/abs/1711.11561

APPROXIMATING CNNS WITH BAG-OF-LOCAL FEATURES MODELS WORKS SURPRISINGLY WELL ON IMAGENET

Brendel & Bethge

ICLR, 2019

https://openreview.net/pdf?id=SkfMWhAqYQ