Practical 3.0

Convolutional Neural Networks – Basics

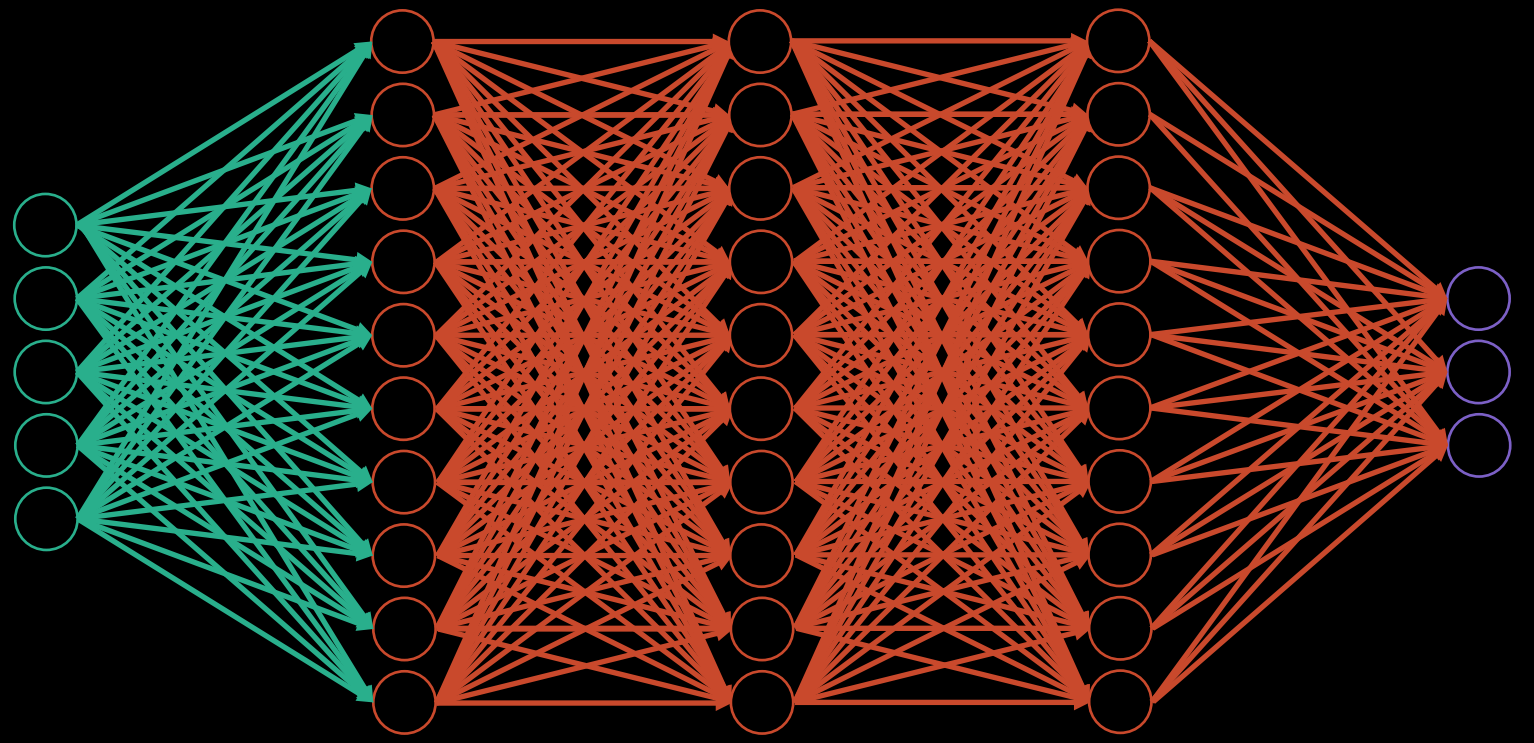
Overview (I)

- Rationale / motivation
- Sparsity (locality)
  - Receptive field
  - Hirarchical view
- Parameters sharing (omogeneity)
- Convolutional layer (3D conv.)
- Non linearity layer (decaying learning speed)
  - Logistic sigmoid
  - Rectifying linear unit

Overview (II)

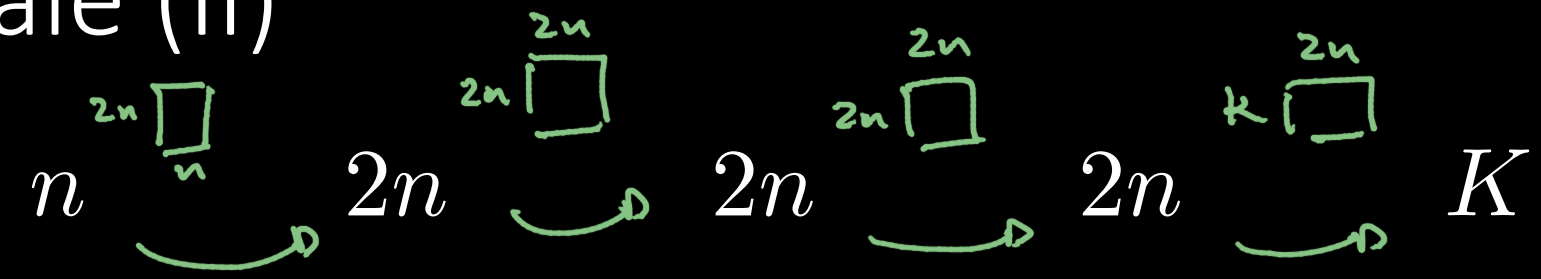
- L-p pooling layer
  - Average pooling
  - Max pooling
- Rationale / conclusion
  - Convolutional benefit
  - Pooling benefit
- Network colour coding with `pretty-nn`
- `e-lab/Torch7-profiling` tool
  - Profile CNN in Torch

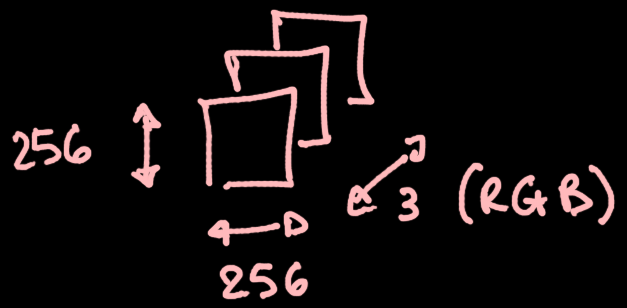
Rationale (I)



IMAGES → ↑ spatial correlation
→ objects are made of parts

Rationale (II)





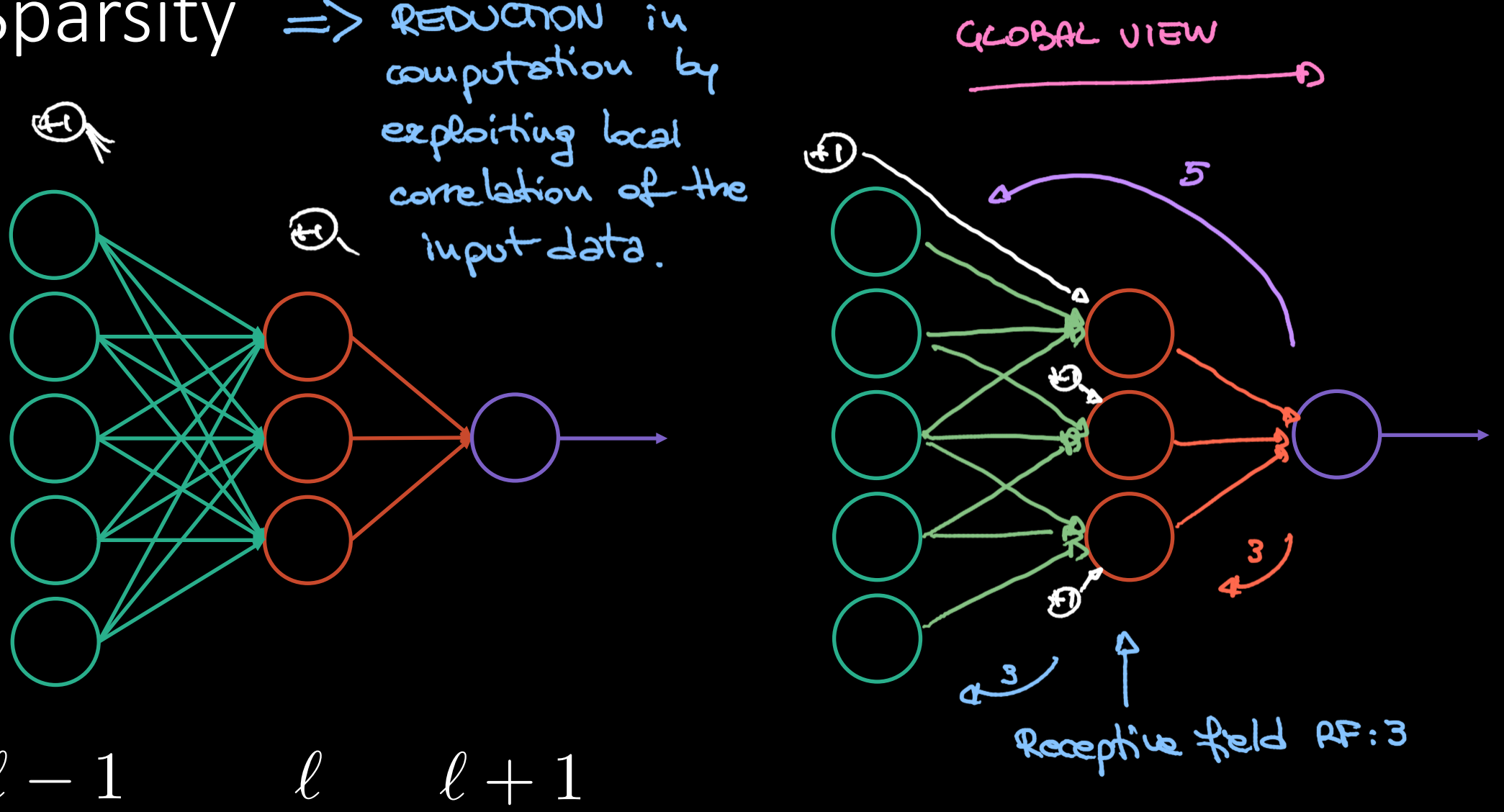
$n = 3 \cdot 256 \cdot 256 = 3 \cdot 2^8 \cdot 2^8 = 3 \cdot 2^{16}$

$7n + K \simeq 7 \cdot 3 \cdot 2^{16} \simeq 1{,}4\,M$

$2n^2 + 2 \cdot 4n^2 + 2Kn \simeq 10n^2 = 2 \cdot 5 \cdot 2^{33} \cdot 3^2 \simeq 390\,G\,MAC$ (Multiplication and accumulation)

MacBook PRO : Intel Core i7 2,2 GHz ⇒ $390\,G / 2{,}2\,GHz \simeq 180\,s \simeq 3\,min$

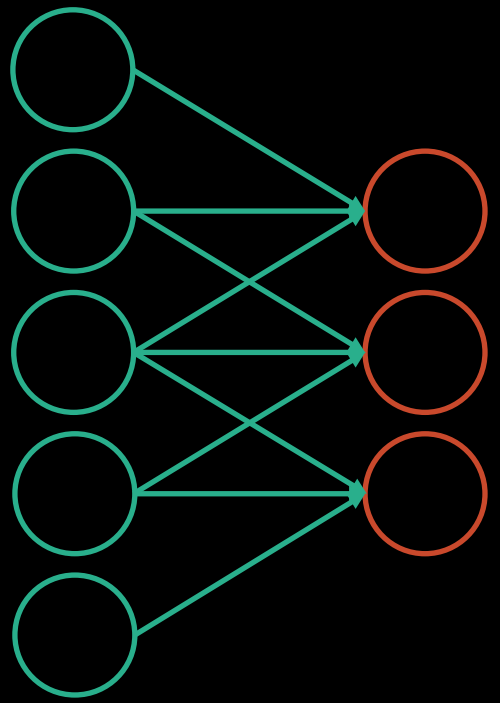
Sparsity

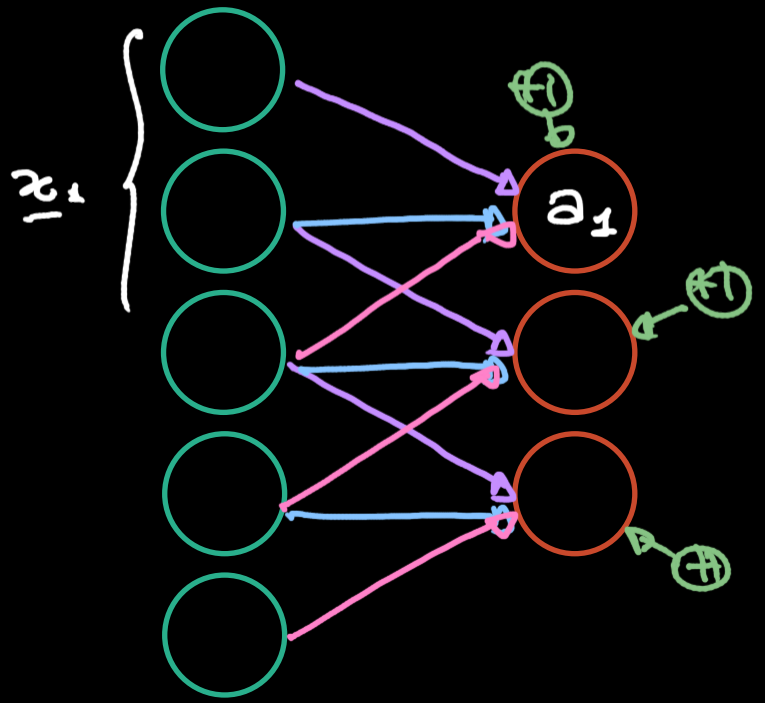


Parameters sharing =>

Reduce $\#\theta$ => helps convergence -> time ↓ -> error ↓

$$a_1 = \sigma(\theta^T x_1)$$





Convolutional layer (I)

256×256

x 3D: n_1 2D $n_2 \times n_3$ (image input => $n_1 = 3$, RGB)

y 3D: m_1 2D $m_2 \times m_3$ y_i feature maps ($m_2 \times m_3$)

k 4D: m_1 3D $n_1 \times p_1 \times p_2$ kernels k_i, $i = 1, \dots, m_1$

$\underline{b}$ 1D: m_1 bias term

CONVOLUTION

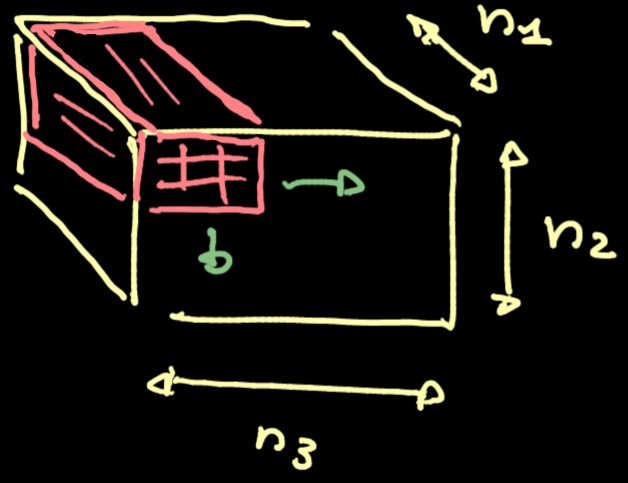
$$y_i = \underline{x} \circledast \underline{k_i} + \underline{b_i}, \quad i = 1, \dots, m_1$$

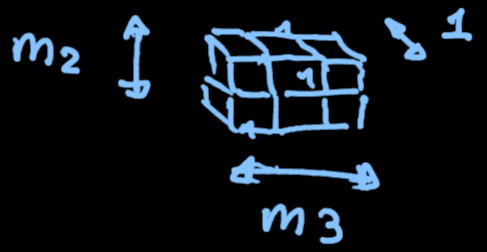
$(\underline{z} = \Theta \cdot \underline{a}$ or $\Theta \cdot \underline{x})$

Convolutional layer (II)

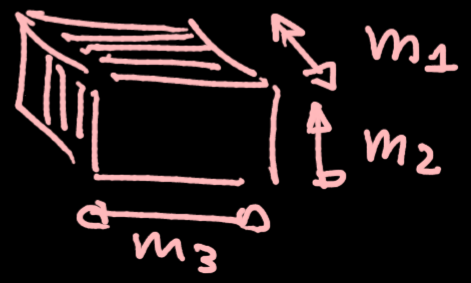
$$y_i = x \star k_i + b_i, \quad i = 1, 2, \cdots, m_1$$

$$f[l,m,n] \star g[l,m,n] = \sum_u \sum_v \sum_w f[u,v,w] \cdot g[l-u, m-v, n-w]$$



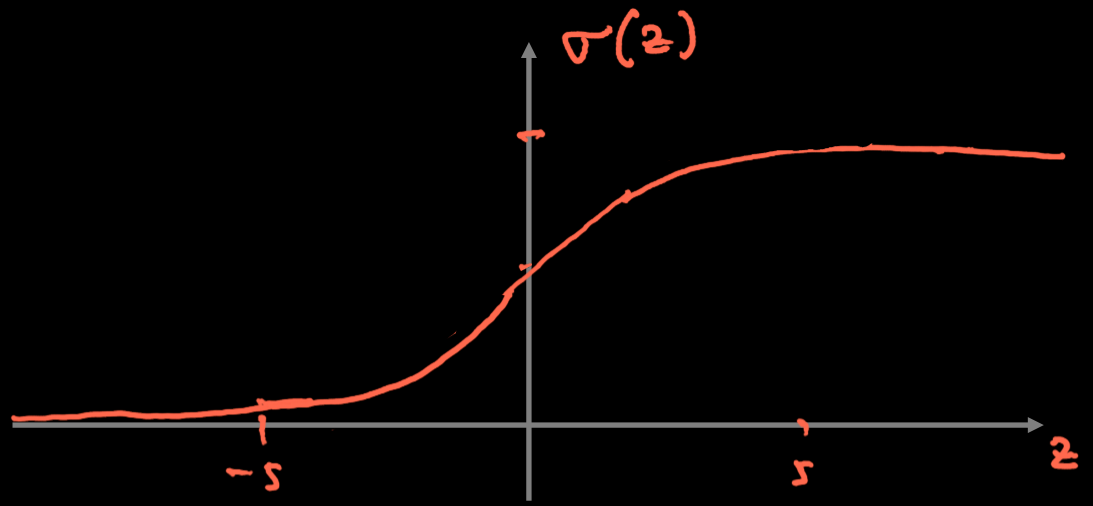


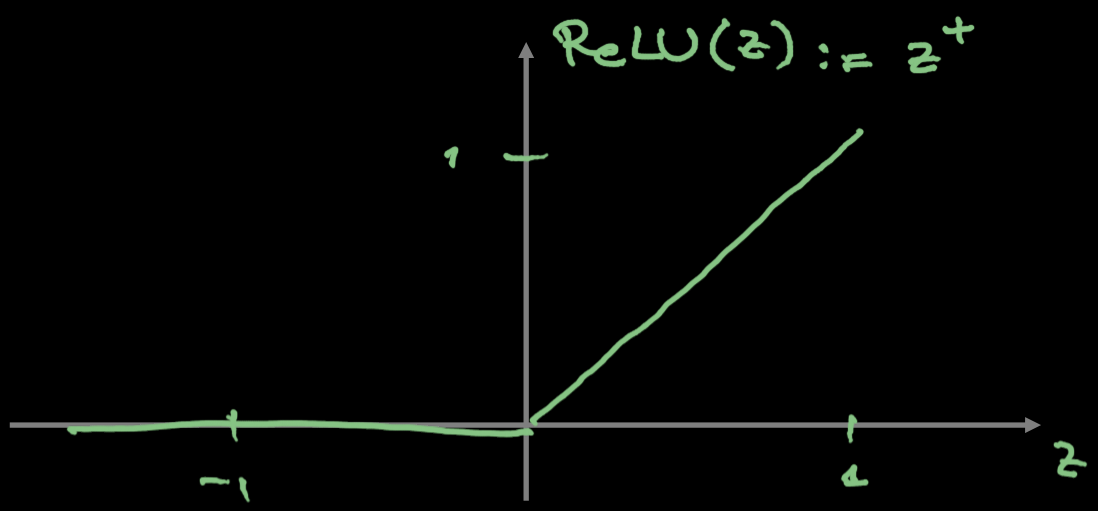
X m_1 times



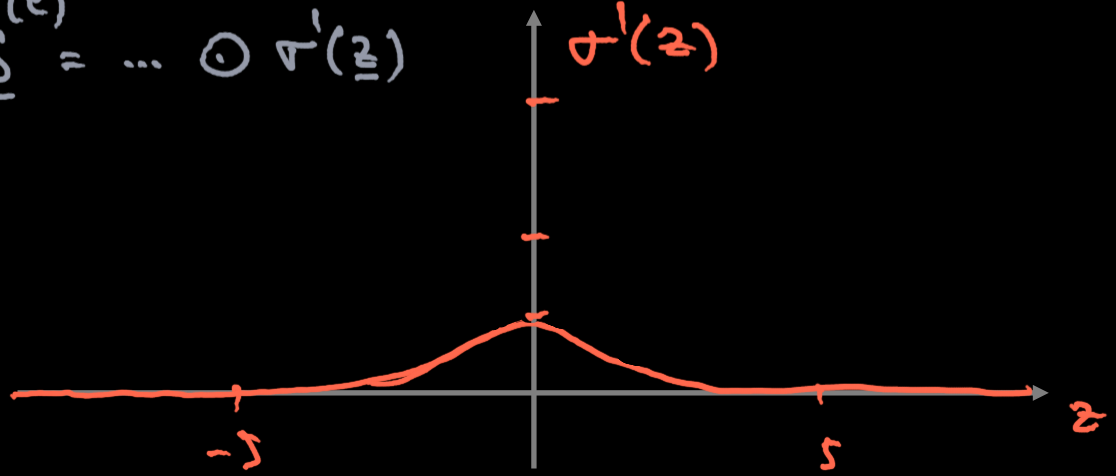
Non linearity

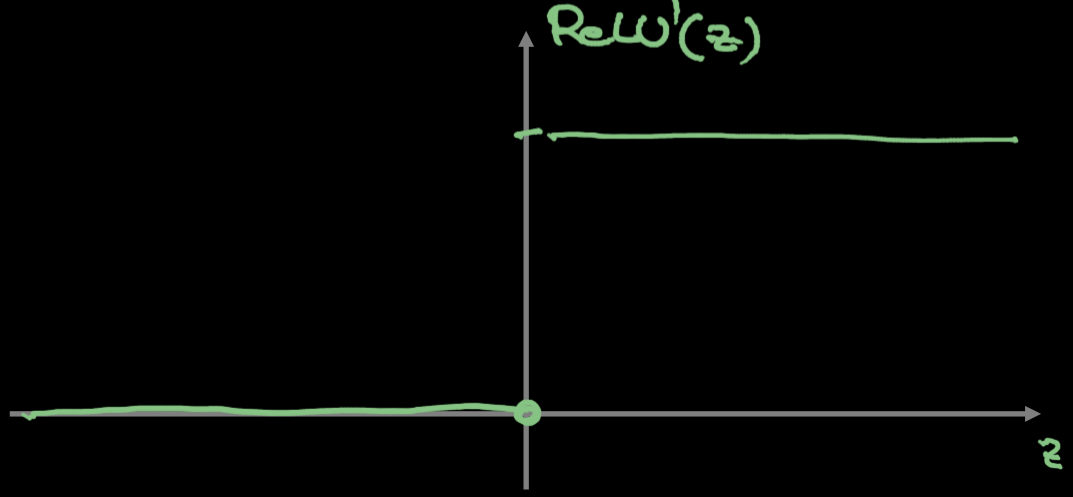
$\underline{z} = \ominus\, \underline{\hat{a}}^{(\ell-1)}$ projection by conv. $\rightsquigarrow$ $\underline{a} = \sigma(\underline{z})$ feature map.



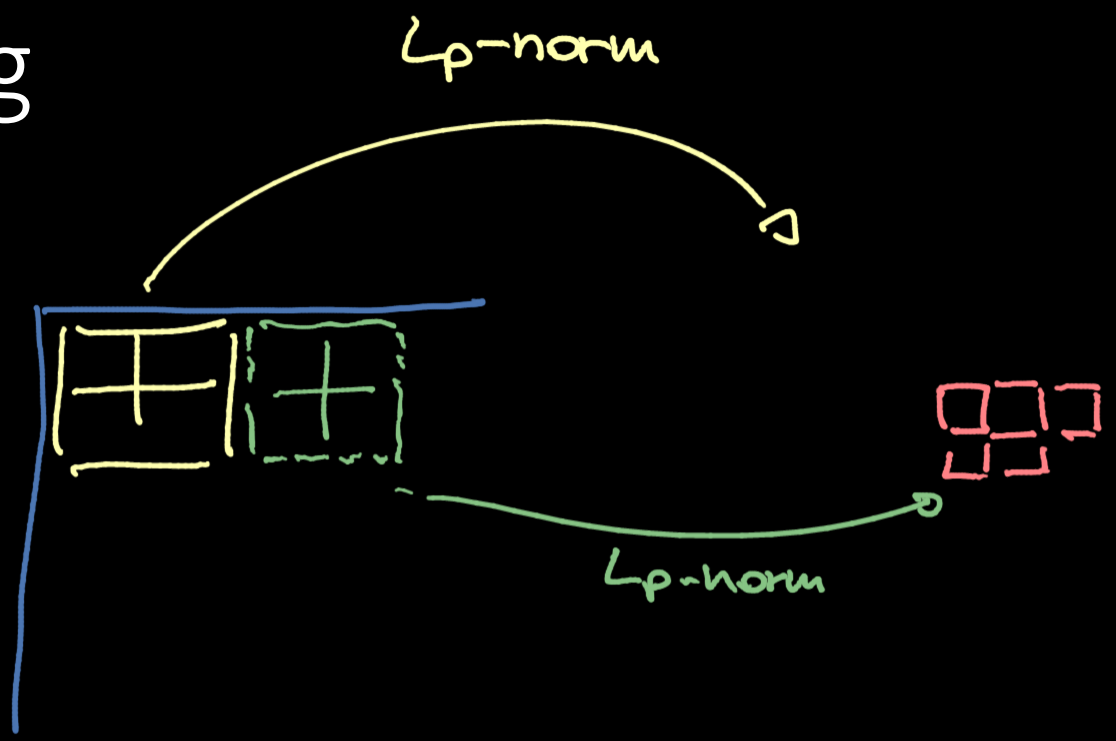


$\underline{\delta}^{(\ell)} = \ldots \odot \sigma'(\underline{z})$



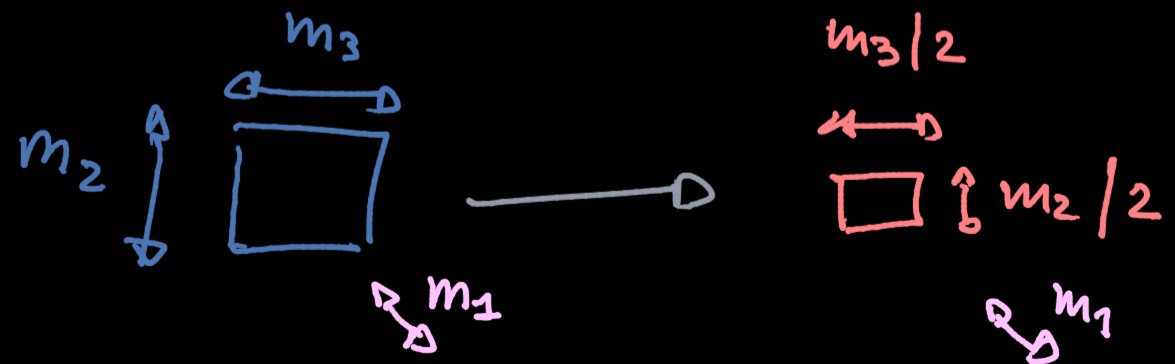


Pooling



$$\| \underline{x} \|_p = \left(\sum_i |x_i|^p \right)^{1/p}$$

$$\| \underline{x} \|_p \to \max(\underline{x}), \; p \to +\infty$$

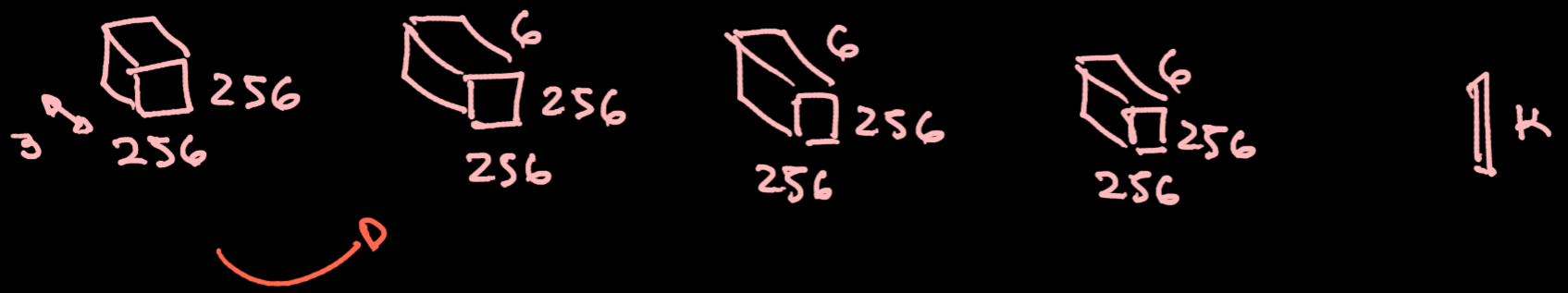


Rationale (III)

$3 \mid 6$

$k_i : 5 \times 5$

n $2n$ $2n$ $2n$ K



$$(256 \times 256 \times 6 \times 3 \times 5 \times 5) \times \underbrace{(1+2+2)}_{5} = 147\text{M MAC}$$

\# activations

540 M MAC

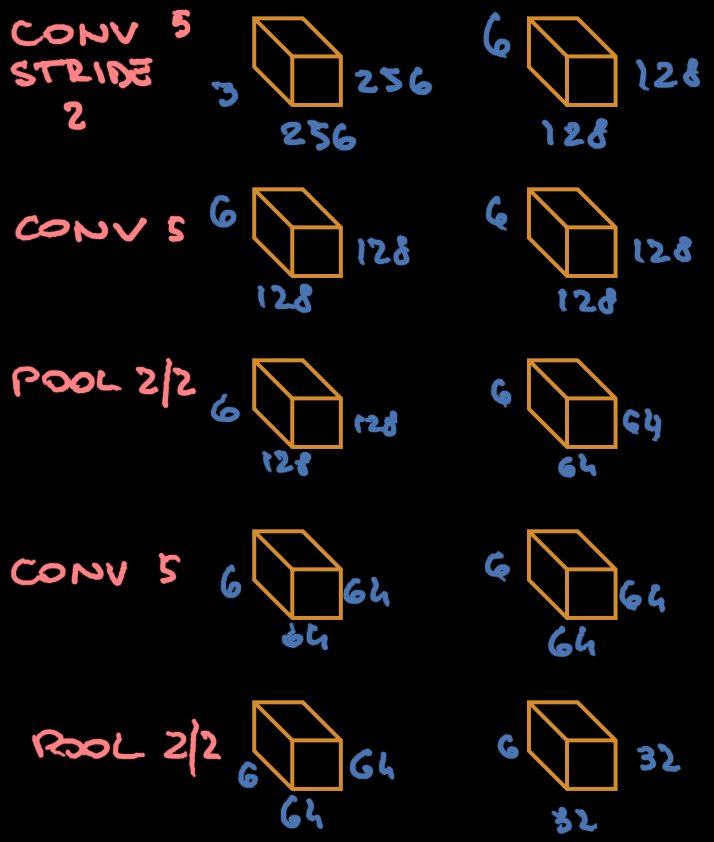
/ 2.2 GHz → 250ms

→ $K \cdot 6 \cdot 256 \cdot 256 = K \cdot 393k = 393M$

ImageNet : $K = 1k$

Rationale (IV)

31,7M MAC / 2.2GHz = 14 ms



$128 \times 128 \times 6 \times 5 \times 5 \times 3 \simeq 7{,}3M$ MAC

98k NL MAC

25,7M CONV

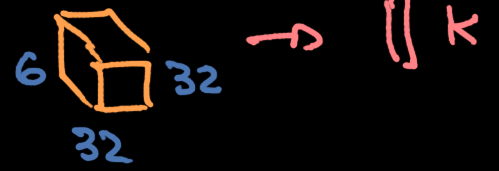
$128 \times 128 \times 6 \times 5 \times 5 \times 6 \simeq 14{,}7M$ MAC

$64 \times 64 \times 6 \times 2 \times 2 \simeq 98k$ MAC

$64 \times 64 \times 6 \times 5 \times 5 \times 6 = 3{,}7M$ MAC

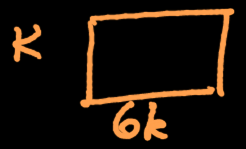
$32 \times 32 \times 6 \times 2 \times 2 \simeq 25k$ MAC

LINEAR



$32 \times 32 \times 6 = 6k$

K = 1k ImageNet



6M MAC