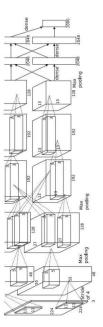
## Case Study: AlexNet

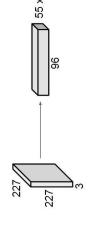
[Krizhevsky et al. 2012]



Input: 227x227x3 images

First layer (CONV1): 96 11x11 filters applied at stride 4 => Output volume [55x55x96]

Τεχνητή Νοημοσύνη



## Architectures

## Case Study; AlexNet

[Krizhevsky et al. 2012]

Architecture: CONV1 — Rel U. MAX POOL1

CONV2 MAX POOL2 ← **NORM2** 

W' = (W - F + 2P) / S + 1

CONV3 CONV5 SONV4

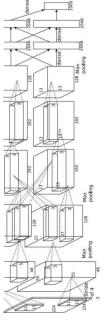
Max POOL3 ⊕ FC6 > FC7 FC8

75

### Architectures

# Case Study: AlexNet

[Krizhevsky et al. 2012]



First layer (CONV1): 96 11x11 filters applied at stride 4

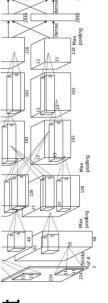
Input: 227x227x3 images

Output volume [55x55x96]

## Architectures

## Case Study: AlexNet

[Krizhevsky et al. 2012]



Input 227x227x3 images

First layer (CONV1):(96(11x11) filters applied at stride 4

W = (W - F + 2P) / S + 1

Q: what is the output volume size? Hint: (227-11)/4+1 (55)

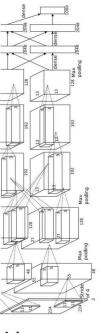
1×11×10 × 06

Q: What is the total number of parameters in this layer?

Τεχνητή Νοημοσύνη

## Case Study: AlexNet

[Krizhevsky et al. 2012]



Input: 227x227x3 images

After CONV1: 55x55x96

W' = (W - F + 2P) / S + 1

**Second layer** (POOL1): 3x3 filters applied at stride 2 Output volume: 27x27x96

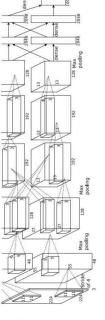
Q: what is the number of parameters in this layer?

Τεχνητή Νοημοσύνη

### Architectures

# Case Study: AlexNet

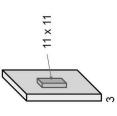
[Krizhevsky et al. 2012]



Input: 227x227x3 images

First layer (CONV1): 96 11x11 filters applied at stride 4

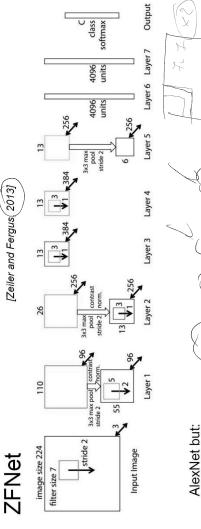
Parameters: (11\*11\*3 + 1)\*96 ≠ 35K Output volume [55x55x96]



Τεχνητή Νοημοσύνη

http://www.vvr.ece.upatras.gr

### Architectures

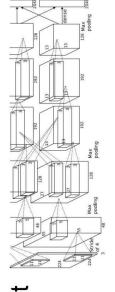


ImageNet top 5 error: 16.4% -> 11.7% CONV1: change from (11x11) stride 4) to (7x7) stride 2) CONV3,4,5: instead of 384, 384, 256 filters use 512, 1024, 512

### Architectures

# Case Study: AlexNet

[Krizhevsky et al. 2012]



Input: 227x227x3 images After CONV1:\55x55x96

Second layer (POOL1)/ 3x3 filters/applied at stride 2

Q: what is the output volume size? Hint:  $(55-3)/2+1 \neq 27$ 

27×CT 49C

W' = (W - F + 2P) / S + 1

Τεχνητή Νοημοσύνη

Τεχνητή Νοημοσύνη

http://www.vvr.ece.upatras.gr

## Case Study: ResNet

He et al., 2015]

Fact: Deep models have more representation power (more parameters) than shallower models.

Hypothesis: the problem is an optimization problem, deeper models are harder to optimize

What should the deeper model learn to be at least as good as the shallower model?

A solution by construction is copying the learned layers from the shallower model and setting additional layers to identity mapping.

Τεχνητή Νοημοσύνη



http://www.vvr.ece.upatras.gr



### Architectures

## Case Study: ResNet

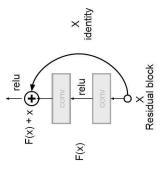
[He et al., 2015]

Very deep networks using residual connections

- **ILSVRC'15** classification winner 152-layer model for ImageNet (3.57% top 5 error)
  - Swept all classification and ILSVRC'15 and COCO'15! detection competitions in

<u>e</u>

 $\stackrel{(\times)}{\vdash}$ 









83

### Τεχνητή Νοημοσύνη

### http://www.vvr.ece.upatras.gr

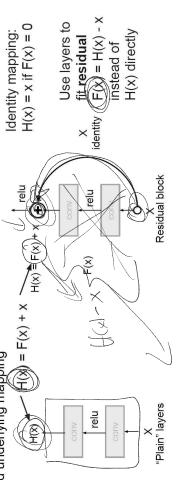
**Architectures** 

## Case Study: ResNet

(He et al., 2015)

20 8 R

Solution: Use network layers to fit a residual mapping instead of directly trying to fit a desired underlying mapping



## Architectures

## Case Study: ResNet

[He et al., 2015]

What happens when we continue stacking deeper layers on a "plain" convolutional neural network?



56-layer model performs worse on both test and training error

-> The deeper model performs worse, but it's not caused by overfitting

Τεχνητή Νοημοσύνη

84

http://www.vvr.ece.upatras.gr

Τεχνητή Νοημοσύνη



http://www.vvr.ece.upatras.gr

Architectures

[He et al., 2015]

No FC layers besides FC 1000 to output classes

average pooling layer after last conv layer

Global

## Case Study: ResNet

[He et al., 2015]

Full ResNet architecture:

Stack residual blocks

relu

F(x) + x

- Every residual block has two 3x3 conv layers
- Periodically, double # of filters and downsample spatially using stride 2 (/2 in each dimension)

identity

le Le

E(×)

No FC layers at the end (only FC 1000 to output Additional conv layer at the beginning (stem)

Residual block

(In theory, you can train a ResNet with input image of variable sizes) classes)

Τεχνητή Νοημοσύνη

http://www.vvr.ece.upatras.gr

Τεχνητή Νοημοσύνη

volume by half.

3x3 conv, 128 filters, /2 spatially with stride 2 3x3 conv, 64 filters identity Residual block relu relu F(x) + x E(X Case Study: ResNet Every residual block has Periodically, double # of filters and downsample spatially using stride 2 /2 in each dimension) Reduce the activation Stack residual blocks two 3x3 conv layers Full ResNet architecture:

### Architectures

## Case Study: ResNet

[He et al., 2015]

 Stack residual blocks Full ResNet architecture:

- Every residual block has
  - Periodically, double # of filters and downsample spatially using stride 2 two 3x3 conv layers
- No FC layers at the end Additional conv layer at (/2 in each dimension) the beginning (stem)
- (In theory, you can train a ResNet with input image of variable sizes) classes)

Τεχνητή Νοημοσύνη

(only FC 1000 to output

### Global average pooling layer after last conv layer identity Residual block relu 들 $\mathbf{F}(\mathbf{x}) + \mathbf{x} + \mathbf{\Phi}$ E(X

http://www.vvr.ece.upatras.gr

## Architectures

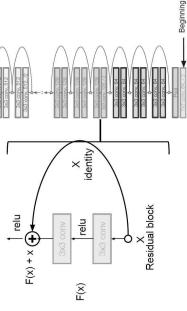
### Case Study: ResNet [He et al., 2015]

No FC layers besides FC 1000 to output classes

- Full ResNet architecture:
- Every residual block has two 3x3 conv layers

Stack residual blocks

- Periodically, double # of filters and downsample spatially using stride 2 /2 in each dimension)
- Additional conv layer at the beginning (stem)



Τεχνητή Νοημοσύνη

88

http://www.vvr.ece.upatras.gr