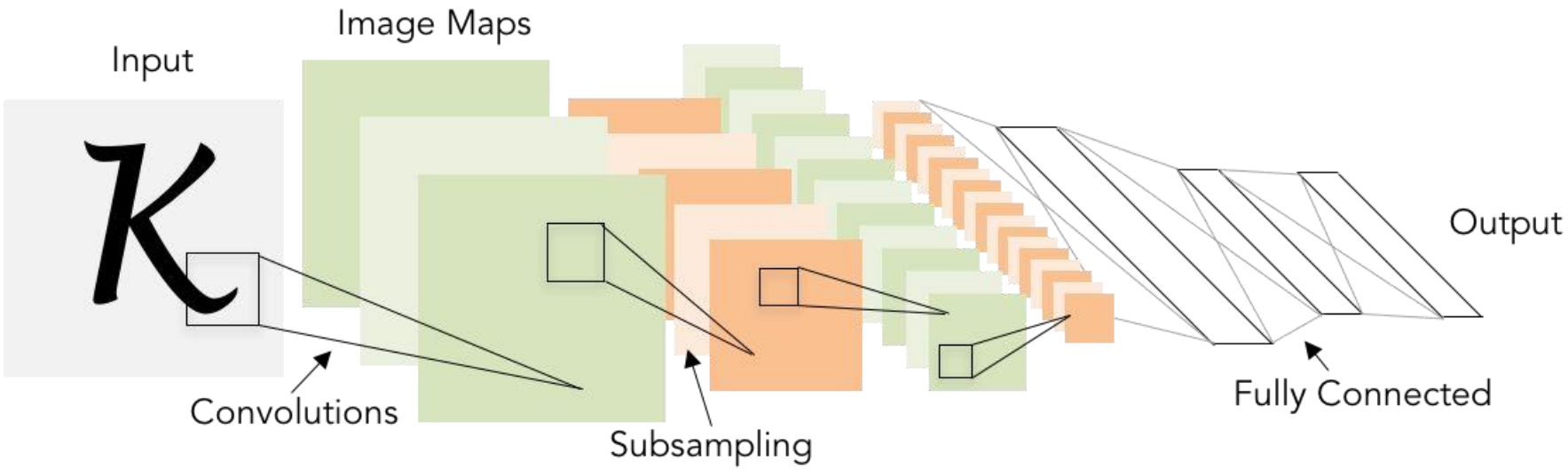
Various CNN Architectures

Content

- **Introduction**
- **VGG**
- **Resnet**

Introduction

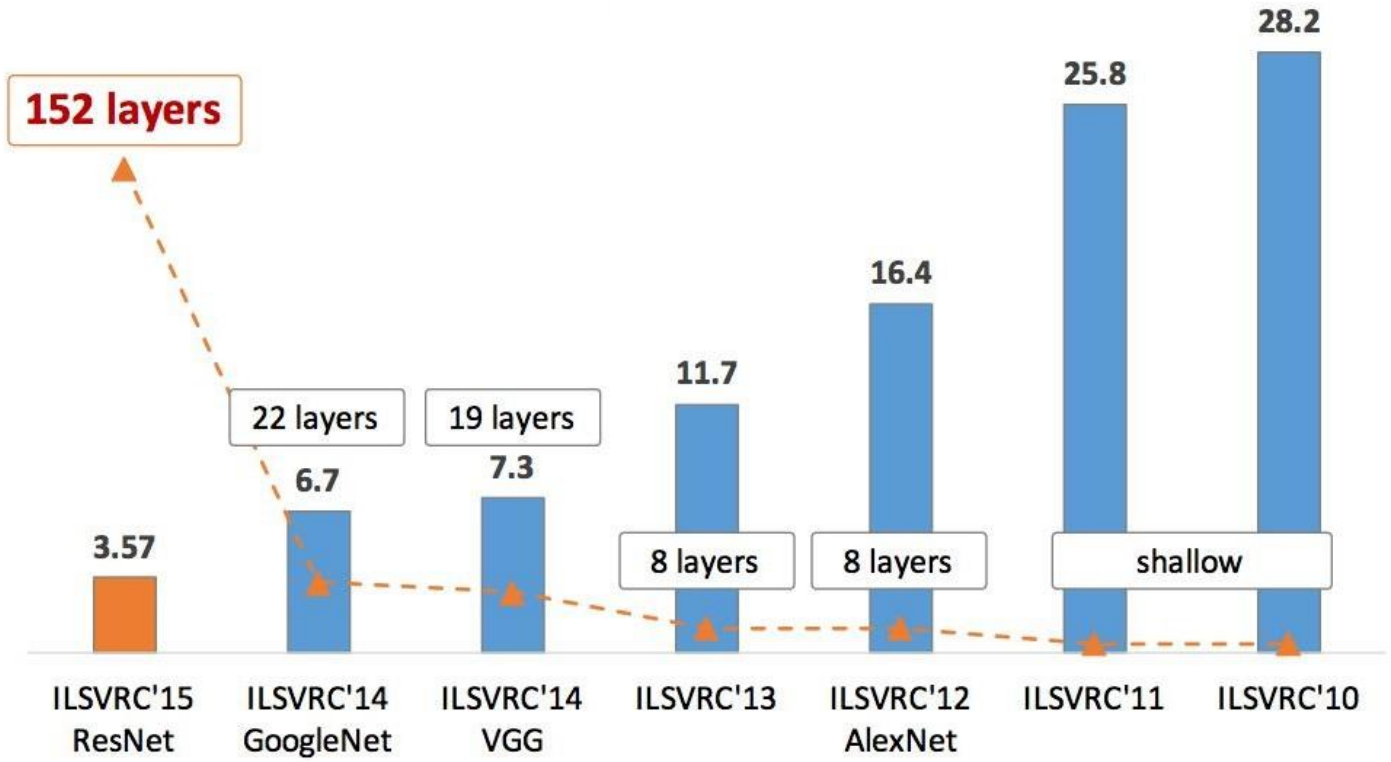
- **LeNet-5**
  - LeCun et al., 1998



Conv filters were 5x5, applied at stride 1
Subsampling (Pooling) layers were 2x2 applied at stride 2
i.e. architecture is [CONV-POOL-CONV-POOL-FC-FC]

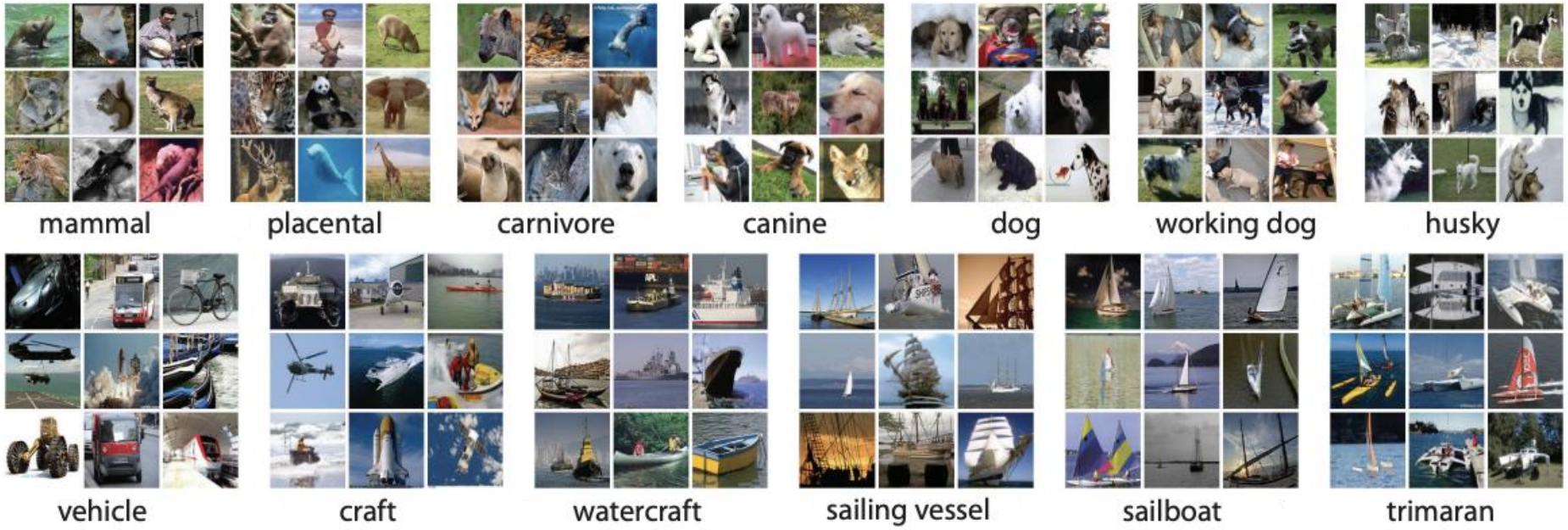
Introduction

- **ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners**



Introduction

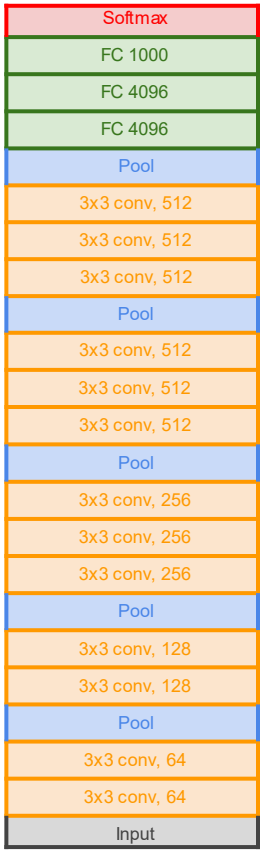
- **ImageNet Dataset**
  - Images from the real world
  - 1000 classes and about 1000 images per class



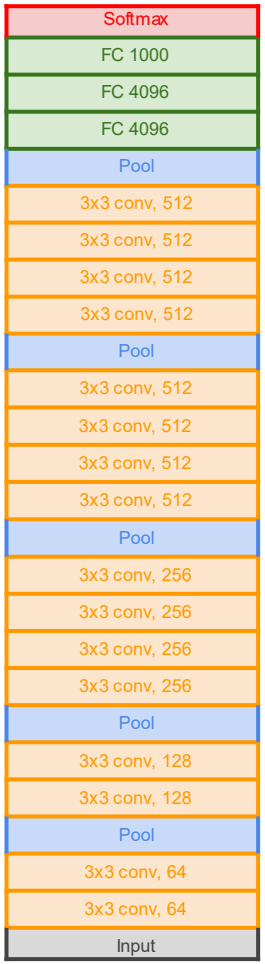
VGGNet

- **Small filters, Deeper networks**
  - Only 3x3 CONV stride 1, pad 1 and 2x2 MAX POOL stride 2

AlexNet



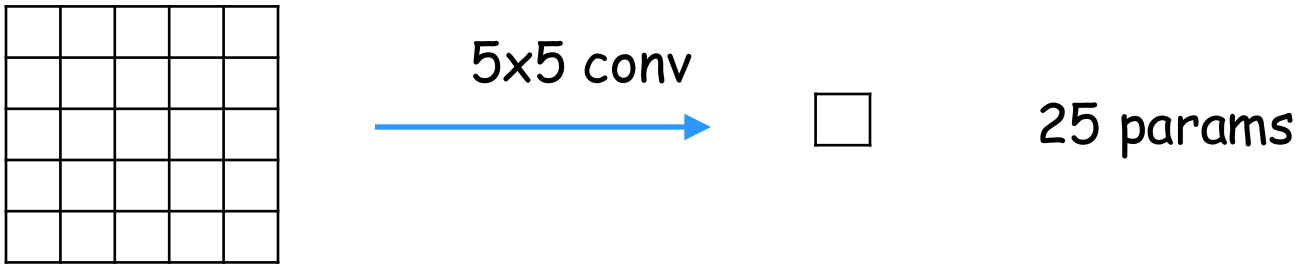
VGG16



VGG19

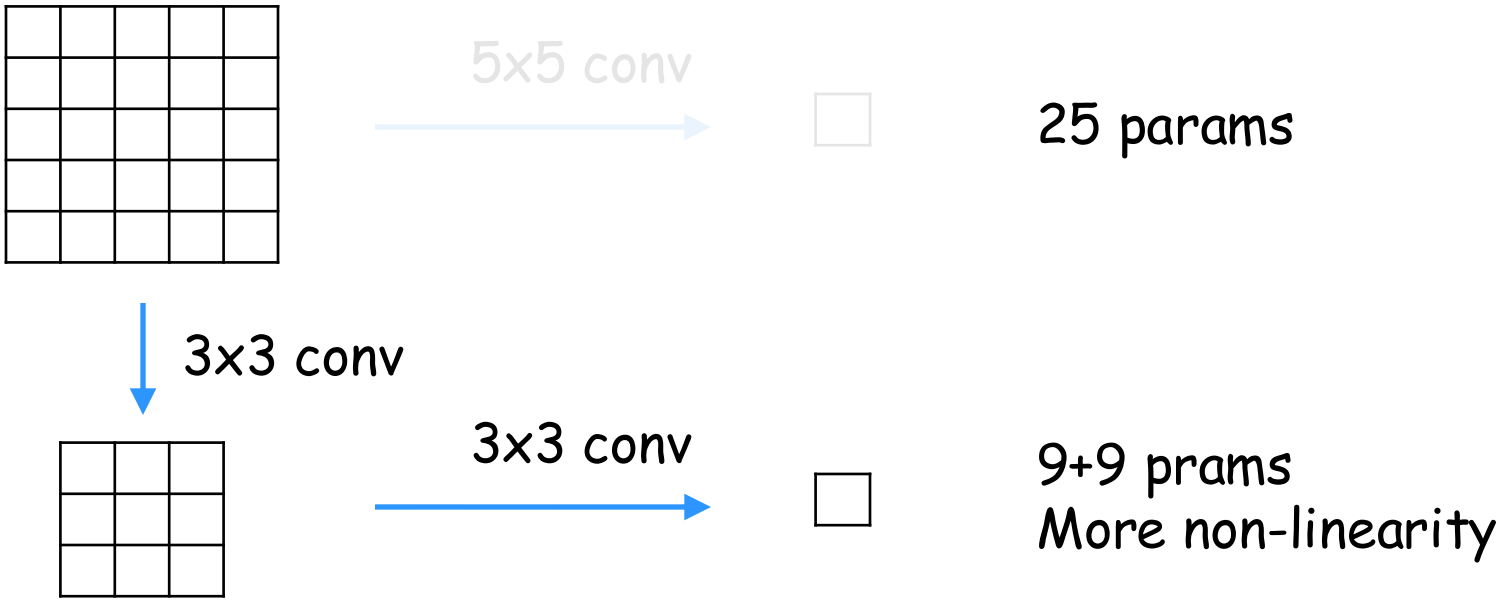
VGGNet

- **Large Filters vs Small Filters**



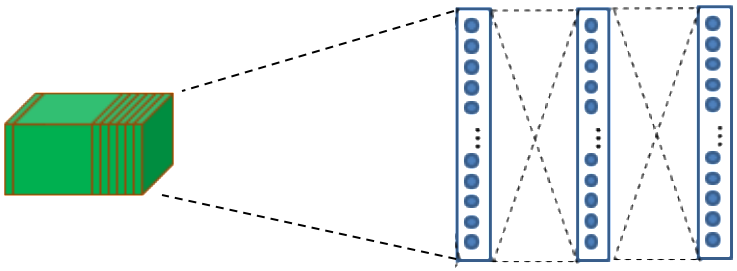
VGGNet

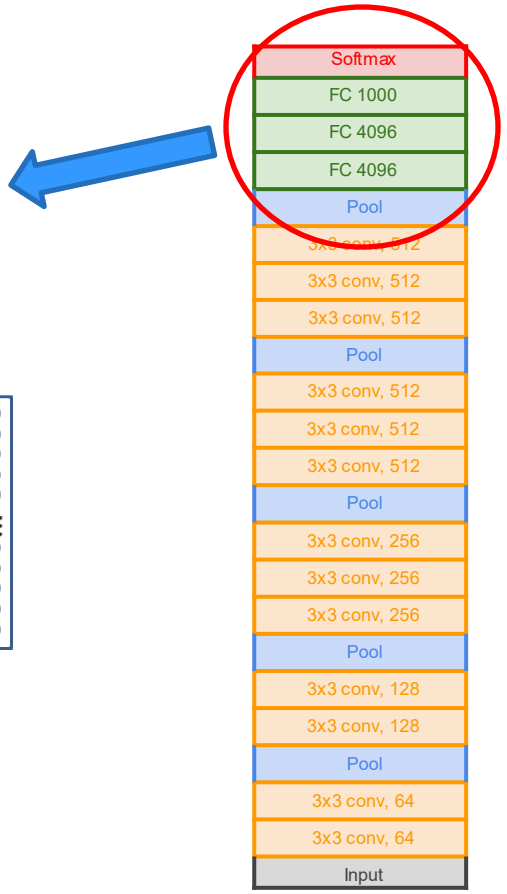
- **Large Filters vs Small Filters**



VGGNet

- **Too many parameters**
  - Especially in FC layers

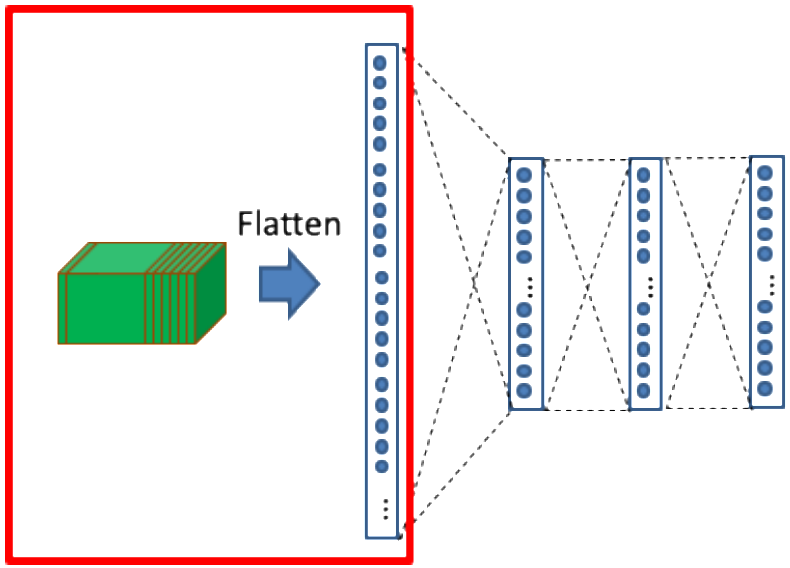


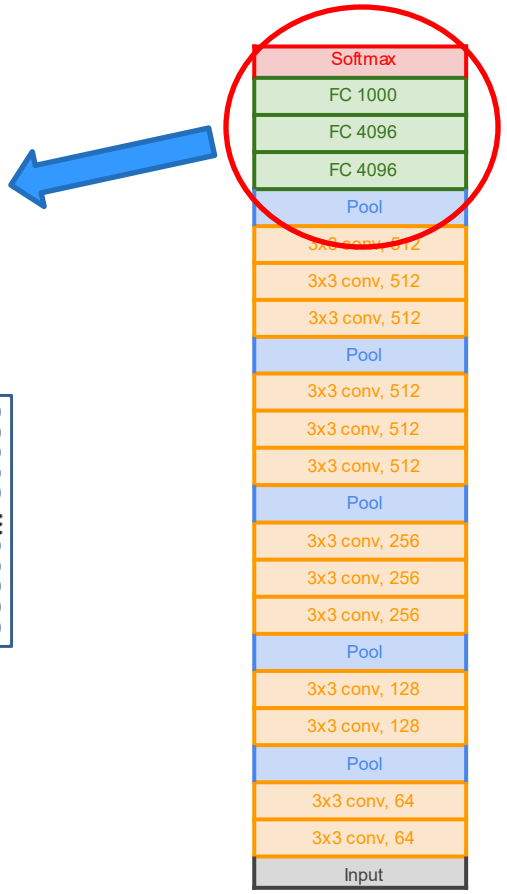


VGG16

VGGNet

- **Too many parameters**
  - Especially in FC layers

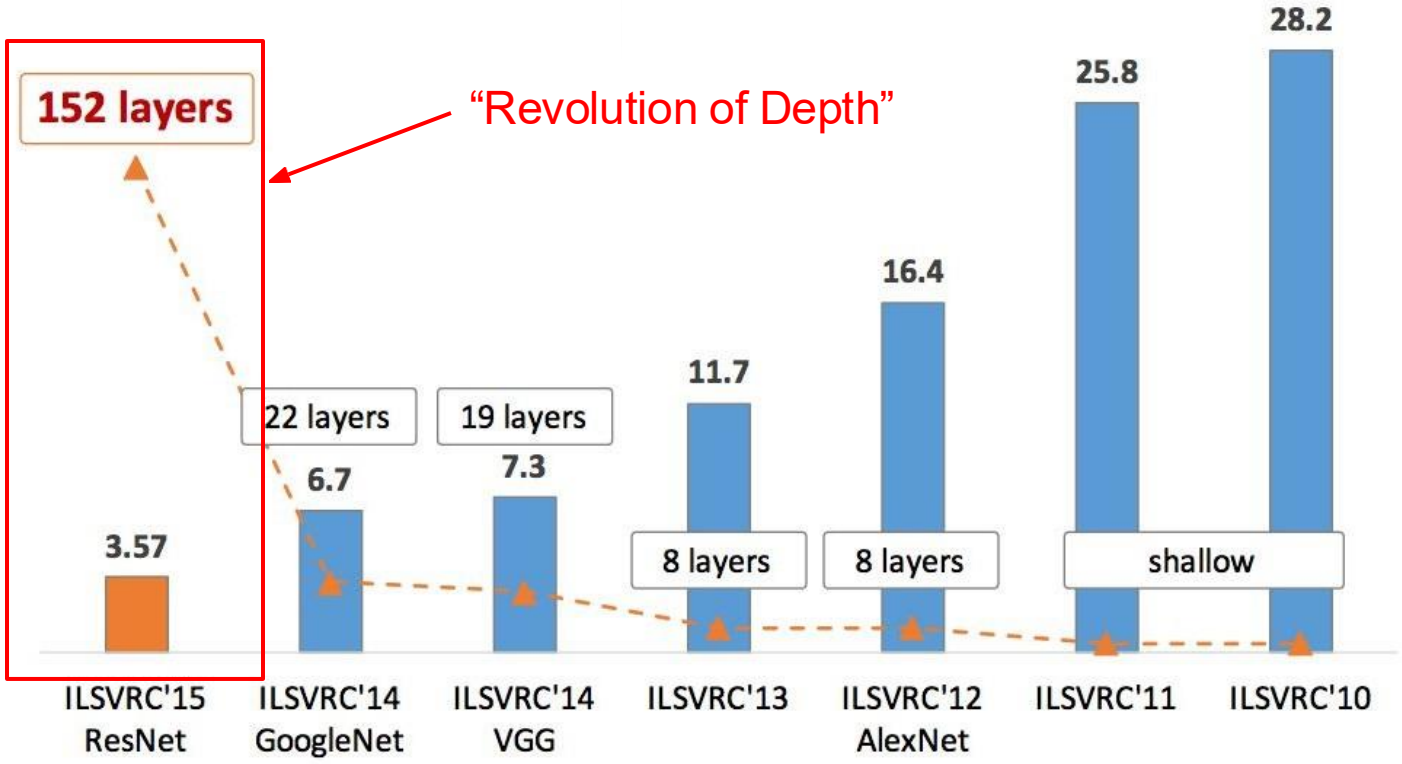




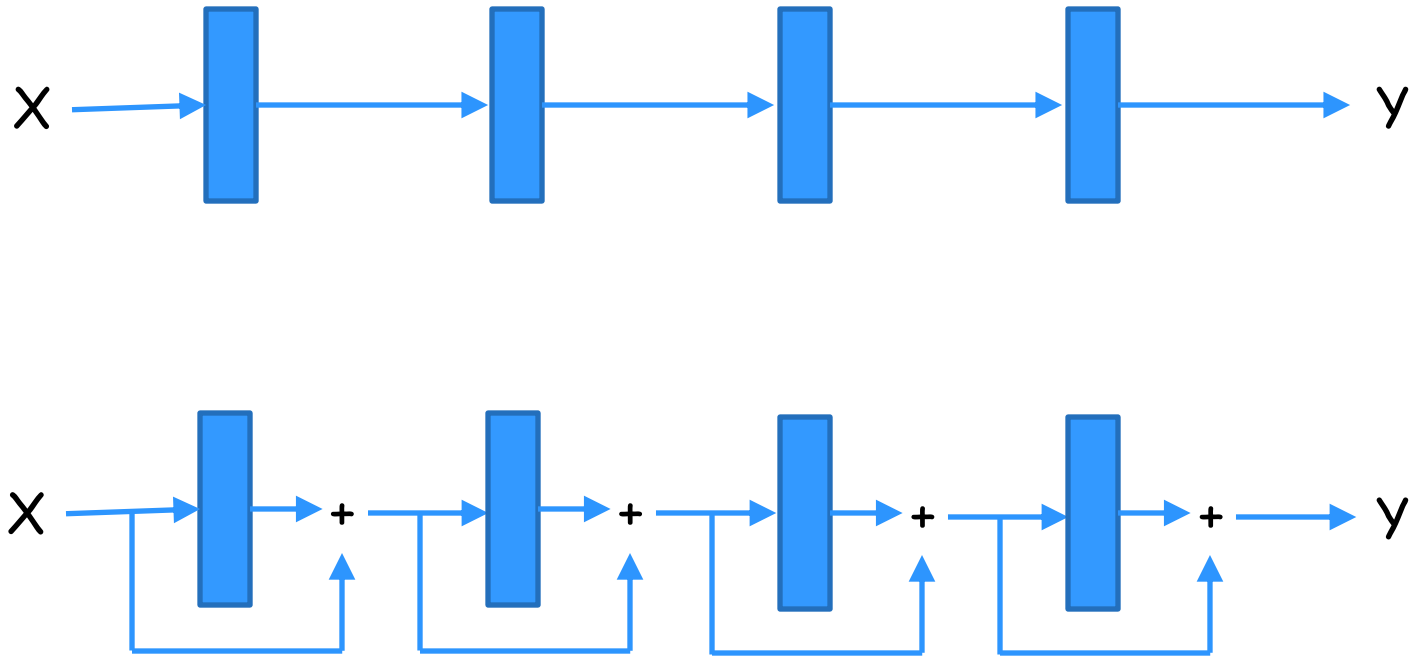
VGG16

ResNet

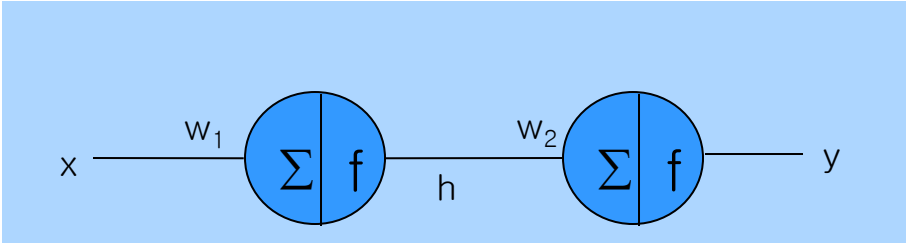
- **ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners**

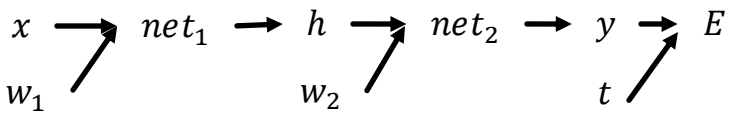


ResNet



ResNet

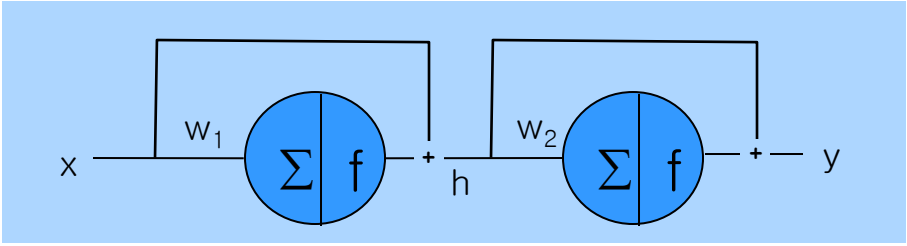


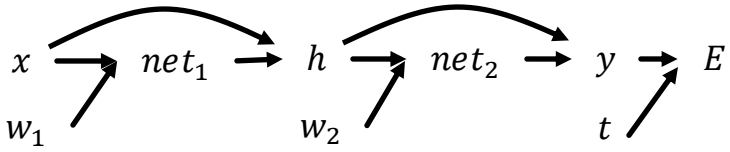


$$net_1 = x \cdot w_1$$
$$h = f(net_1)$$
$$net_2 = h \cdot w_2$$
$$y = f(net_2)$$
$$E = \frac{1}{2}(t - y)^2$$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial net_2}\frac{\partial net_2}{\partial h}\frac{\partial h}{\partial net_1}\frac{\partial net_1}{\partial w_1}$$

ResNet

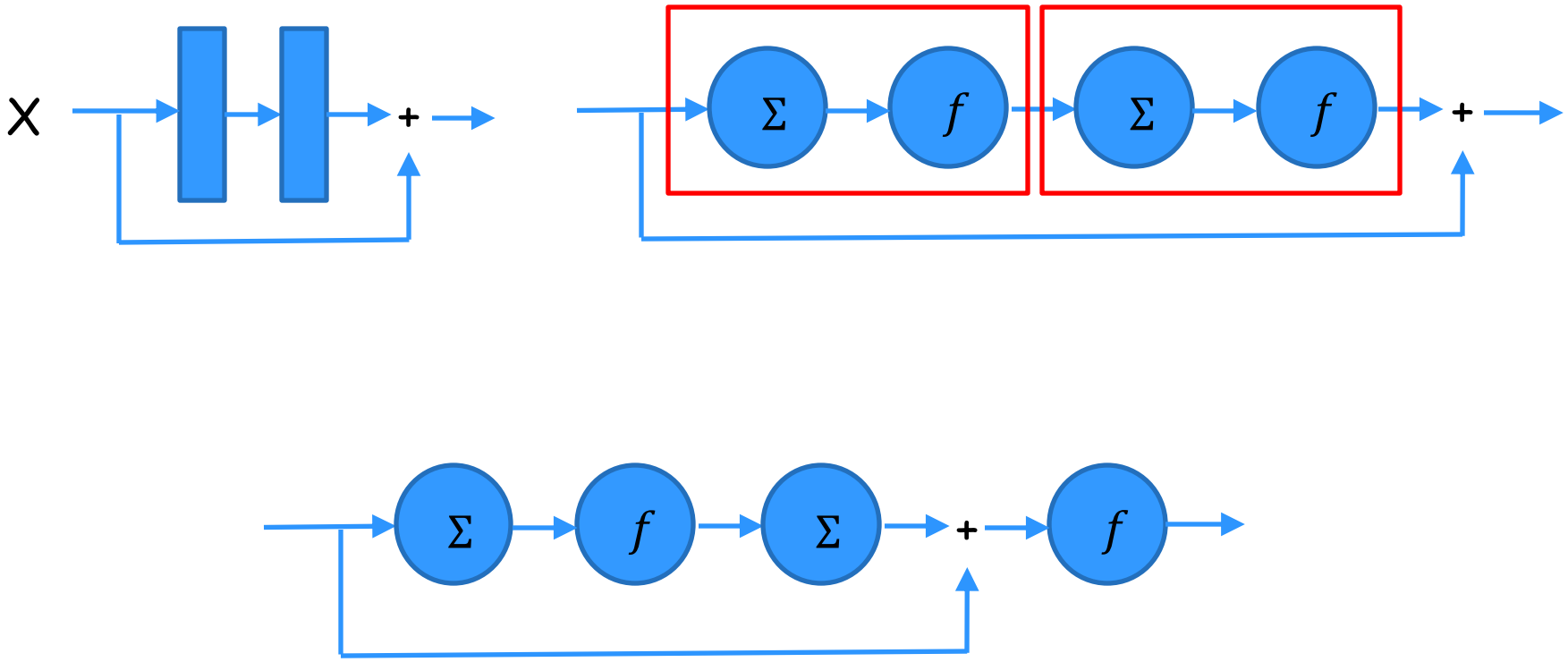




$$net_1 = x \cdot w_1$$
$$h = f(net_1) + x$$
$$net_2 = h \cdot w_2$$
$$y = f(net_2) + h$$
$$E = \frac{1}{2}(t - y)^2$$

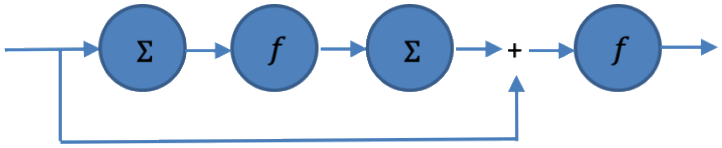
$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial y}\frac{\partial y}{\partial net_2}\frac{\partial net_2}{\partial h}\frac{\partial h}{\partial net_1}\frac{\partial net_1}{\partial w_1} + \frac{\partial E}{\partial y}\frac{\partial y}{\partial h}\frac{\partial h}{\partial net_1}\frac{\partial net_1}{\partial w_1}$$

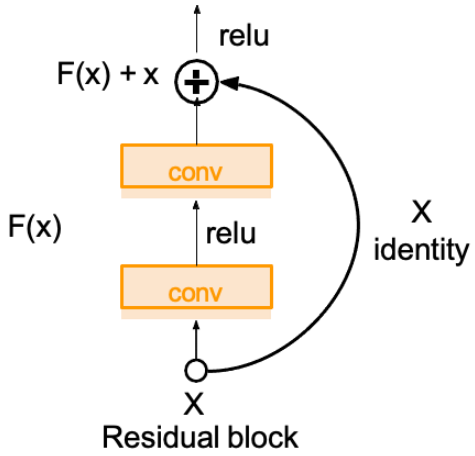
ResNet

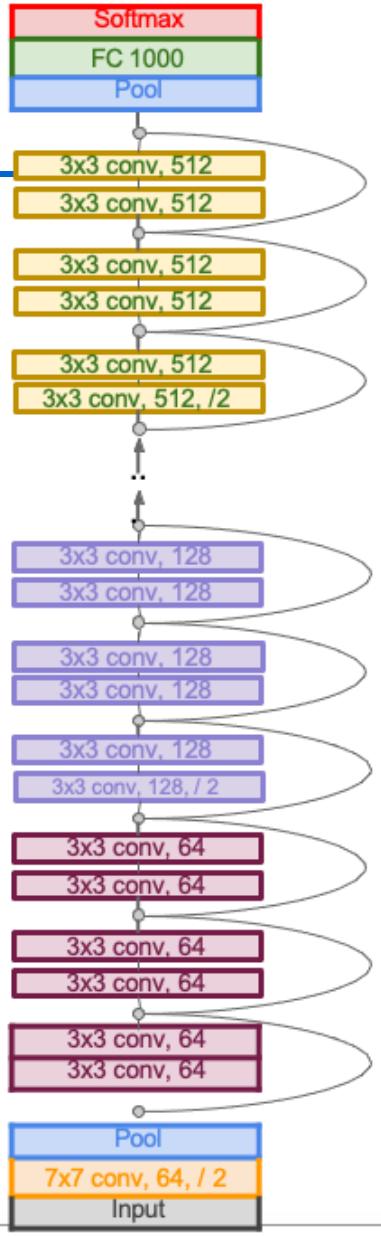


ResNet

- **Very deep networks using residual connections**
  - Every residual block has two 3x3 conv layers
  - Periodically, double # of filters and downsample spatially using stride 2
  - Global average pooling layer after last conv. layer

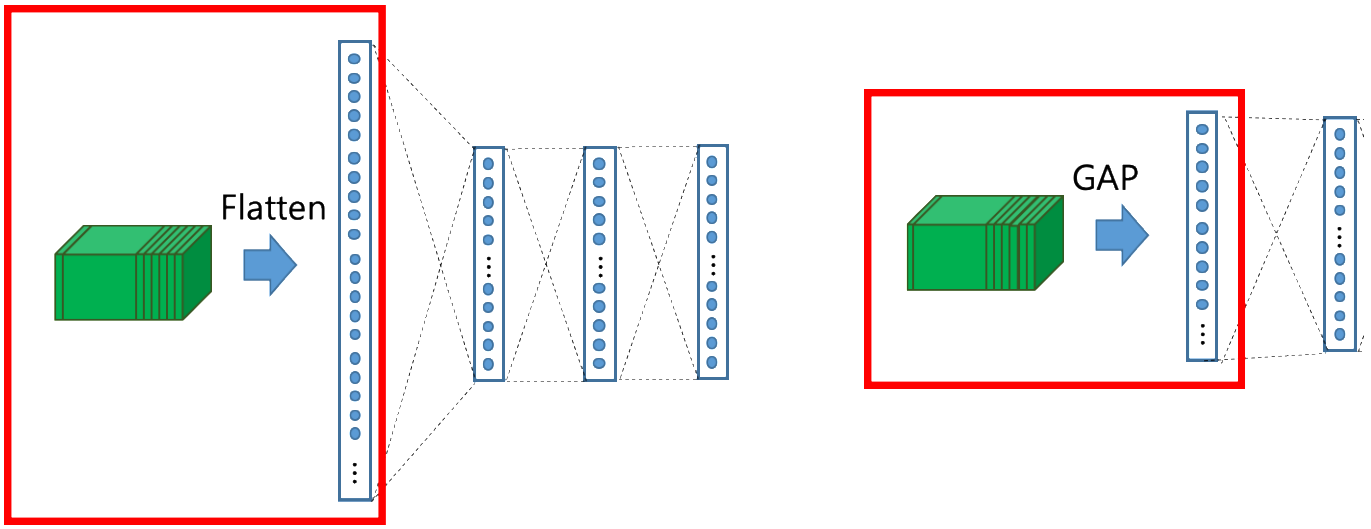






ResNet

- **GAP vs Flatten**



Comparing Complexity