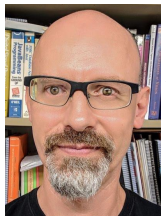
telecom BCN

UPC

Day 4 Lecture 3

Architectures from Imagenet

https://telecombcn-dl.github.io/idl-2021/



Josep Ramon Morros
ramon.morros@upc.edu
Associate Professor,
Universitat Politècnica de Catalunya

Acknowledgements

Xavier Giro-i-Nieto

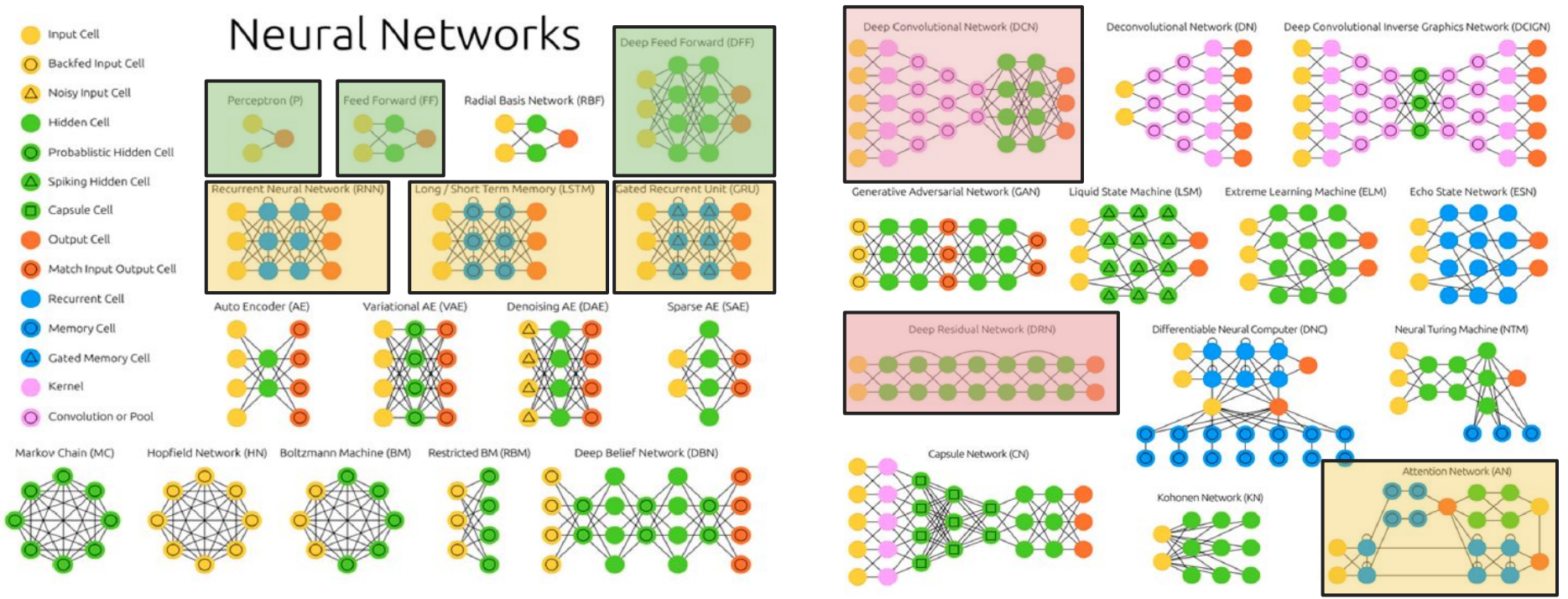
xavier.giro@upc.edu

Associate Professor

Universitat Politecnica de Catalunya

Technical University of Catalonia

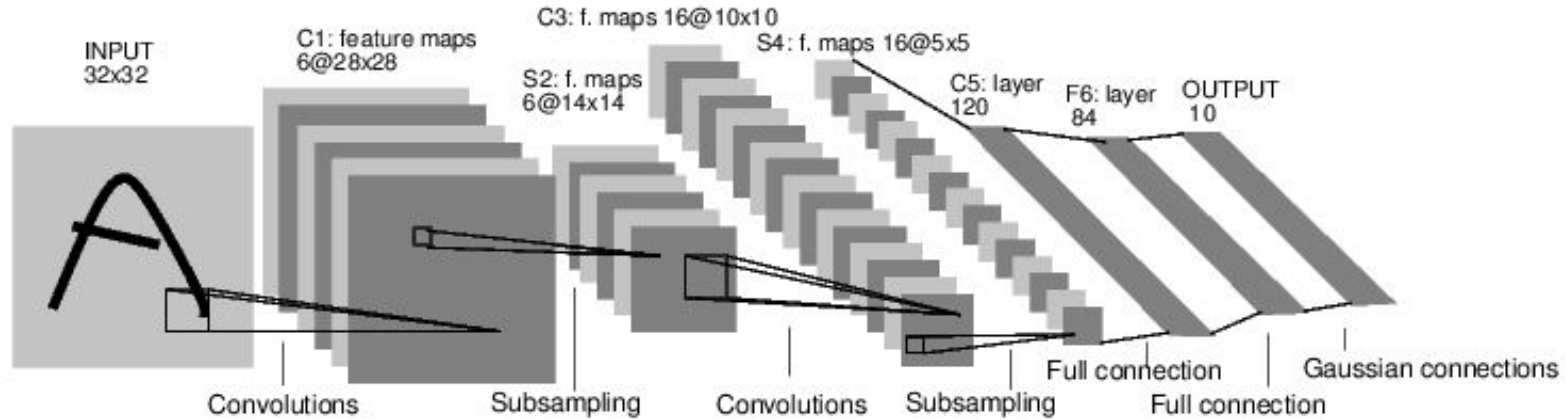
Neural Network architectures



Leijnen S, Veen Fv. The Neural Network Zoo. *Proceedings*. 2020; 47(1):9. https://doi.org/10.3390/proceedings2020047009

Convolutional Neural Networks for Vision

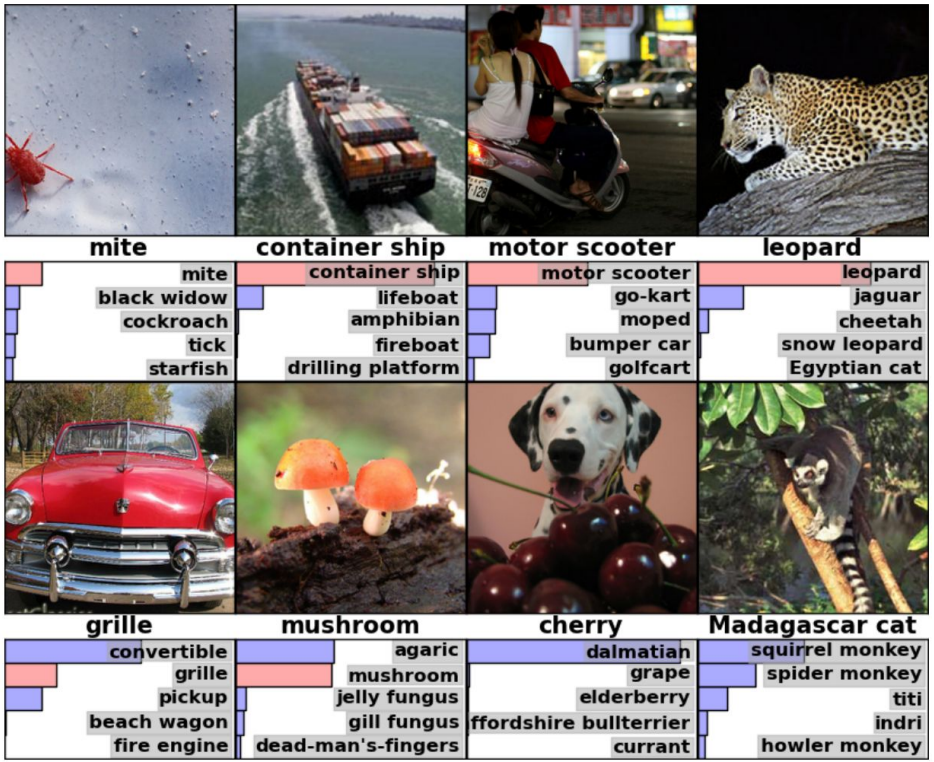
LeNet-5: Several convolutional layers, combined with pooling layers, and followed by one fully connected layer



#LeNet-5 LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

ImageNet Challenge

- **ImageNet** is a large visual database designed for use in visual object recognition research.
- It is organized hierarchically.
- Each node is depicted by hundreds/thousands of images.
- The project has been instrumental in advancing computer vision and deep learning research.
- More than 14 million images have been hand-annotated to indicate what objects are pictured and in at least one million of the images, bounding boxes are also provided.



ImageNet challenge (ILSVRC)

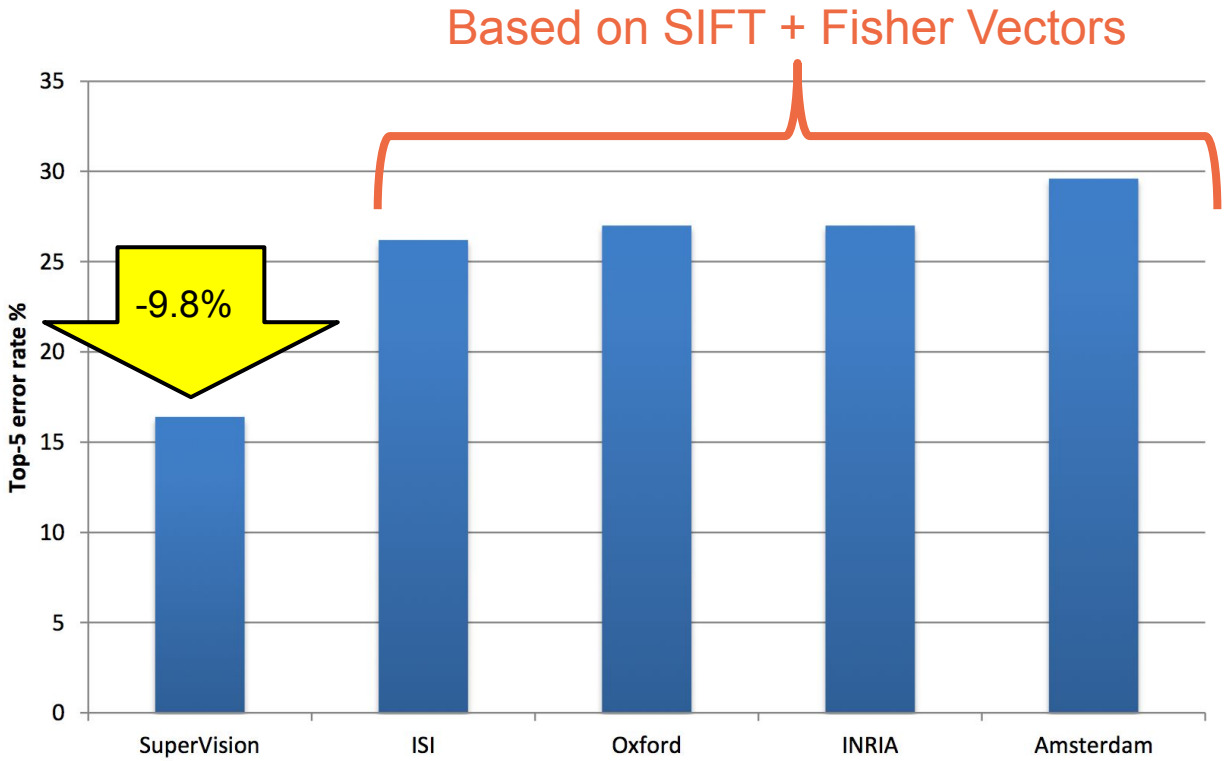
- 1,000 object classes
- Images:
  - 1.2 M train
  - 100k test.
- Metric: Top 5 error rate

Algorithm predicts at most 5 classes in descending order of confidence. If the correct class is among the first 5 predictions, error is 0, otherwise is 1

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "Imagenet large scale visual recognition challenge." International Journal of Computer Vision 115, no. 3 (2015): 211-252. [web]

ImageNet Challenge: 2012

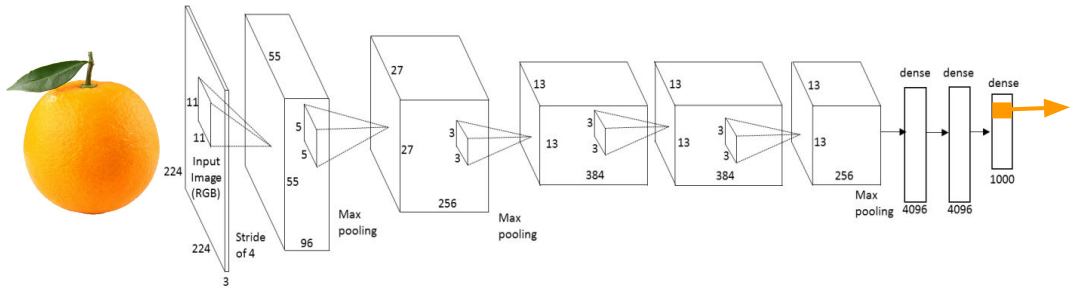
Slide credit:
Rob Fergus (NYU)



Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "Imagenet large scale visual recognition challenge." International Journal of Computer Vision 115, no. 3 (2015): 211-252. [web]

AlexNet (SuperVision)

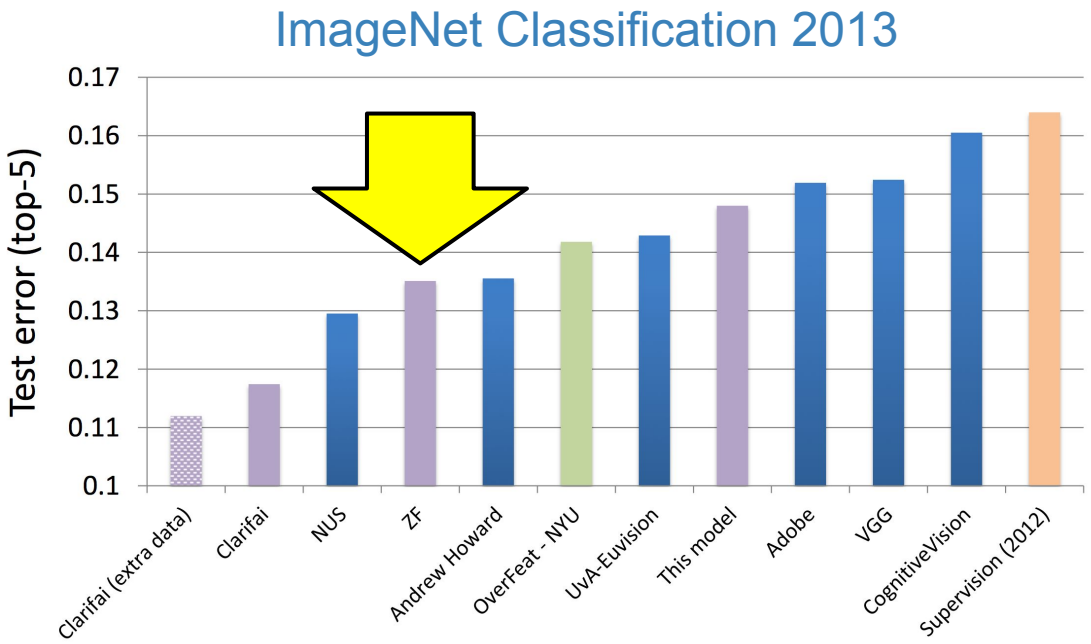
- Main ideas: ReLU nonlinearity, training on multiple GPUs, local response normalization, overlapping pooling, data augmentation, dropout.
- There are 8 trainable layers: 5 convolutional and 3 fully connected.
- 60 million free parameters



#AlexNet Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." NIPS **2012**

ImageNet Challenge: 2013

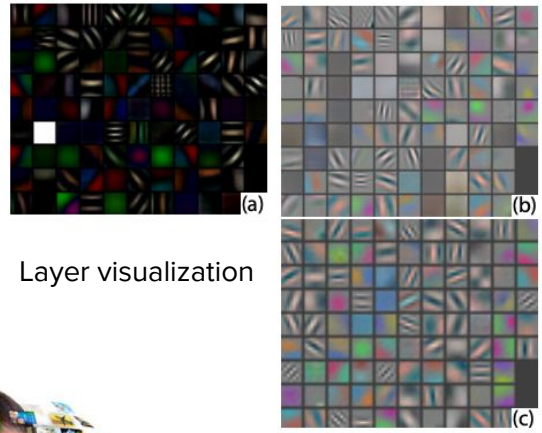
IMAGENET



Slide credit:
Rob Fergus (NYU)

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "Imagenet large scale visual recognition challenge." International Journal of Computer Vision 115, no. 3 (2015): 211-252. [web]

Zeiler-Fergus (ZF)



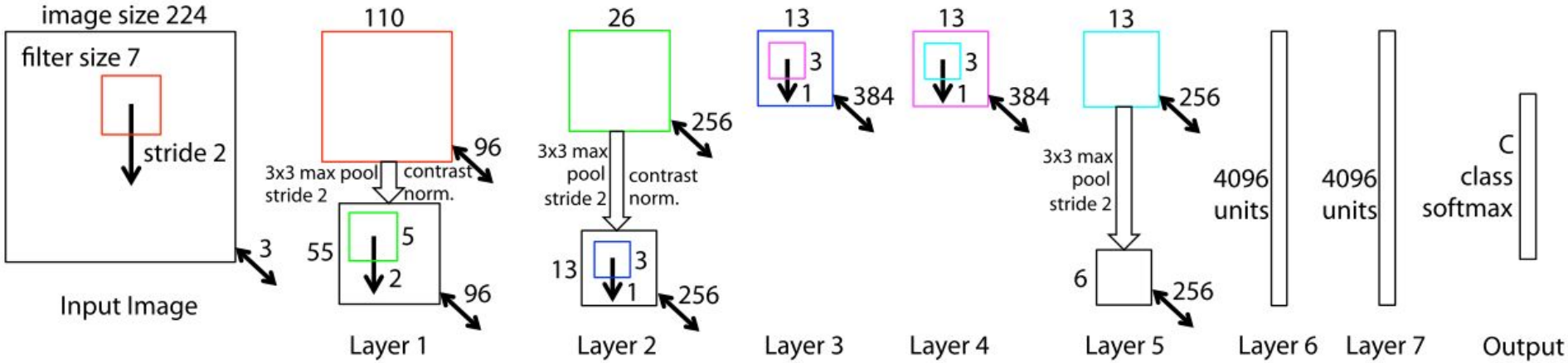
Layer visualization

The development of better convnets is reduced to trial-and-error.

Visualization can help in proposing better architectures.

Zeiler, M. D., & Fergus, R. . Visualizing and understanding convolutional networks. ECCV 2014

Zeiler-Fergus (ZF)

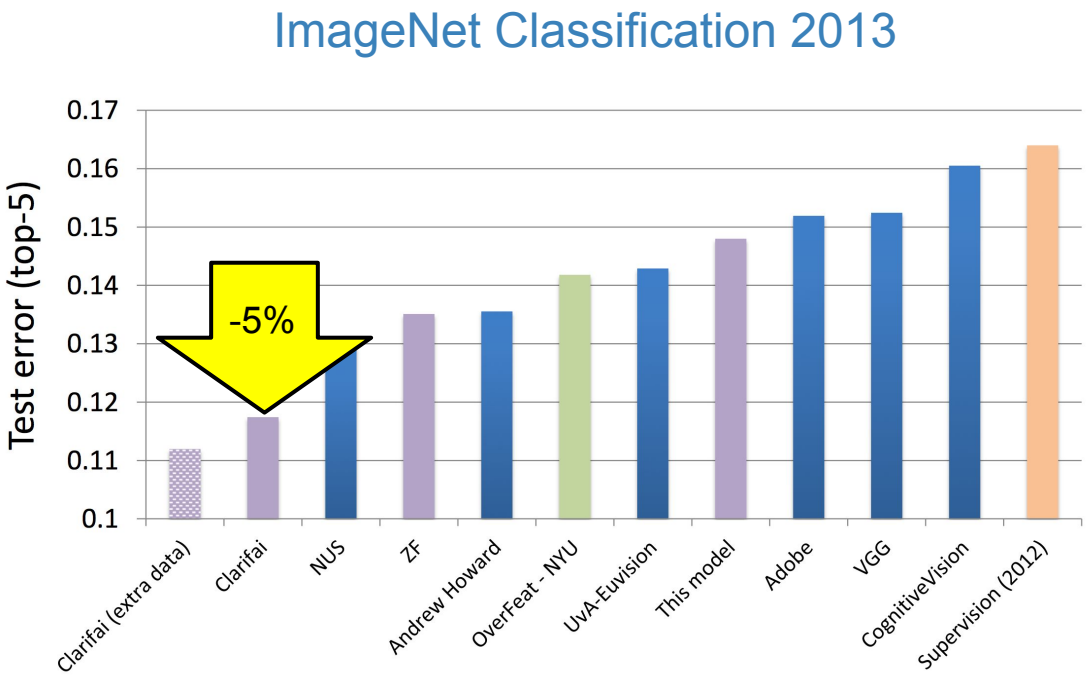


Zeiler, M. D., & Fergus, R. . Visualizing and understanding convolutional networks. ECCV 2014

ImageNet Challenge: 2013

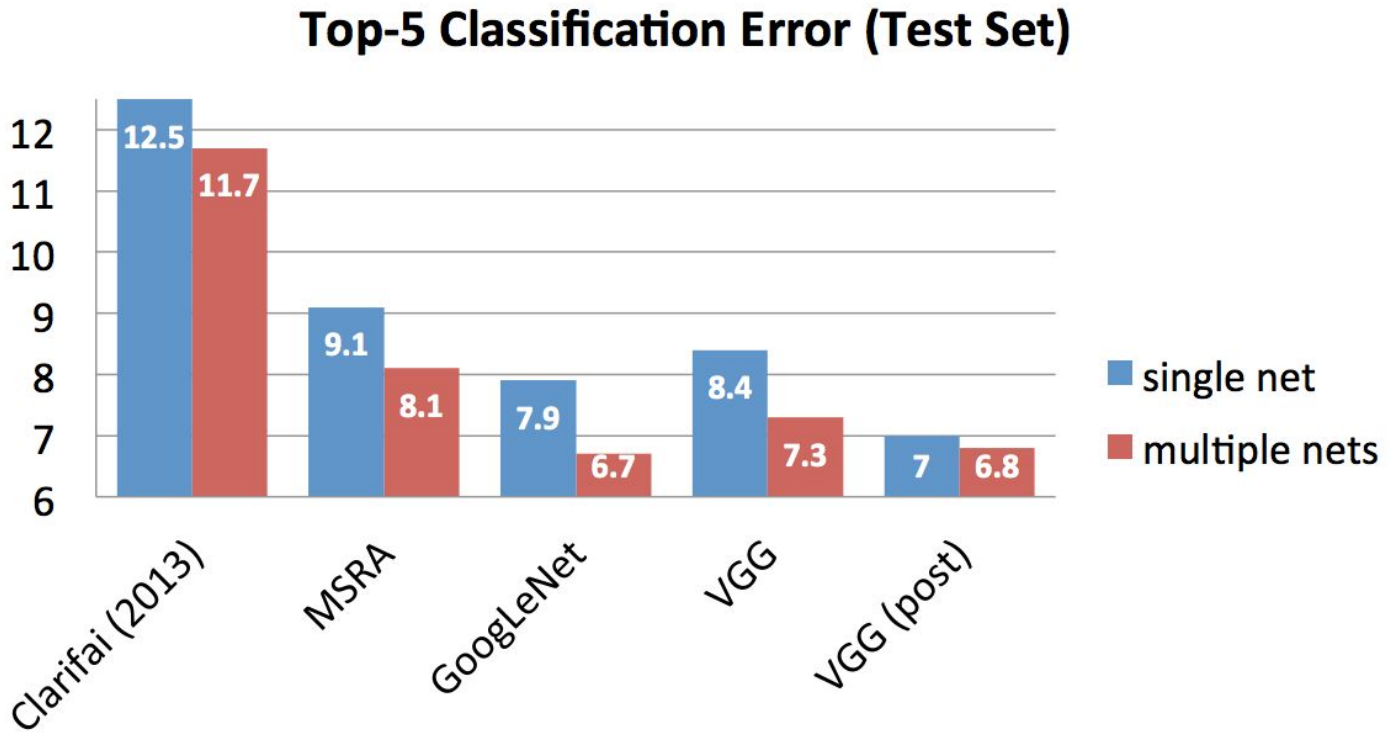
IMAGENET

clarifai



Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "Imagenet large scale visual recognition challenge." IJCV 2015. [web]

VGG



#VGG Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *ICLR 2015*. [video] [slides] [project]

VGG:

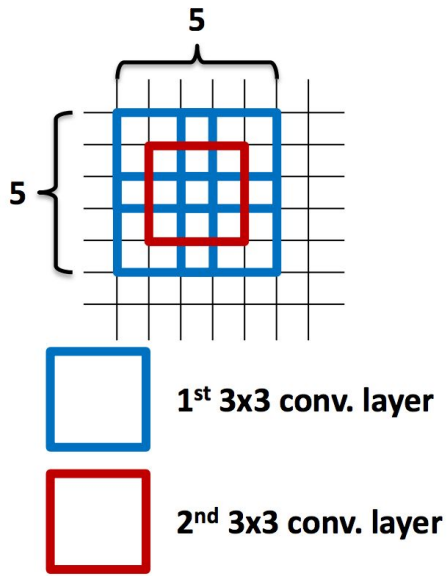
- Add more layers
- Stacked convolutions with smaller filters: 3x3 work better than a large 7x7 convolution
- Use only:
  - 3x3 CONV **stride 1, pad 1** and
  - 2x2 MAX POOL stride 2
- Shows that depth is a critical component for good performance (16 – 19 layers)
- 2 versions
  - VGG-16 (16 parameter layers)
  - VGG-19 (19 parameter layers)

#VGG Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *ICLR 2015*. [video] [slides] [project]

VGG: Stacked 3x3 convolutions

Why 3x3 layers?

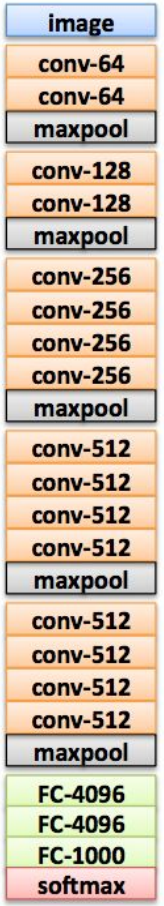
- Stacked conv. layers have a large receptive field
  - two 3x3 layers – 5x5 receptive field
  - three 3x3 layers – 7x7 receptive field
- More non-linearity
- Less parameters to learn
  - ~140M per net



#VGG Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *ICLR 2015*. [video] [slides] [project]

VGG: Other details

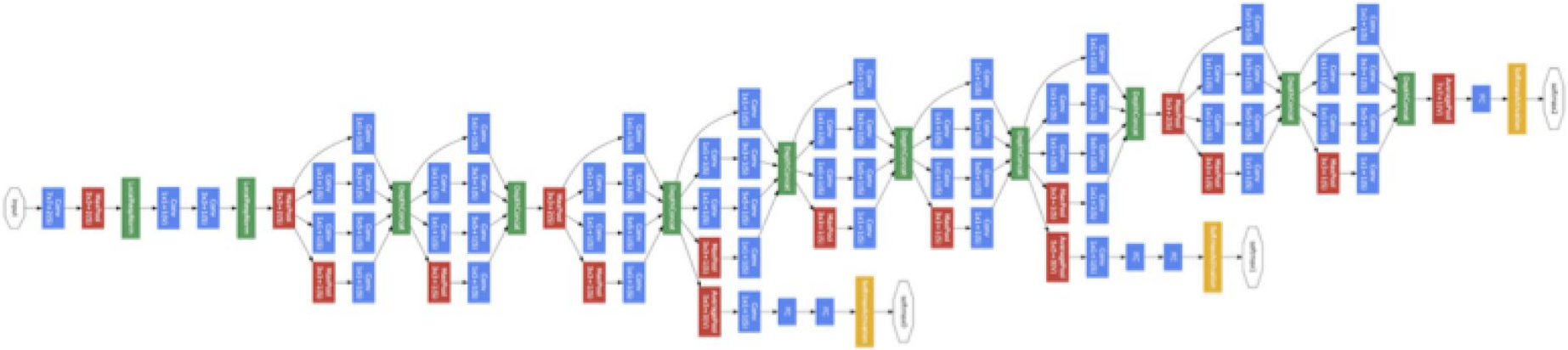
- No poolings between some convolutional layers.
- Convolution strides of 1 (no skipping).



Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *ICLR 2015*. [video] [slides] [project]

GoogleNet (Inception)

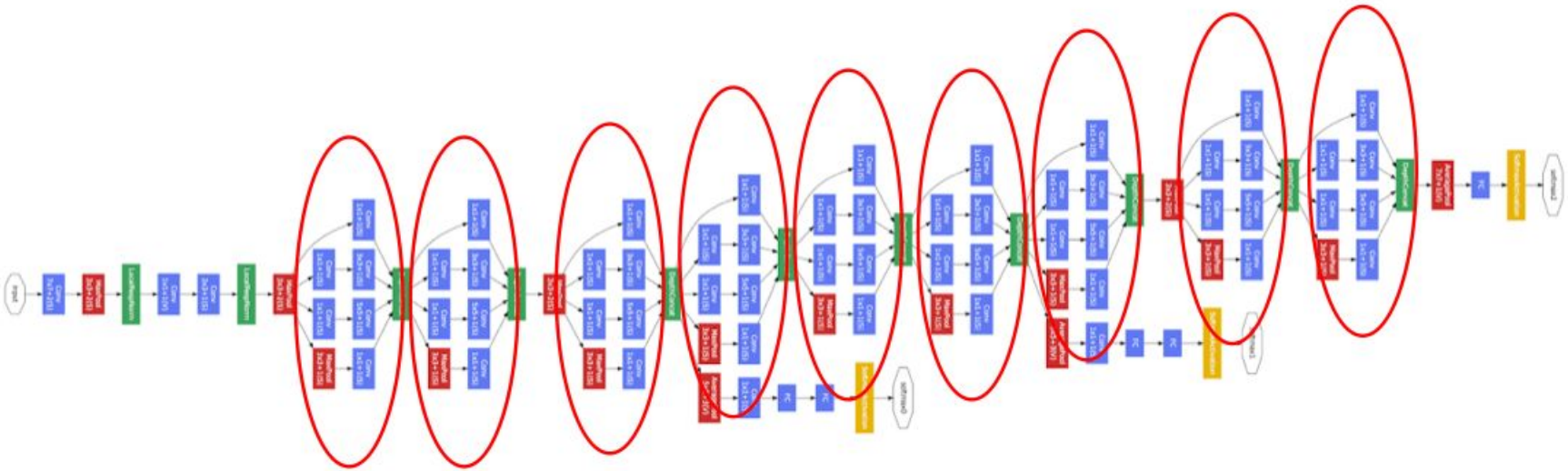
22 layers !



Convolution
Pooling
Softmax
Other

#Inception Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." CVPR 2015. [video] [slides] [poster]

GoogleNet (Inception)

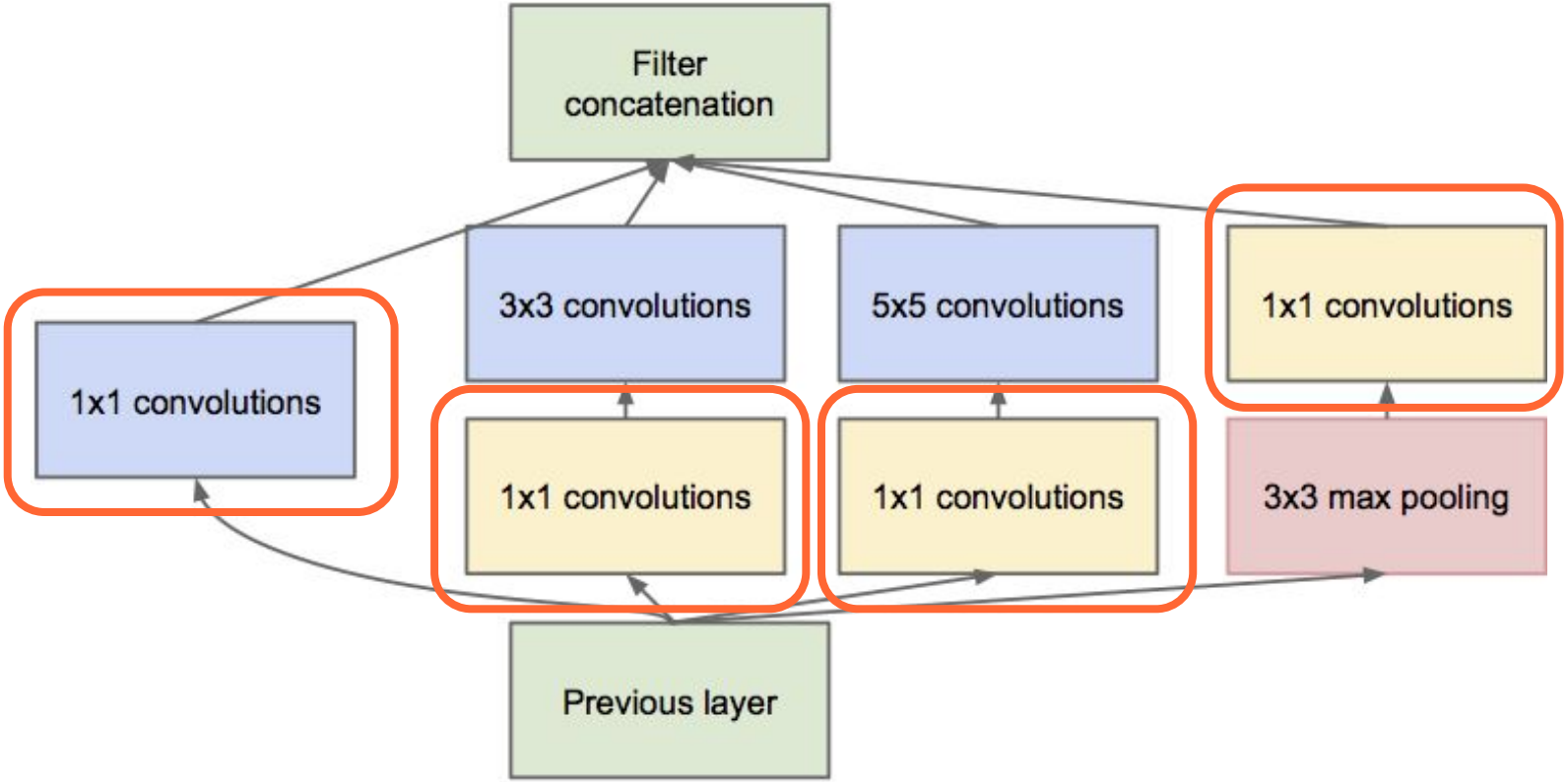


9 **Inception** modules

Network in a network in a network...

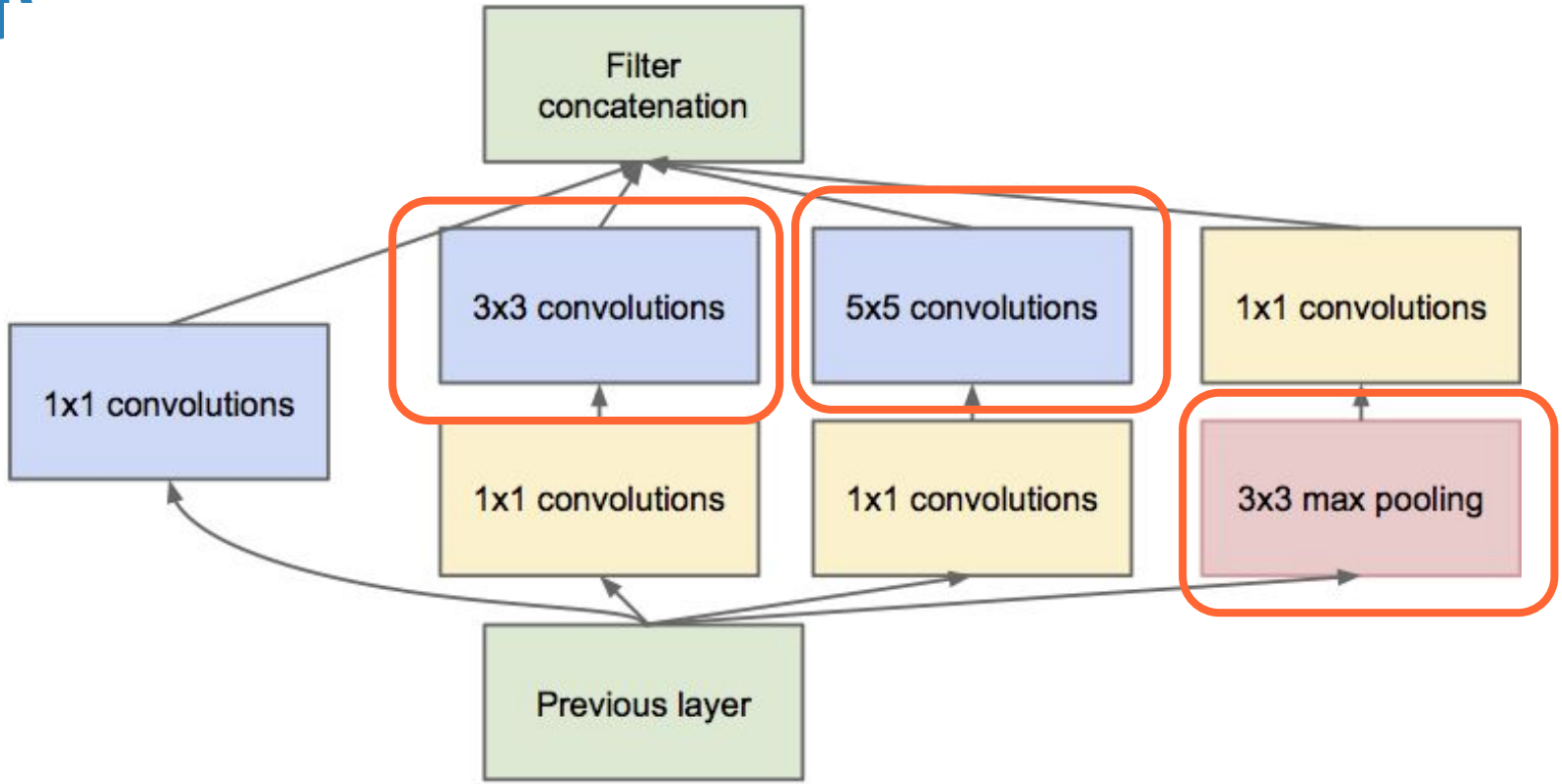
Convolution
Pooling
Softmax
Other

Pointwise convolutions (bottlenecks)



#NiN Lin, Min, Qiang Chen, and Shuicheng Yan. "Network in network." *ICLR 2014.*

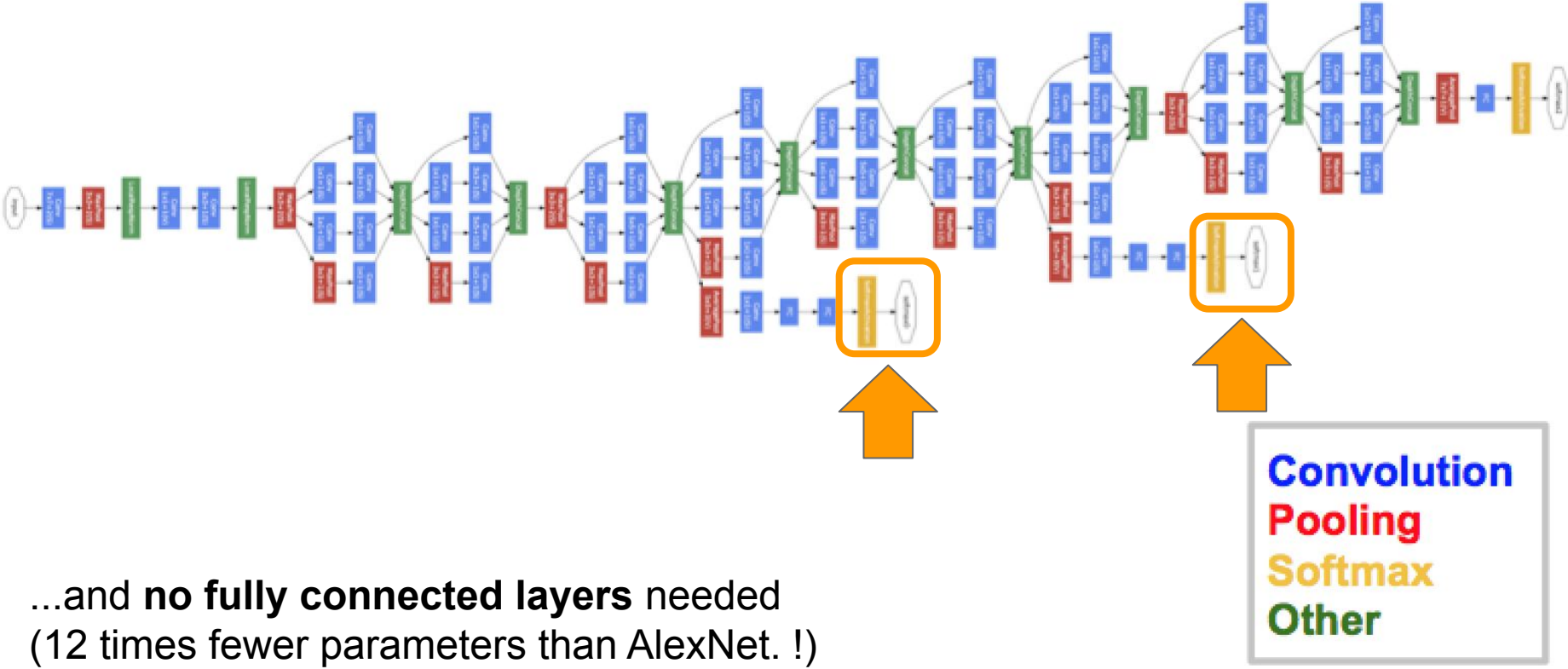
Multiple convolutional filter sizes (receptive field‘



#NiN Lin, Min, Qiang Chen, and Shuicheng Yan. "Network in network." *ICLR 2014.*

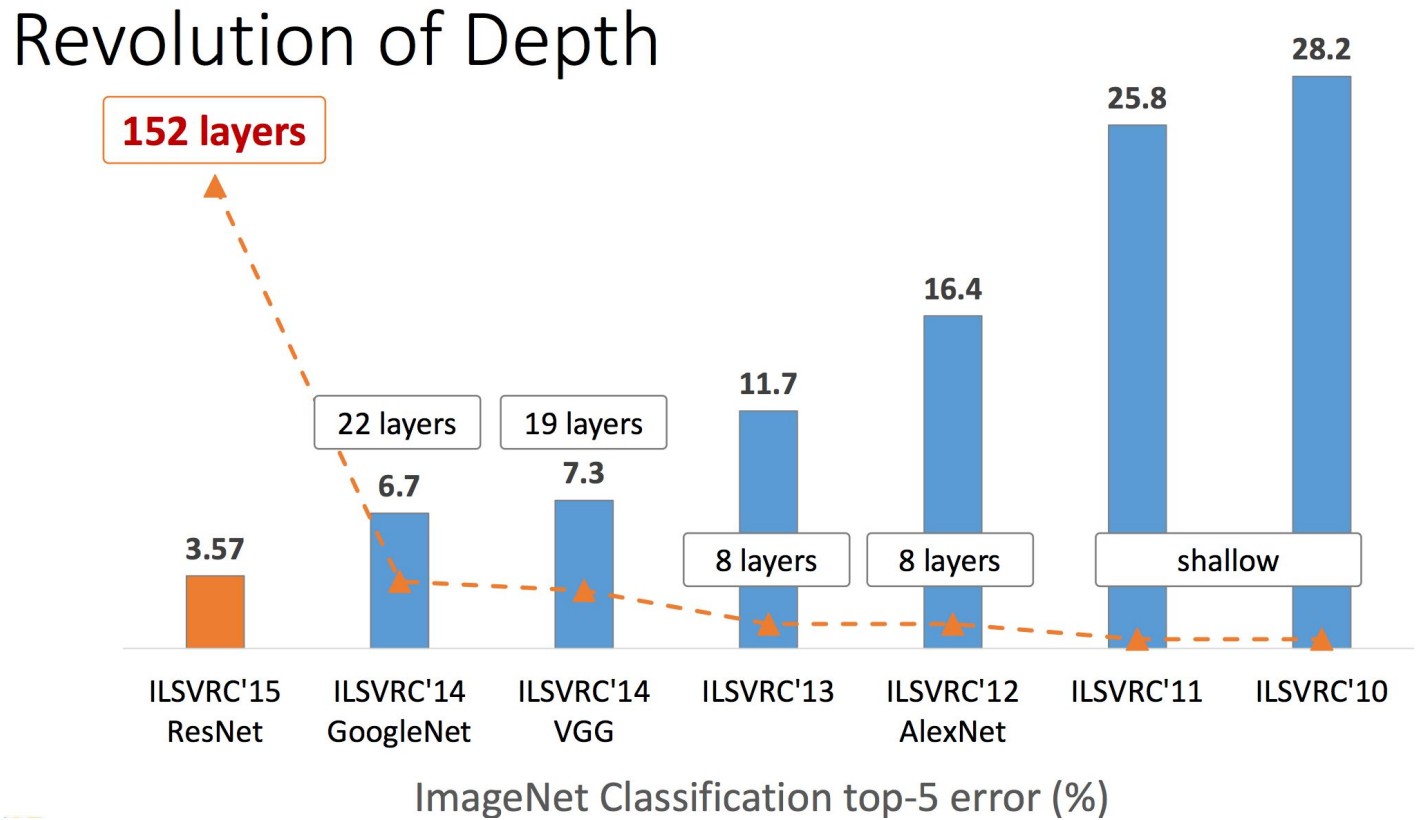
GoogleNet (Inception)

Two Softmax classifiers at intermediate layers combat the vanishing gradient while providing regularization at training time.



...and **no fully connected layers** needed
(12 times fewer parameters than AlexNet. !)

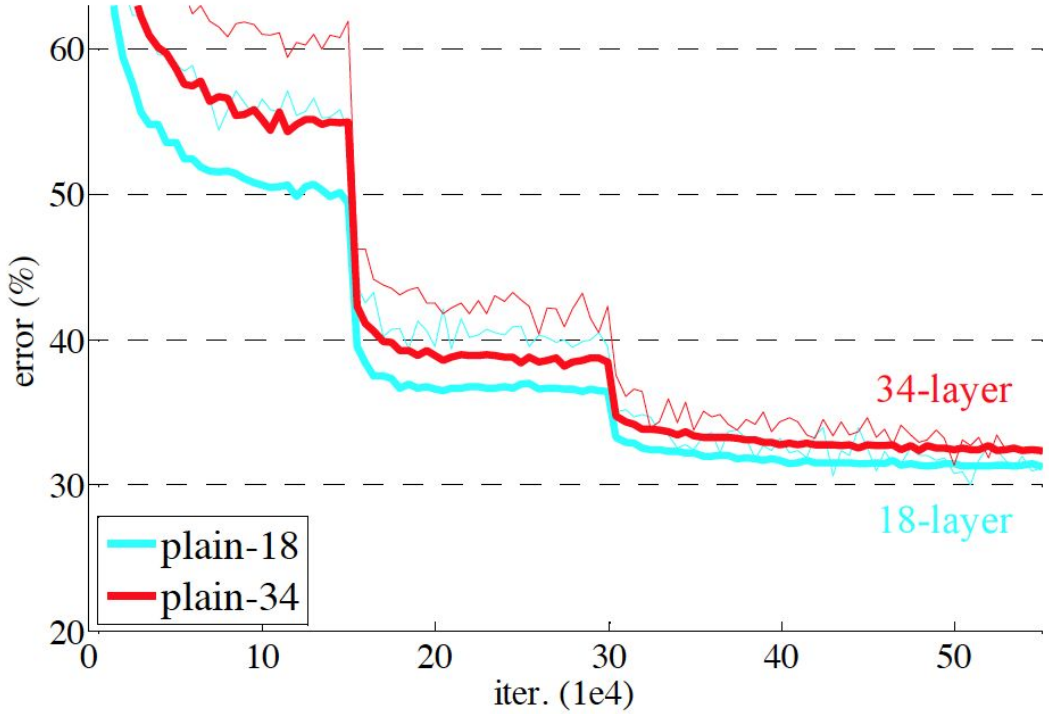
ResNet



He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." CVPR 2016. [slides]

ResNet

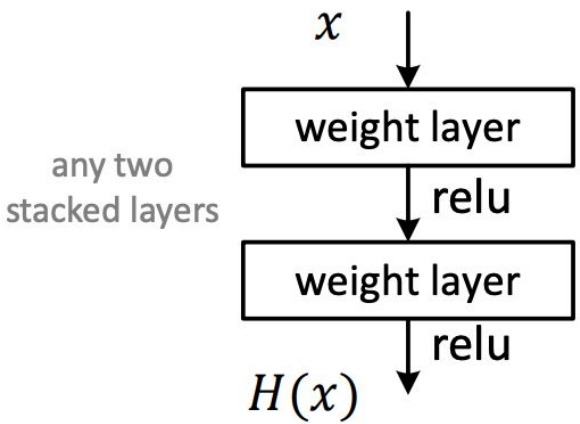
Challenge: Deeper networks (34 is deeper than 18) are more difficult to train.



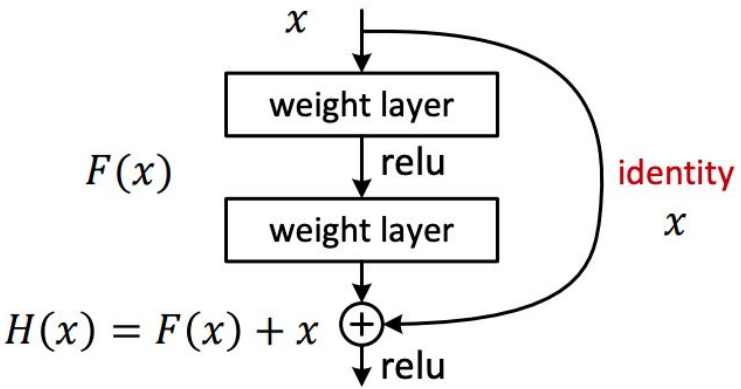
He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." CVPR 2016. [slides]

ResNet

Plain Net



Residual Net



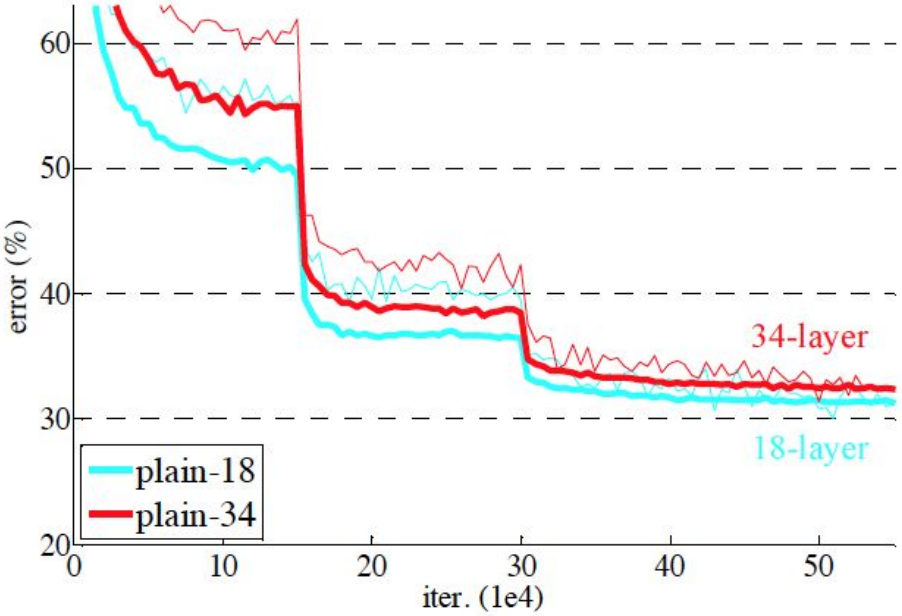
Residual learning: reformulate the layers as learning residual functions with respect to the identity F(x), instead of learning unreferenced functions H(x)

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." CVPR 2016. [slides]

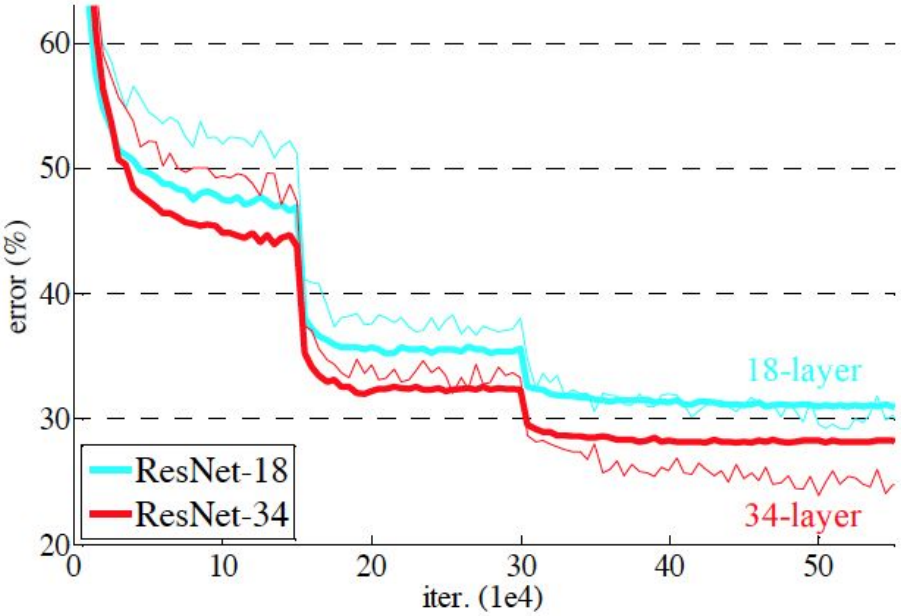
ResNet

Solution: Replace plain layer for residual layers.

Non-Residual connections



Residual connections

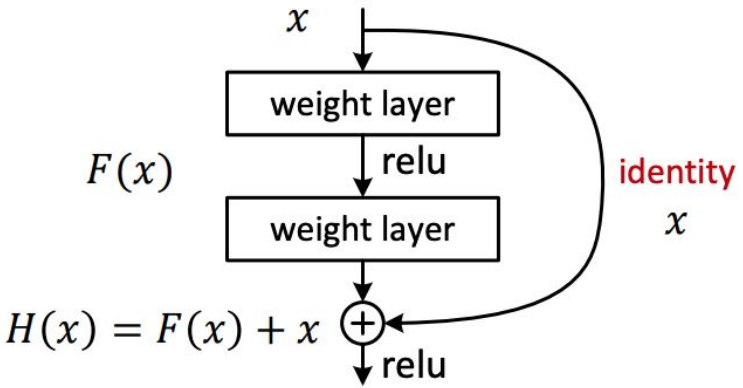


He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." CVPR 2016. [slides]

ResNet

Why ResNets train better ?:

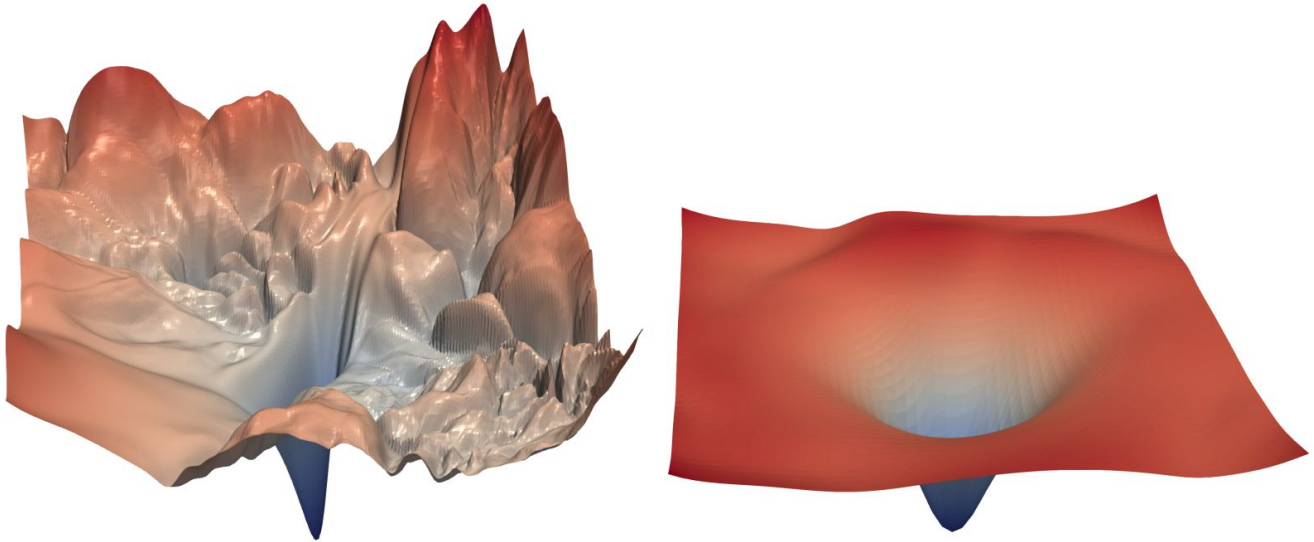
“We hypothesize that it is easier to optimize the residual mapping than to optimize the original, unreferenced mapping. To the extreme, if an identity mapping were optimal (given x return x), it would be easier to push F(x) to zero than to fit an identity mapping by a stack of nonlinear layers”



He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." CVPR 2016. [slides]

ResNet

The loss surfaces of ResNet-56 with/without skip connections



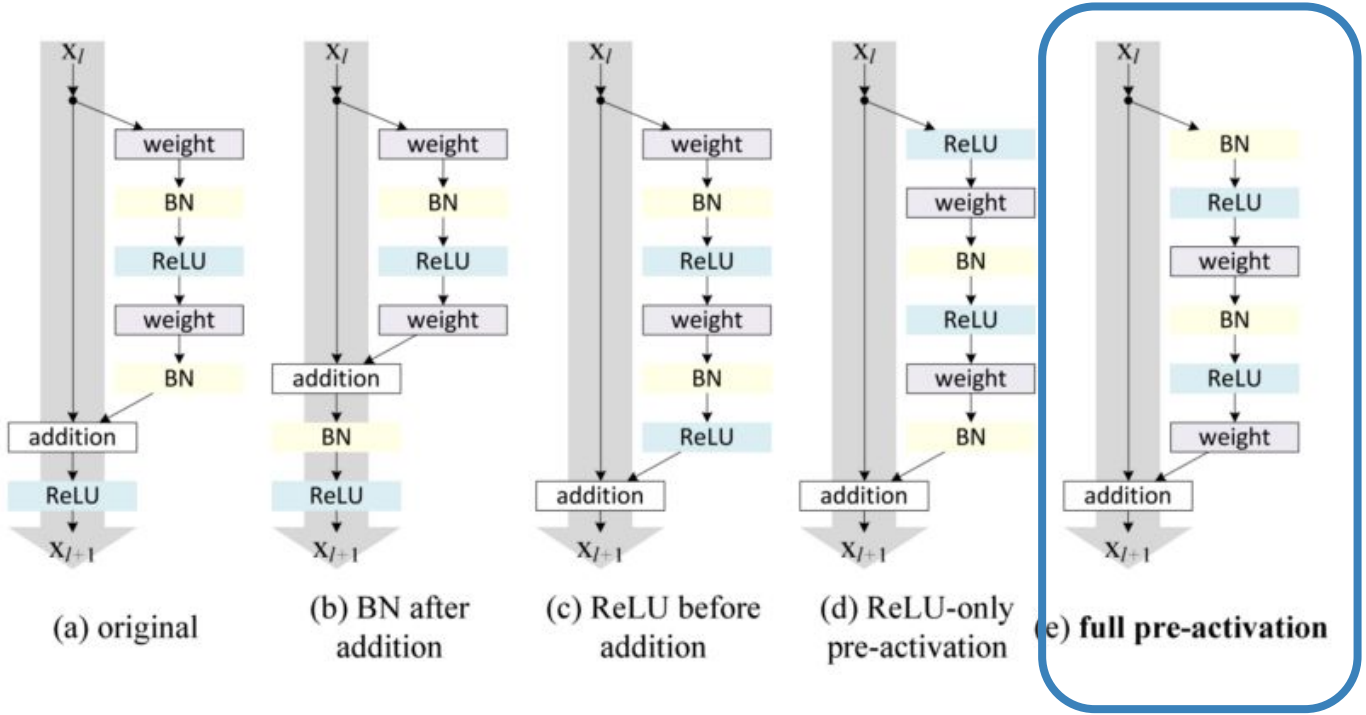
(a) without skip connections

(b) with skip connections

Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer and Tom Goldstein. Visualizing the Loss Landscape of Neural Nets. NeurIPS, 2018. [code] [poster]

ResNet

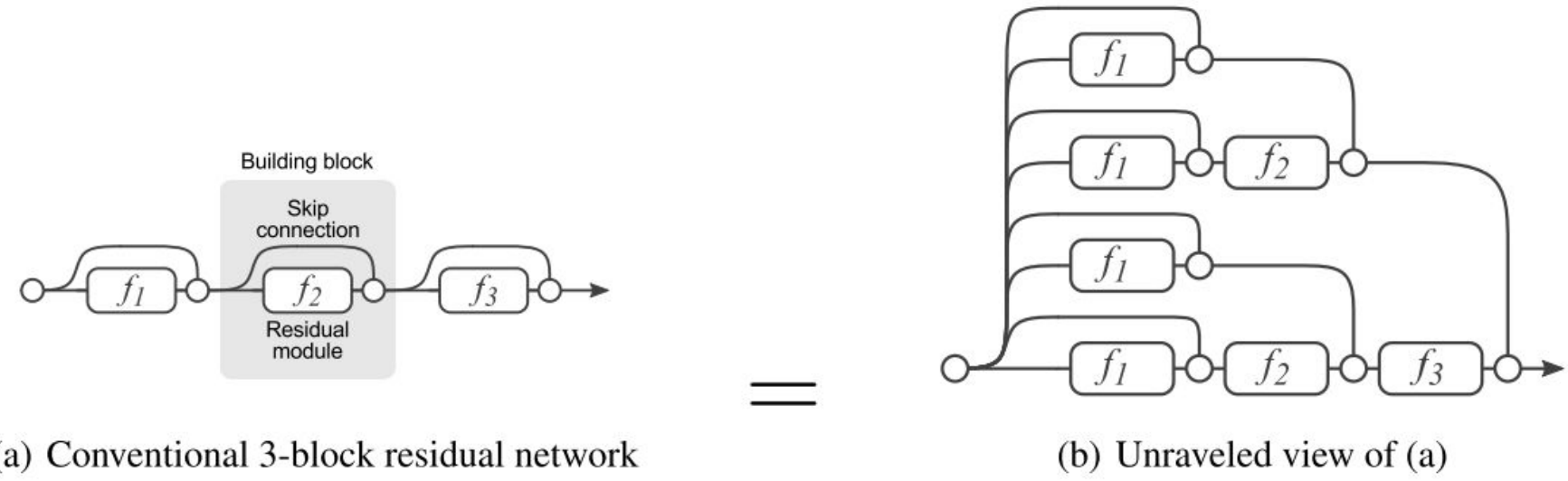
It has been observed that pre-activations with batch normalizations give the best results in general



He, K., Zhang, X., Ren, S., & Sun, J. "Identity mappings in deep residual networks". ECCV 2016.

ResNet

After we unroll the network architecture, it is quite clear that a ResNet architecture with i residual blocks has 2 ** i different paths (because each residual block provides two independent paths).



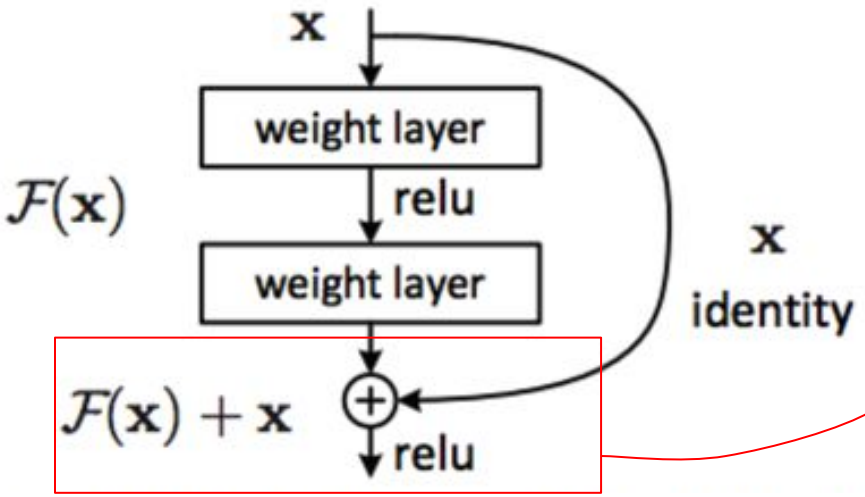
(a) Conventional 3-block residual network

(b) Unraveled view of (a)

Veit, Andreas, Michael J. Wilber, and Serge Belongie. "Residual networks behave like ensembles of relatively shallow networks." NIPS 2016.

ResNet

PyTorch

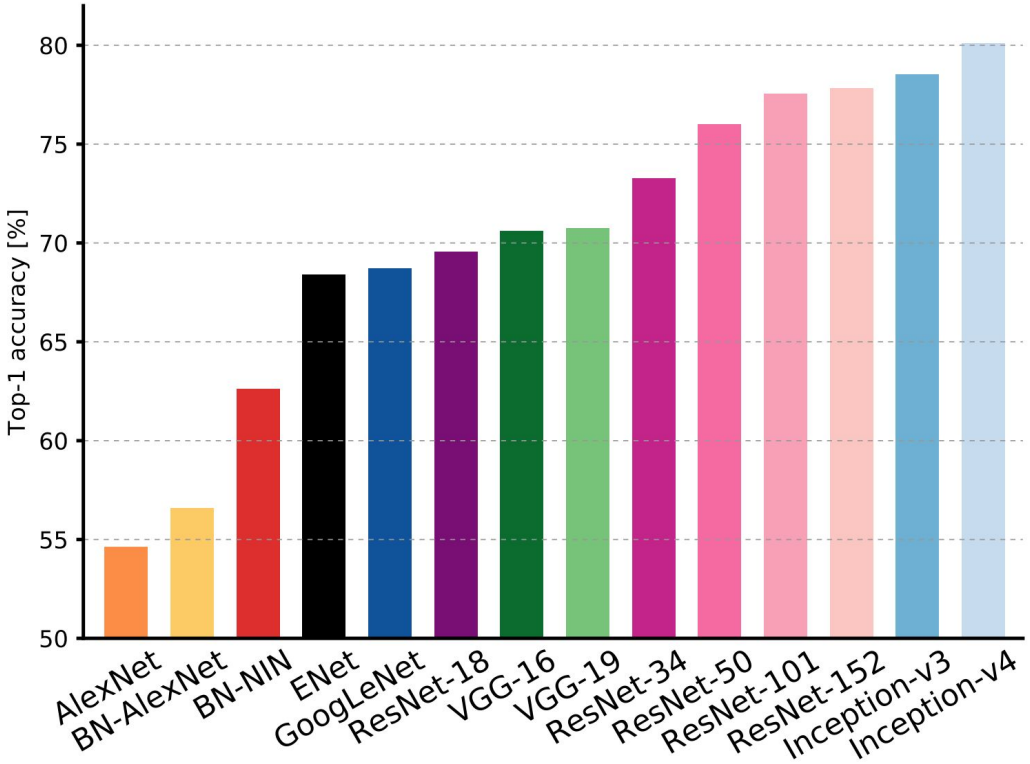



```python
class ResLayer(nn.Module):

  def __init__(self, num_inputs):
    super().__init__()
    self.num_inputs = num_inputs
    num_outputs = num_inputs
    self.num_outputs = num_outputs
    self.conv1 = nn.Sequential(
        nn.Conv2d(num_inputs, num_outputs, 3, padding=1),
        nn.BatchNorm2d(num_outputs),
        nn.ReLU(inplace=True)
    )
    self.conv2 = nn.Sequential(
        nn.Conv2d(num_outputs, num_outputs, 3, padding=1),
        nn.BatchNorm2d(num_outputs),
        nn.ReLU(inplace=True)
    )
    self.out_relu = nn.ReLU(inplace=True)

  def forward(self, x):
    # non-linear processing trunk
    conv1_h = self.conv1(x)
    conv2_h = self.conv2(conv1_h)
    # output is result of res connection + non-linear processing
    y = self.out_relu(x + conv2_h)
    return y

x = torch.randn(1, 64, 100, 100)
print('Input tensor x size: ', x.size())
reslayer = ResLayer(64)
y = reslayer(x)
print('Output tensor y size: ', y.size())
```

```
Input tensor x size:  torch.Size([1, 64, 100, 100])
Output tensor y size:  torch.Size([1, 64, 100, 100])
```

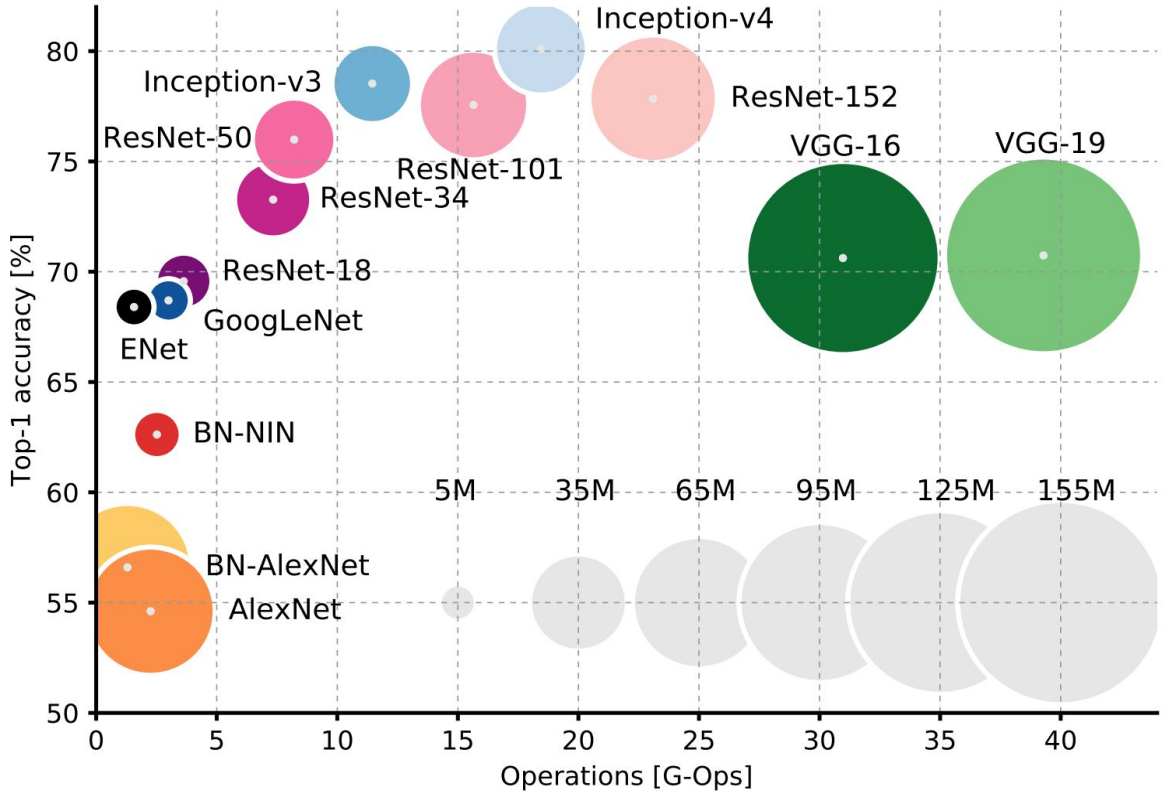

Slide credit: Santiago Pascual (UPC School 2019)

ResNet



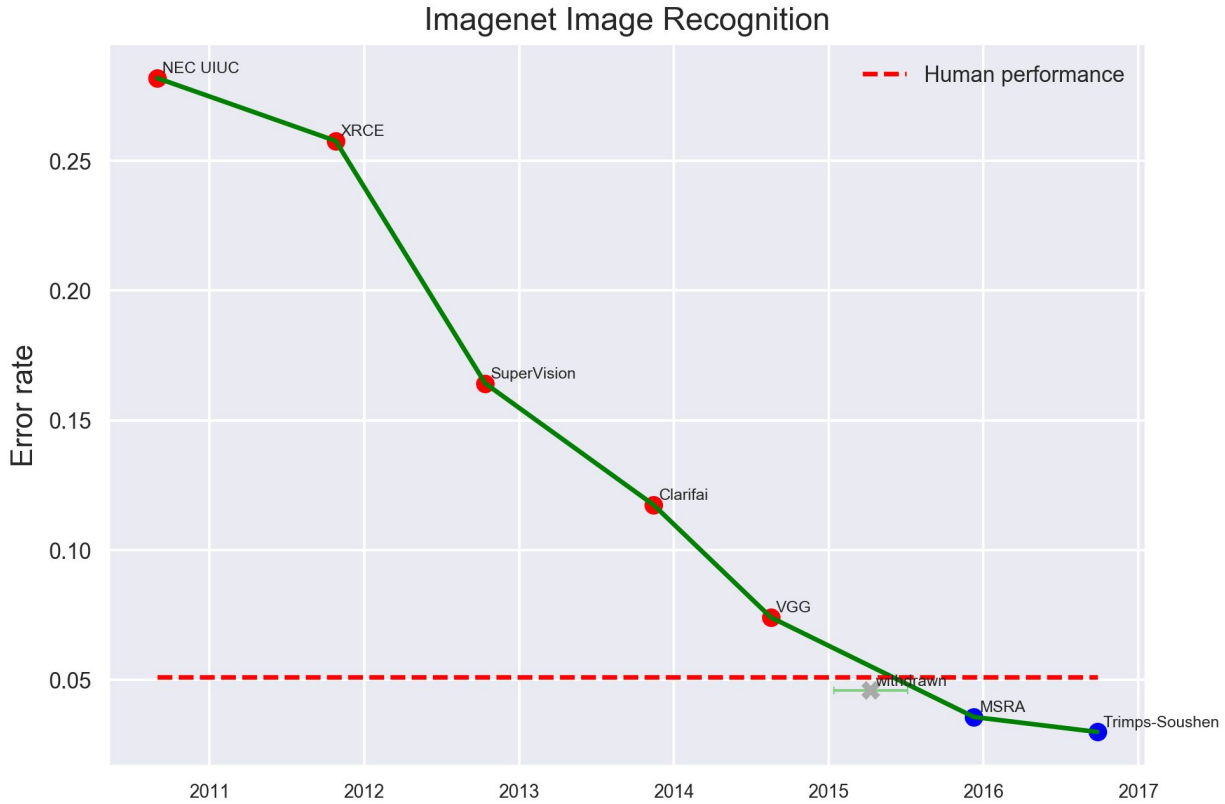
Canziani, Alfredo, Adam Paszke, and Eugenio Culurciello. "An analysis of deep neural network models for practical applications." arXiv preprint arXiv:1605.07678 (2016).

ResNet



Canziani, Alfredo, Adam Paszke, and Eugenio Culurciello. "An analysis of deep neural network models for practical applications." arXiv preprint arXiv:1605.07678 (2016).

ResNet



Electronic Frontier Foundation: [“Measuring the Progress of AI Research”](#) (2017)

Xception

Depth-wise (in channels) separable convolutions.

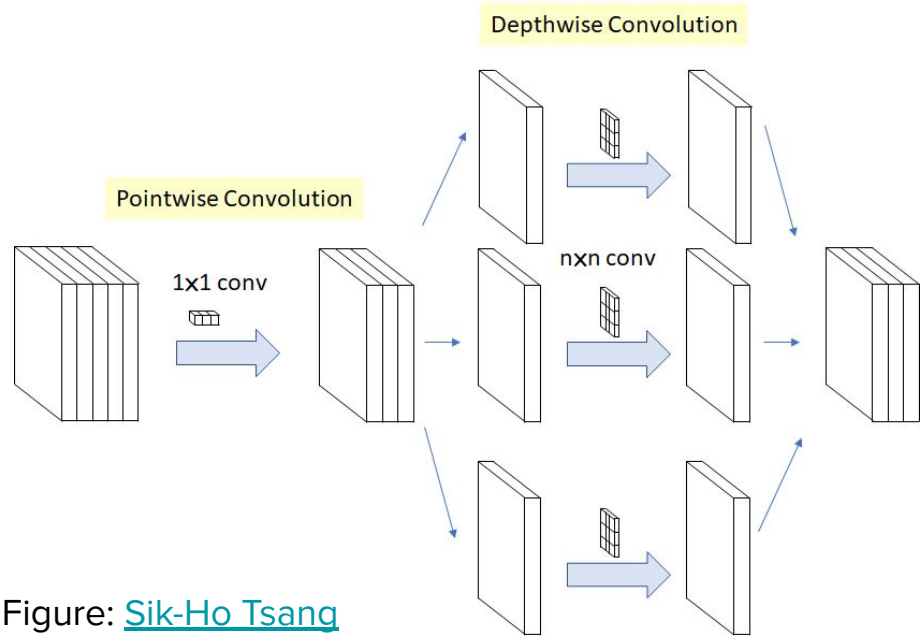
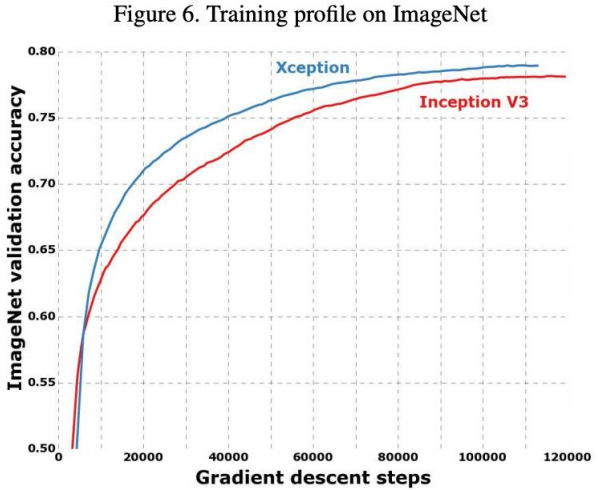


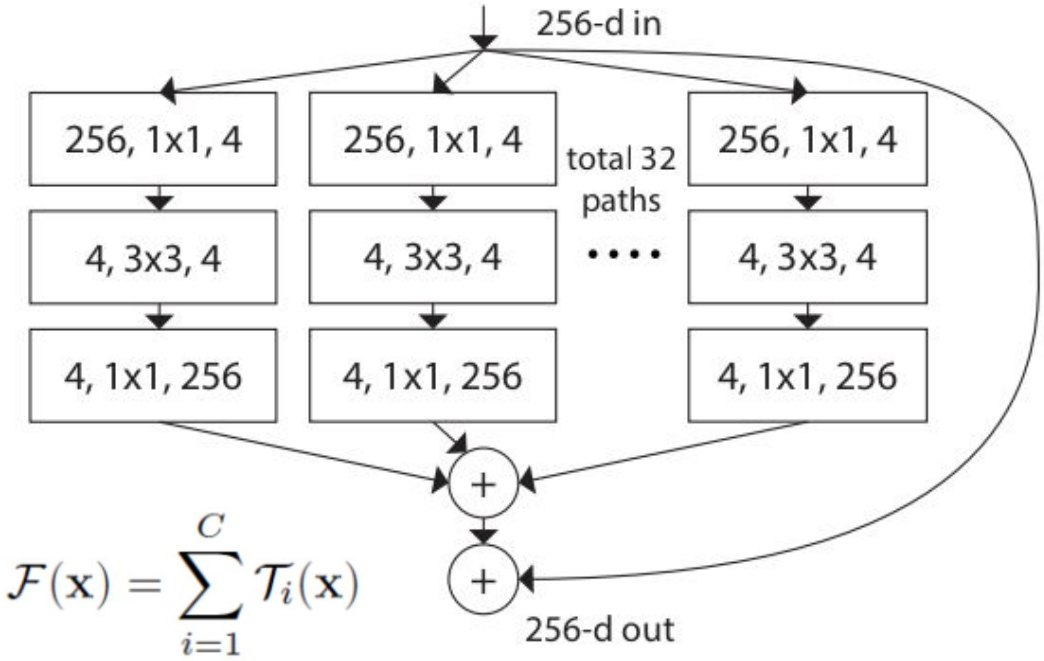
Figure: Sik-Ho Tsang



#XCeption Chollet, François. "Xception: Deep learning with depthwise separable convolutions." CVPR 2017.

ResNext

Depth-wise separable convolutions



#ResNext Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. "Aggregated residual transformations for deep neural networks." CVPR 2017 [code]

ResNext

Results on ImageNet-1K

	224×224		320×320 / 299×299	
	top-1 err	top-5 err	top-1 err	top-5 err
ResNet-101 [14]	22.0	6.0	-	-
ResNet-200 [15]	21.7	5.8	20.1	4.8
Inception-v3 [39]	-	-	21.2	5.6
Inception-v4 [37]	-	-	20.0	5.0
Inception-ResNet-v2 [37]	-	-	19.9	4.9
ResNeXt-101 (**64** × **4d**)	20.4	5.3	**19.1**	**4.4**

Winner of ILSVRC 2015

Pre-Activation ResNet

1st Runner-Up of ILSVRC 2015

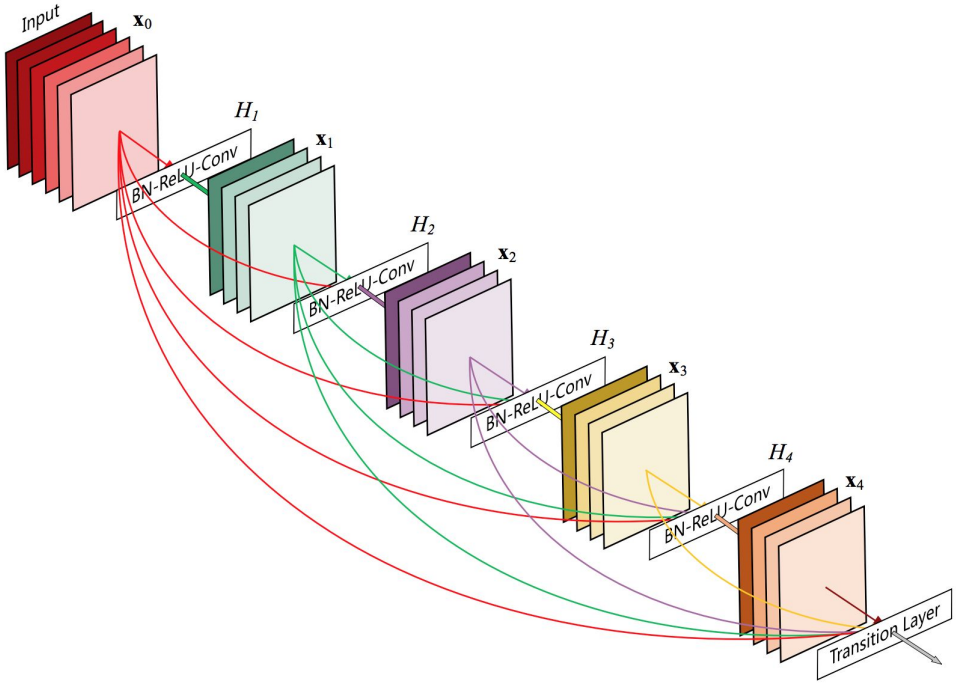
Better Than Inception-v3

Inception-v4 + ResNet

#ResNext Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. "Aggregated residual transformations for deep neural networks." CVPR 2017 [code]

Dense Blocks

The input of each layer consists of the feature maps of all earlier layer, and its output is passed to each subsequent layer.

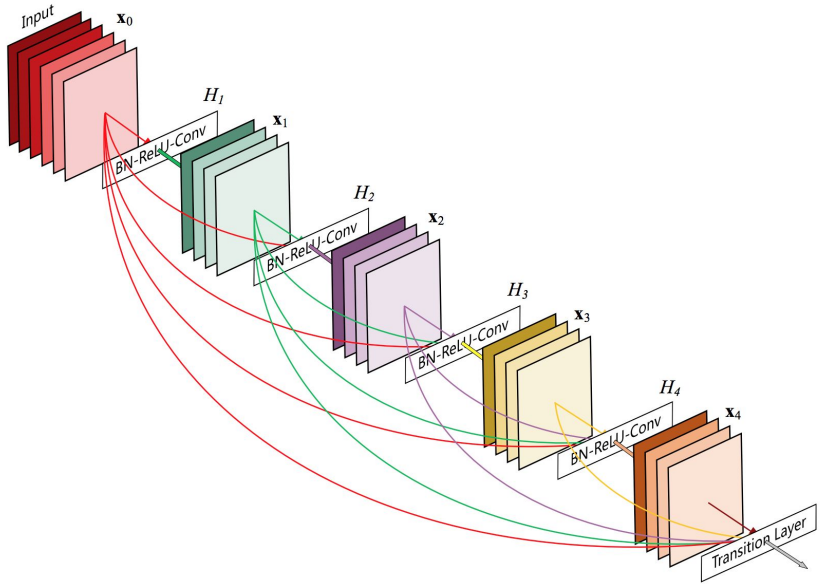


#DenseNet Huang, Gao, Zhuang Liu, Kilian Q. Weinberger, and Laurens van der Maaten. "Densely connected convolutional networks." CVPR 2017. [code]

Dense Blocks

Growth rate (k): If each function H_l produces k feature maps, it follows that the l^{th} layer has $k_0 + k \times (l-1)$ input feature-maps, where k_0 is the number of channels in the input layer.

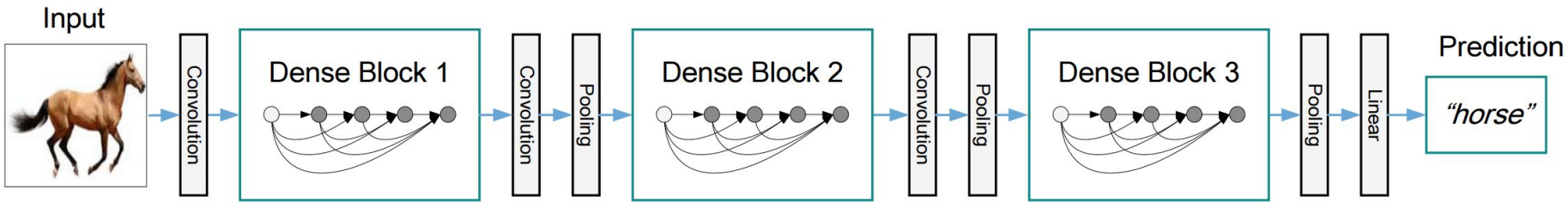
Layer (l)	Growth rate (k)	# input f. maps
0	k_0=5	-
1	4	5
2	4	9
3	4	13
4	4	17



#DenseNet Huang, Gao, Zhuang Liu, Kilian Q. Weinberger, and Laurens van der Maaten. "Densely connected convolutional networks." CVPR 2017. [code]

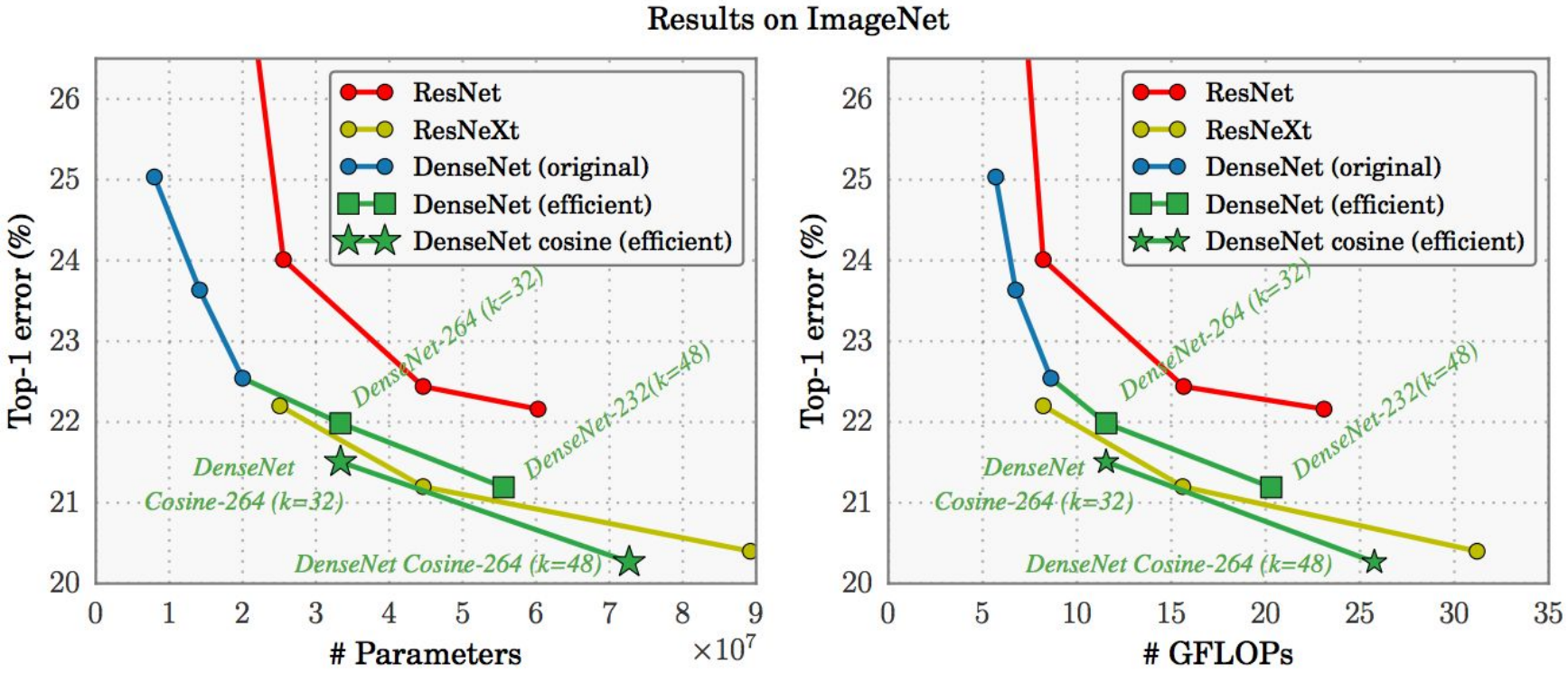
Dense Blocks

DenseNet is divided into multiple densely connected dense blocks to facilitate spatial **pooling**.



#DenseNet Huang, Gao, Zhuang Liu, Kilian Q. Weinberger, and Laurens van der Maaten. "Densely connected convolutional networks." CVPR 2017. [code]

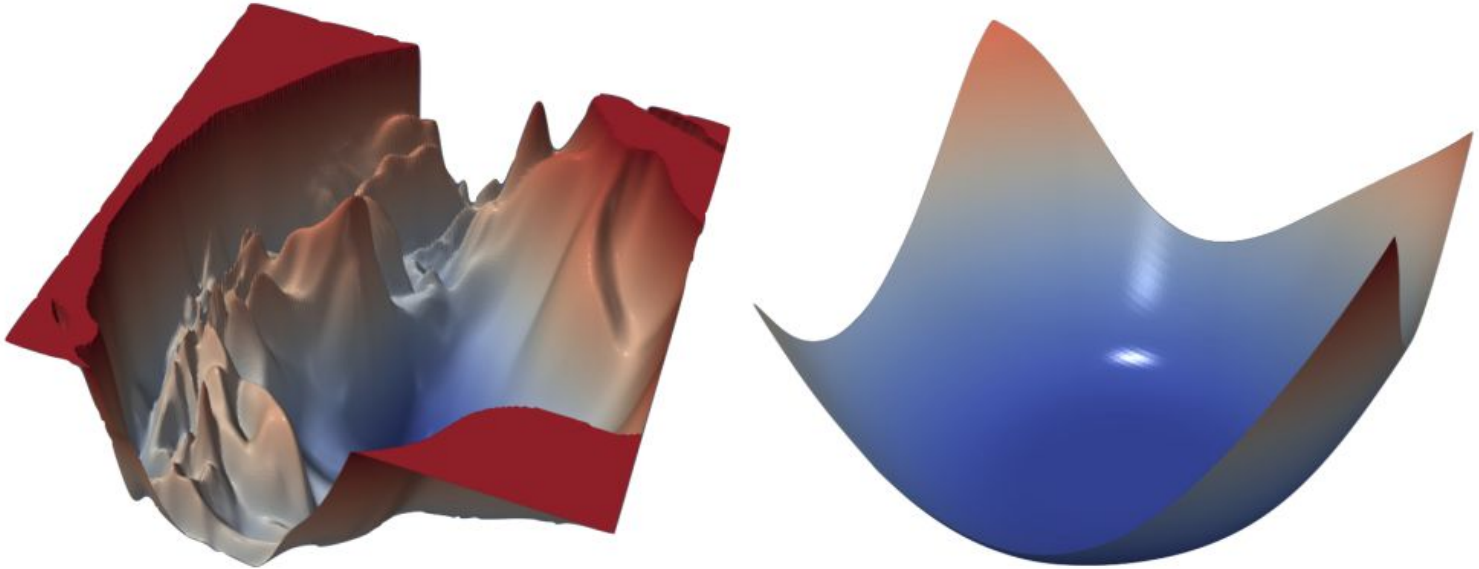
DenseNet



#DenseNet Huang, Gao, Zhuang Liu, Kilian Q. Weinberger, and Laurens van der Maaten. "Densely connected convolutional networks." CVPR 2017. [code]

DenseNets

The loss surfaces of ResNet-110-noshort and DenseNet for CIFAR-10.



(a) ResNet-110, no skip connections

(b) DenseNet, 121 layers

Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer and Tom Goldstein. Visualizing the Loss Landscape of Neural Nets. NeurIPS, 2018. [code] [poster]