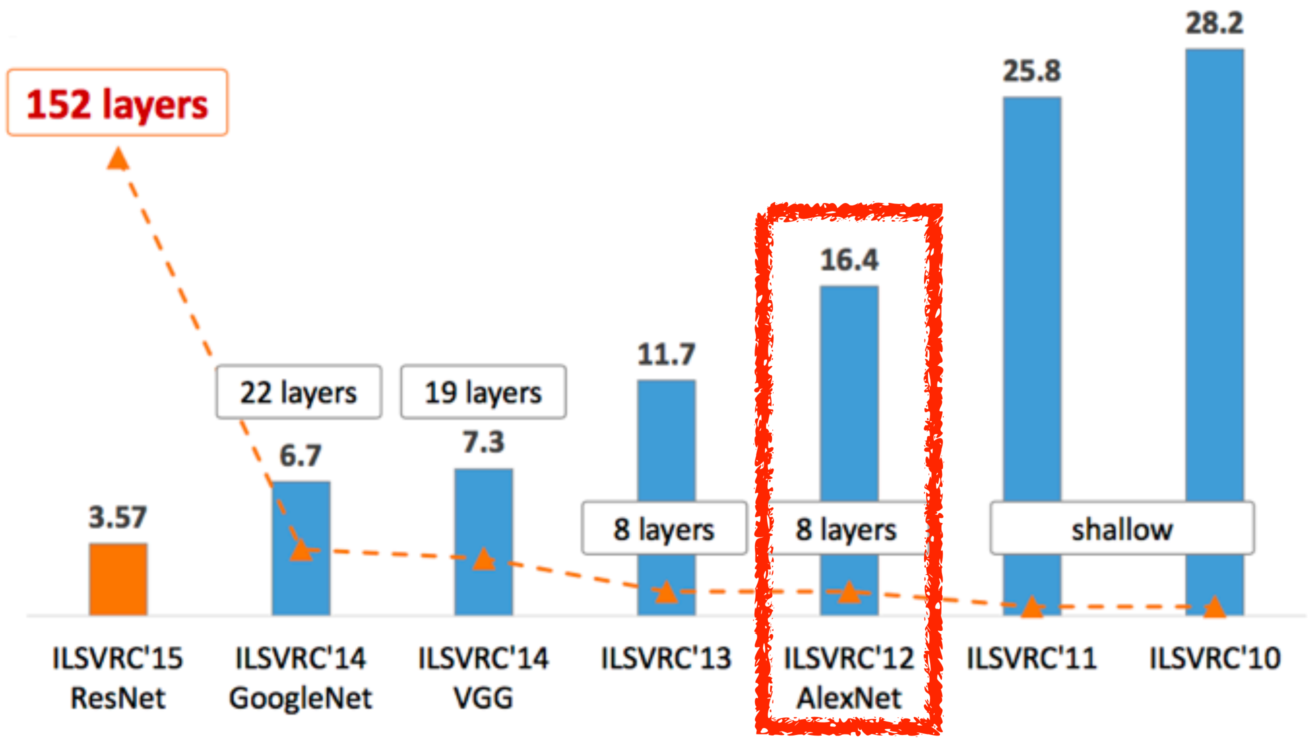
Deep Learning Basics

Lecture 5: Modern Convolutional Neural Networks

최성준 (고려대학교 인공지능학과)

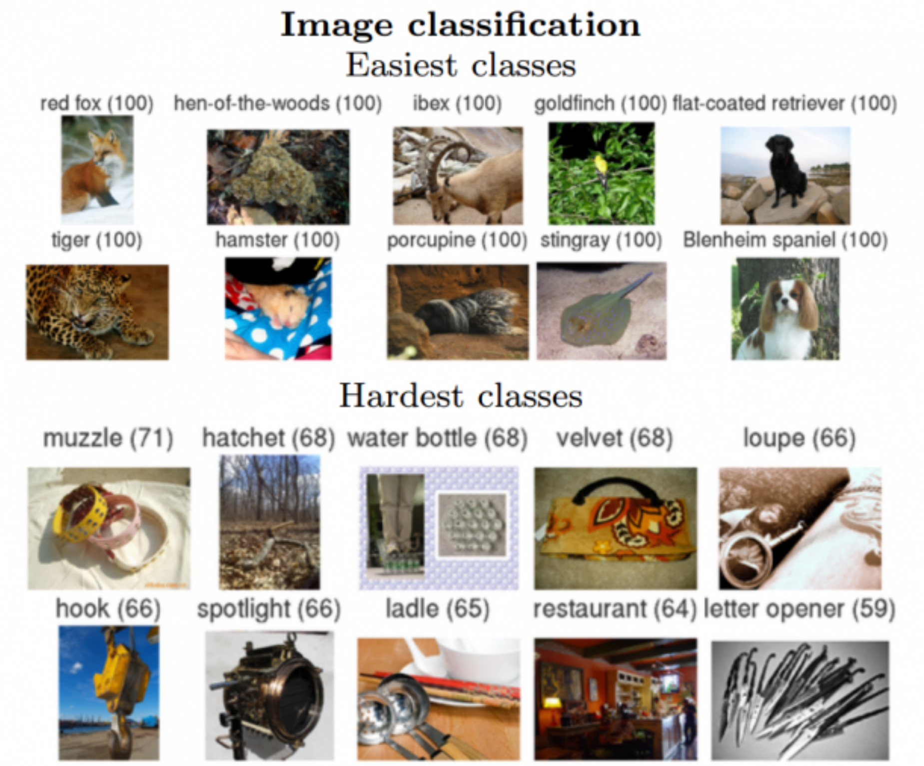
AlexNet



Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," NIPS, 2012

ILSVRC

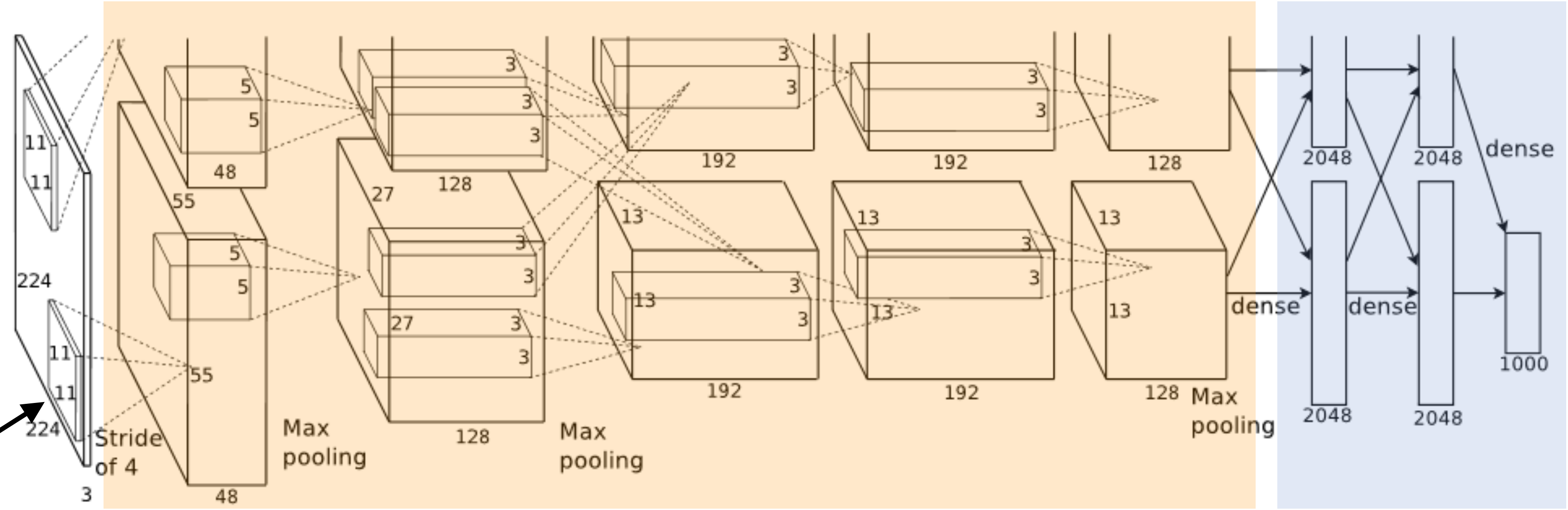
- ImageNet Large-Scale Visual Recognition Challenge
  - Classification / Detection / Localization / Segmentation
  - 1,000 different categories
  - Over 1 million images
  - Training set: 456,567 images



ILSVRC

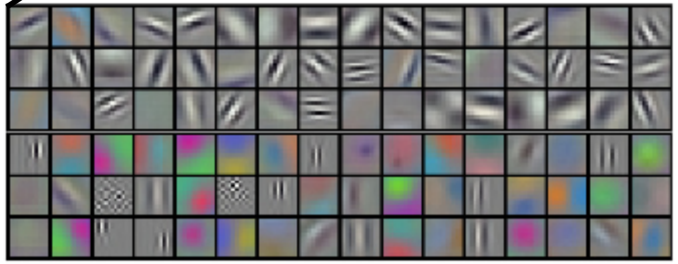
Year	Error Rate
2010	28.2%
2011	25.8%
2012	16.4%
2013	11.2%
2014	6.7%
2015	3.5%
Human	About 5.1%

AlexNet



5 convolutional layers

3 dense layers



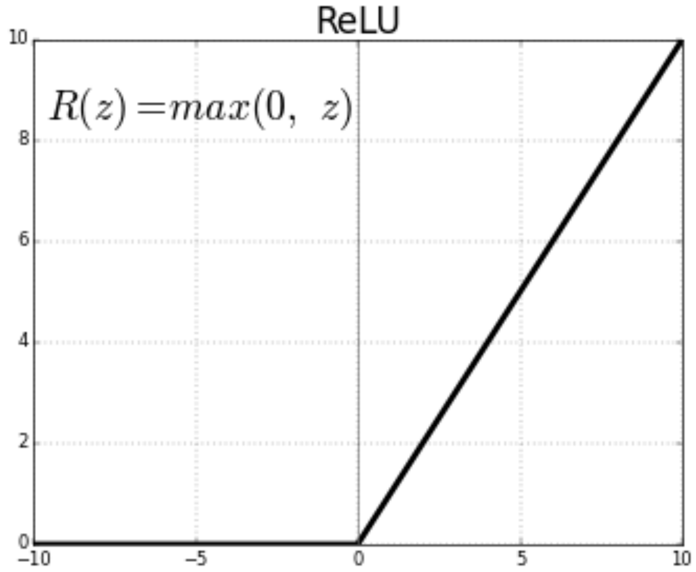
Learned 11x11x3 filters

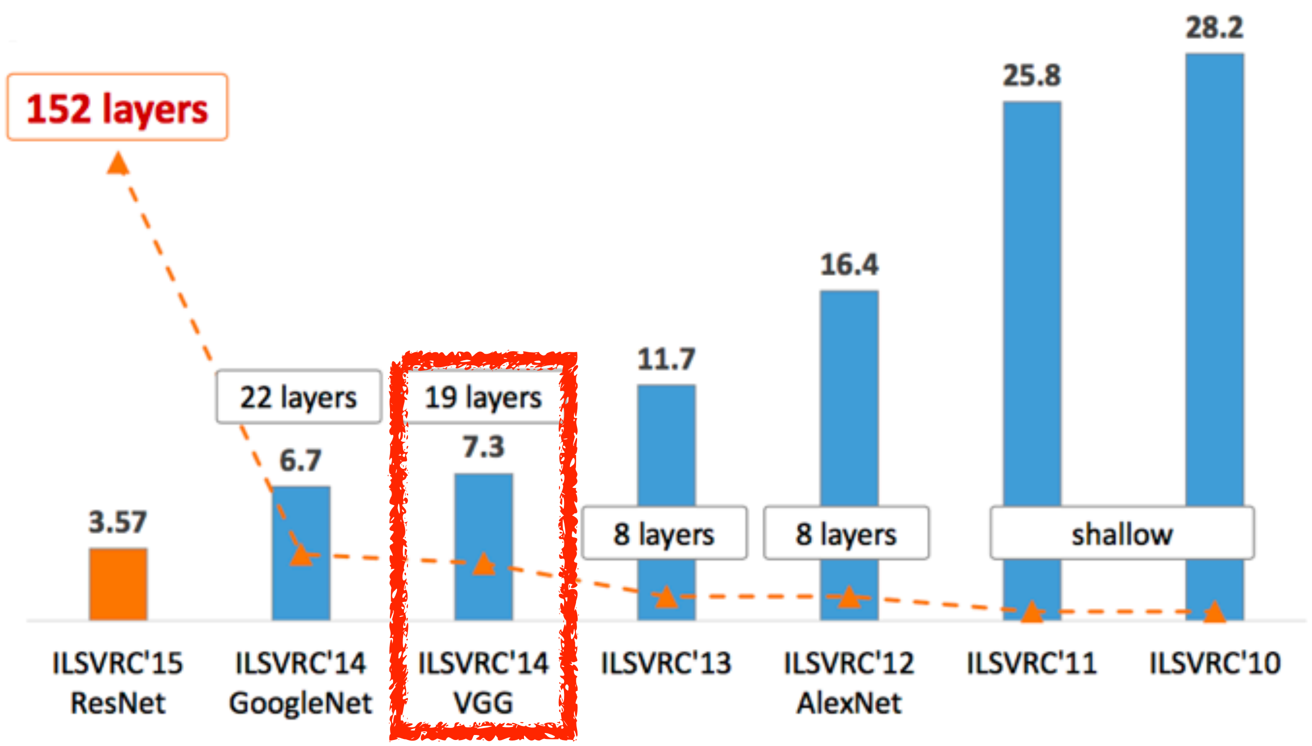
AlexNet

- Key ideas
  - Rectified Linear Unit (ReLU) activation
  - GPI implementation (2 GPUs)
  - Local response normalization, Overlapping pooling
  - Data augmentation
  - Dropout

AlexNet

- ReLU Activation
  - Preserves properties of linear models
  - Easy to optimize with gradient descent
  - Good generalization
  - Overcome the vanishing gradient problem

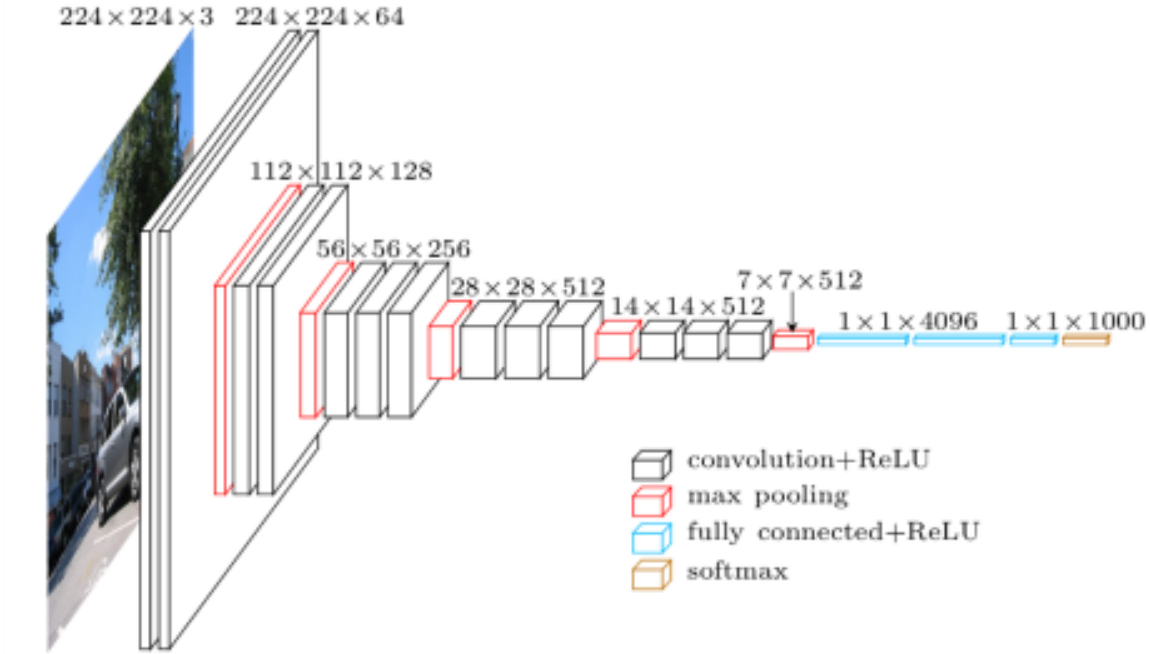




VGGNet

Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," ICLR, 2015

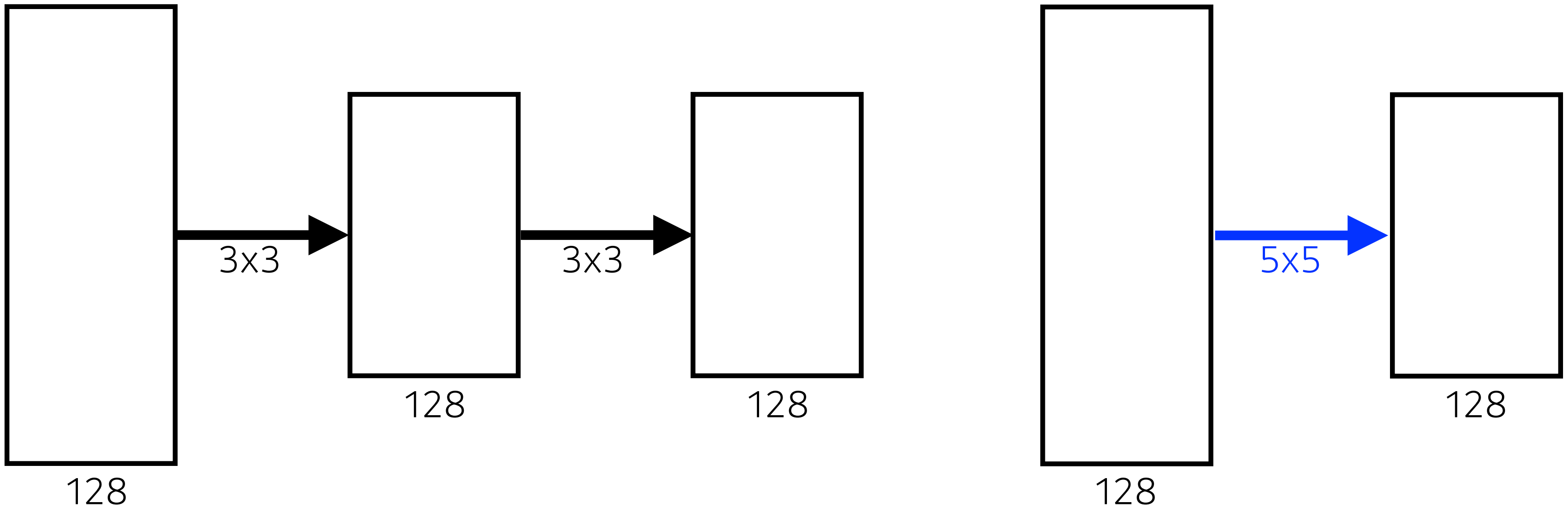
VGGNet



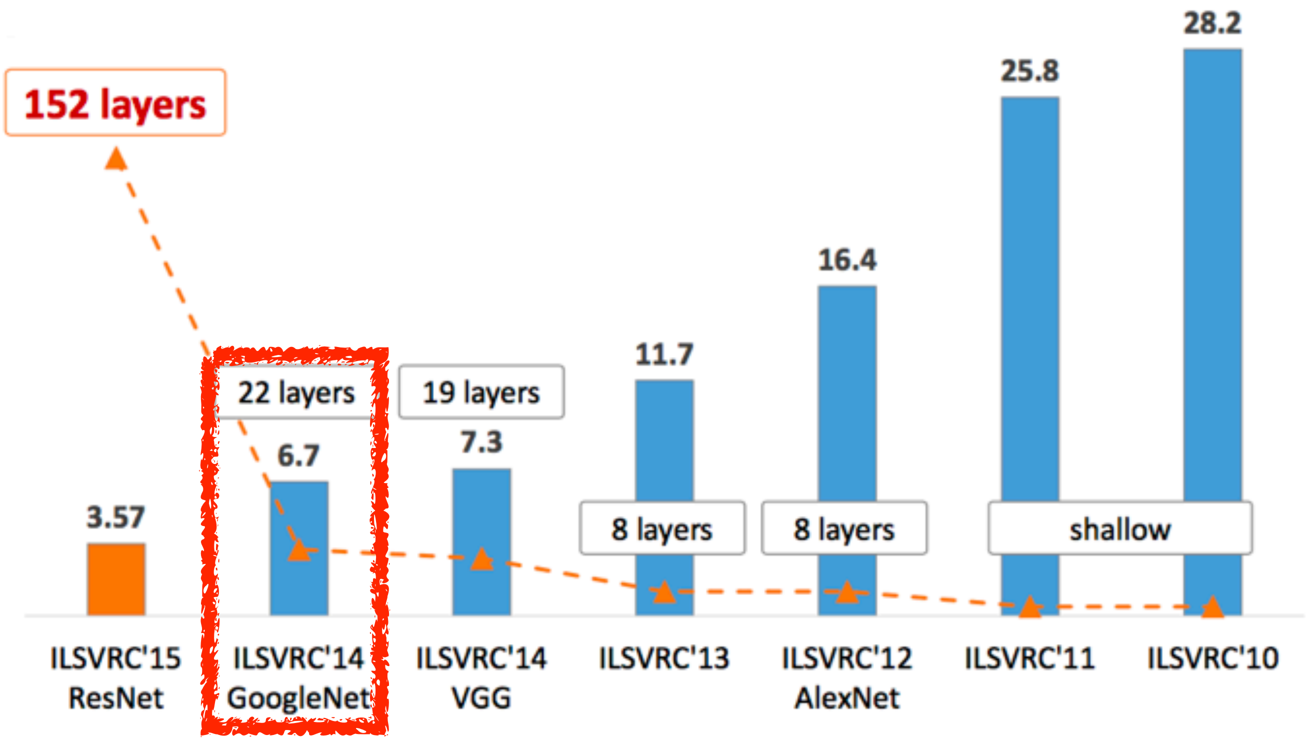
- Increasing depth with 3×3 convolution filters (with stride 1)
- 1x1 convolution for fully connected layers
- Dropout (p=0.5)
- VGG16, VGG19

VGGNet

- Why 3×3 convolution?



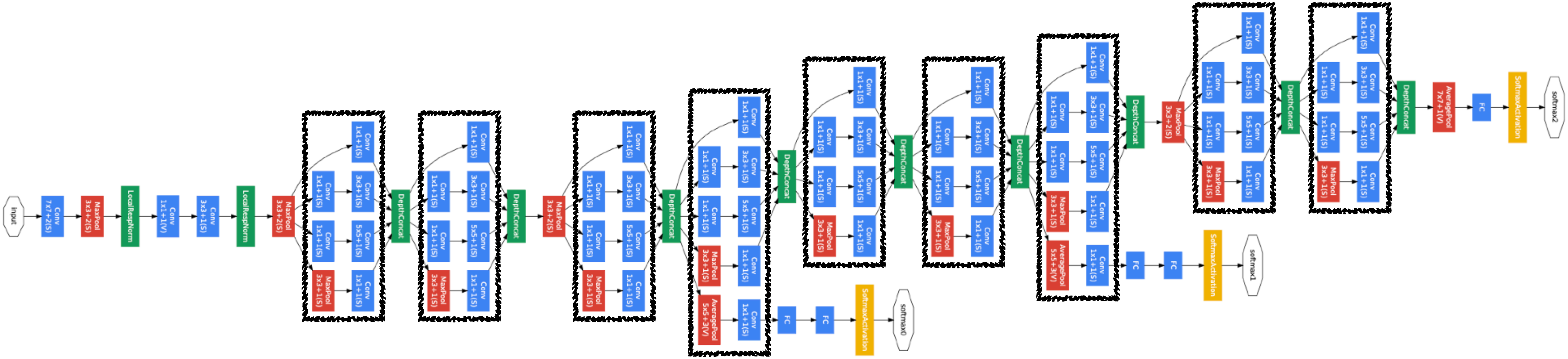
Receptive field	5x5	5x5
# of params	3x3x128x128+3x3x128x128 = **294,912**	5*5*128*128 = **409,600**



GoogLeNet

Christian et al. “Going Deeper with Convolutions”, CVPR, 2015

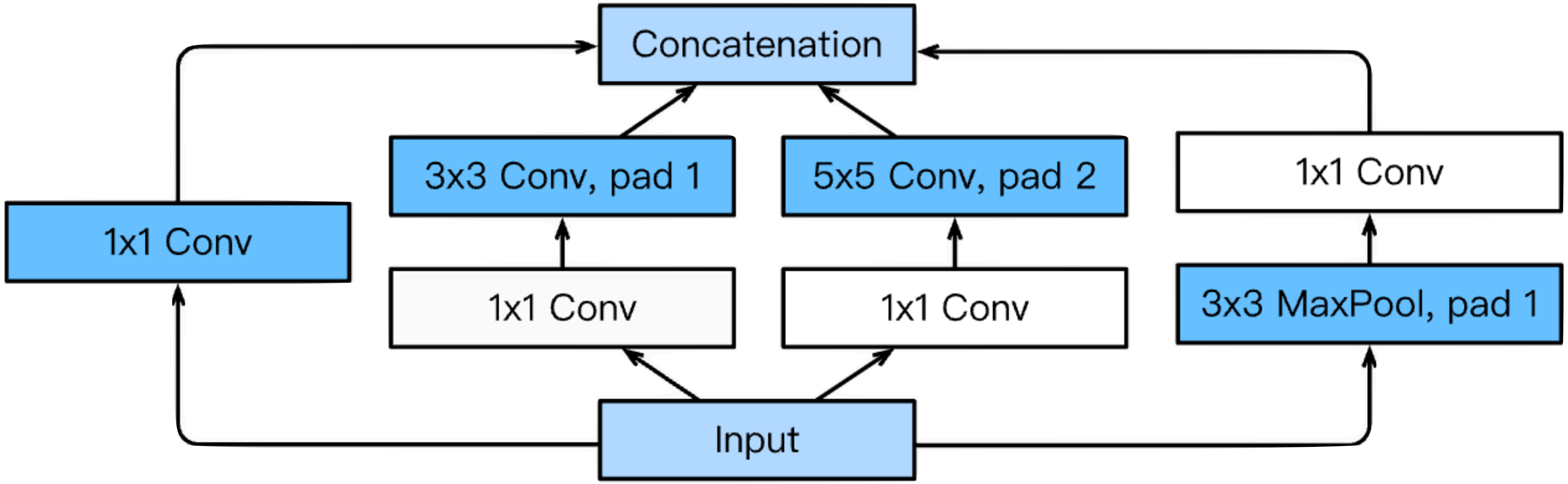
GoogLeNet



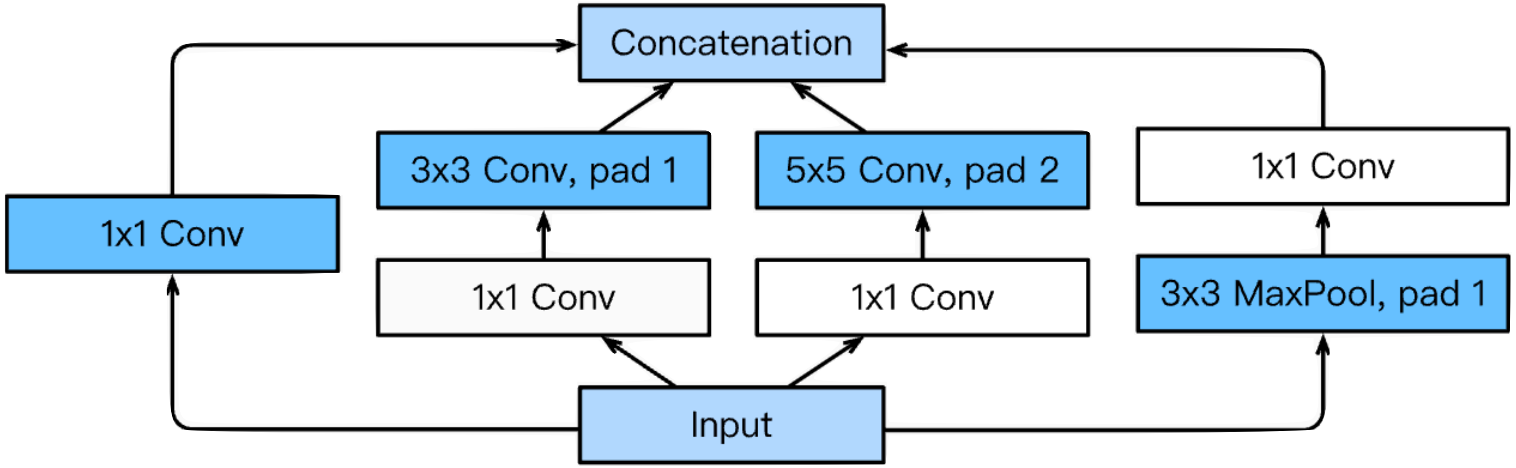
22 layers

GoogLeNet

- GoogLeNet won the ILSVRC at 2014
  - It combined network-in-network (NiN) with inception blocks.
- Inception blocks



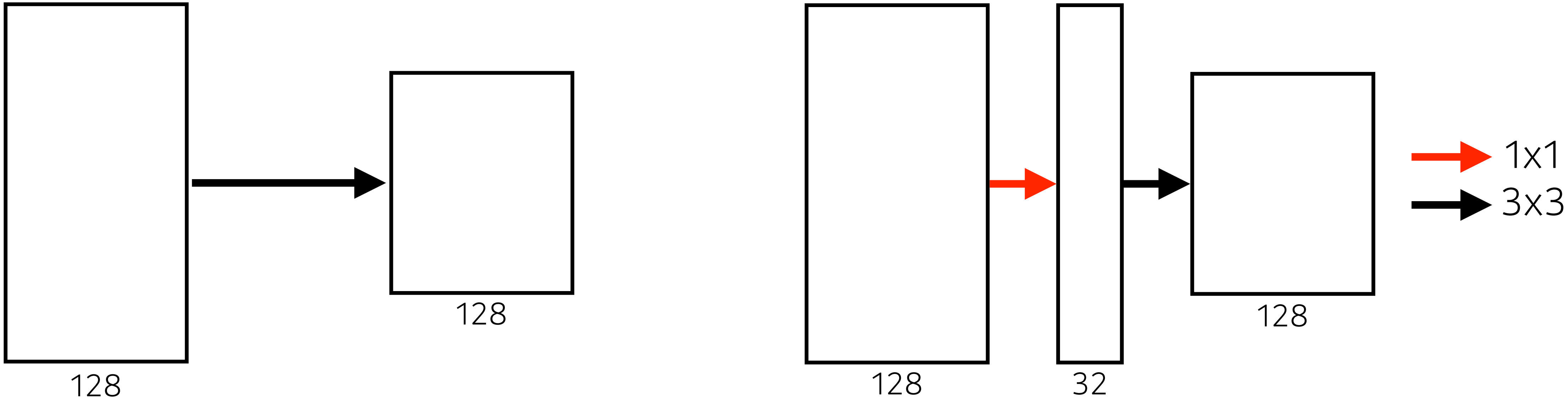
Inception Block



- What are the benefits of the inception block?
  - Reduce the number of parameter.
- How?
  - Recall how the number of parameters is computed.
  - 1x1 convolution can be seen as channel-wise dimension reduction.

Inception Block

- Benefit of 1x1 convolution



$$3 \times 3 \times 128 \times 128 = 147,456$$

$$1 \times 1 \times 128 \times 32 = 4,096$$

$$3 \times 3 \times 32 \times 128 = 36,864$$

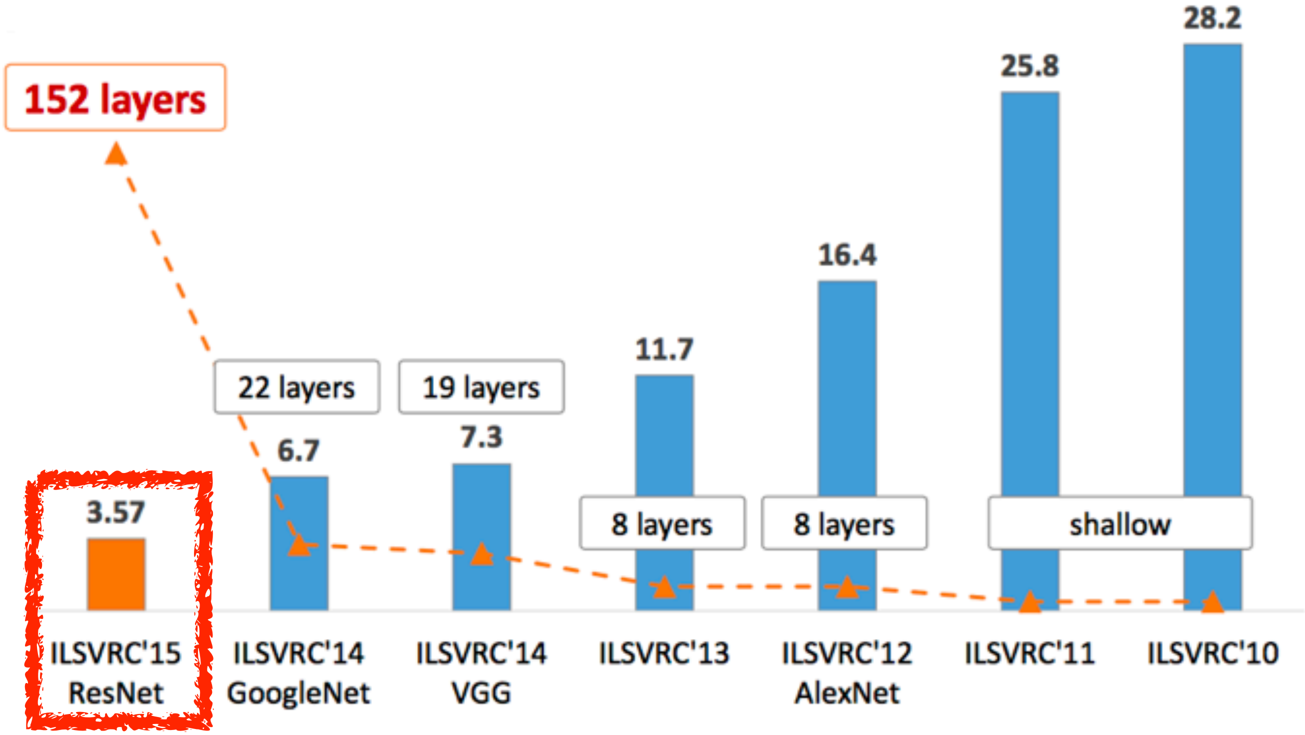
$$4,096 + 36,864 = 40,960$$

1x1 convolution enables about 30% reduce of the number of parameters!

Quiz

- Which CNN architecture has the least number of parameters?
  1. AlexNet (8-layers) **(60M)**
  2. VGGNet (19-layers) **(110M)**
  3. GoogLeNet (22-layers) **(4M)**

- The answer is GoogLeNet.

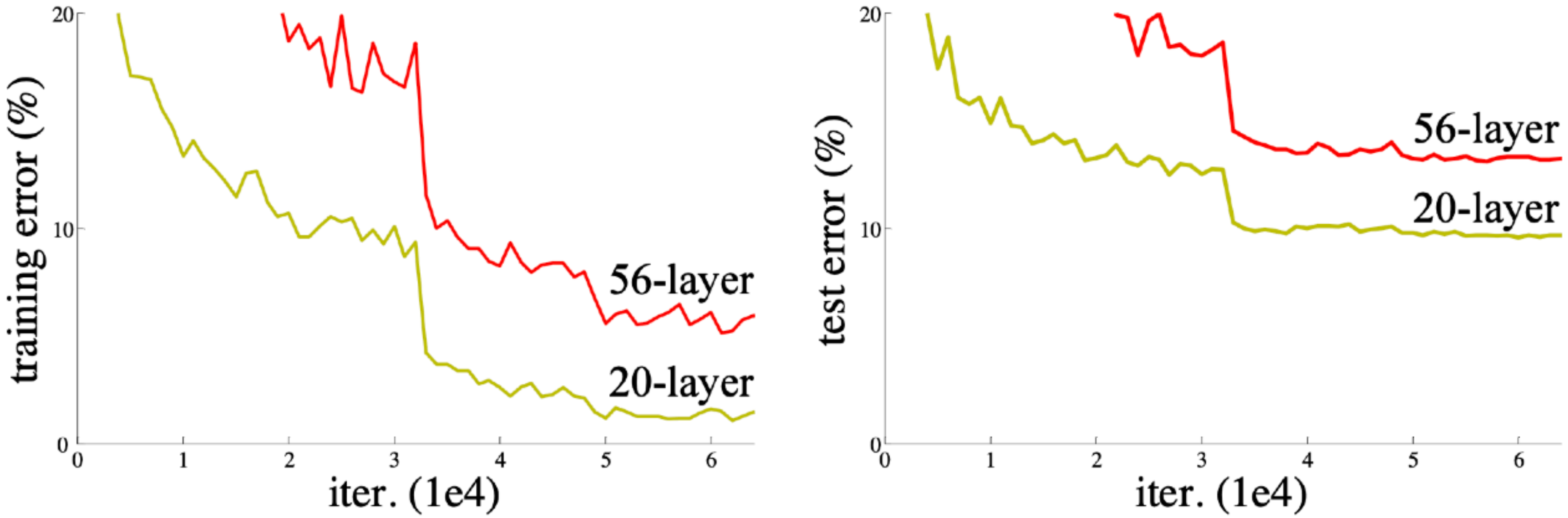


ResNet

Kaiming He, Xiangyu Zhang, Shaoquing Ren, Jian Sun, "Deep Residual Learninig for Image Recognition,", CVPR, 2015

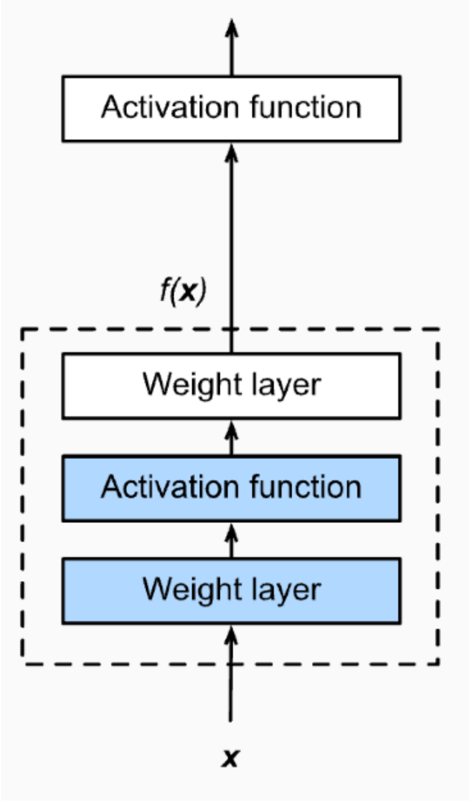
ResNet

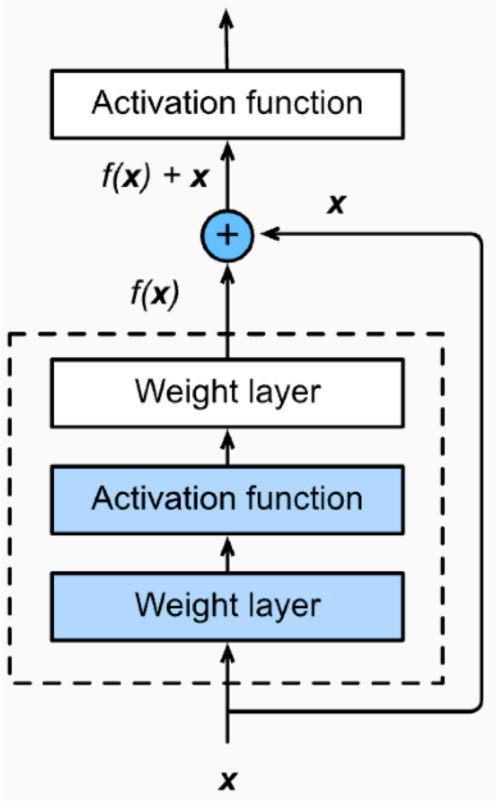
- Deeper neural networks are hard to train.
  - Overfitting is usually caused by an excessive number of parameters.
  - But, not in this case.



ResNet

- Add an identity map (skip connection)

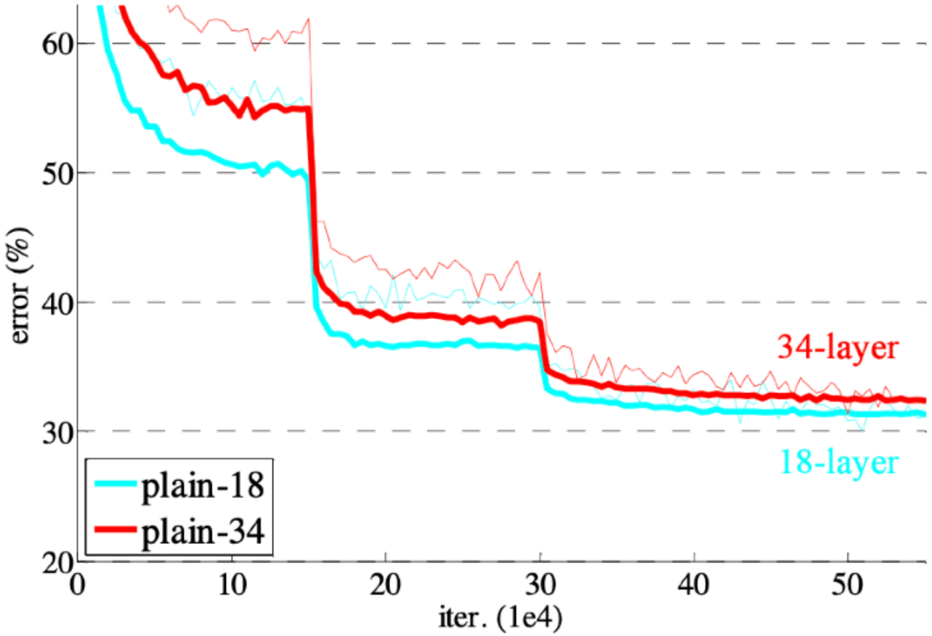


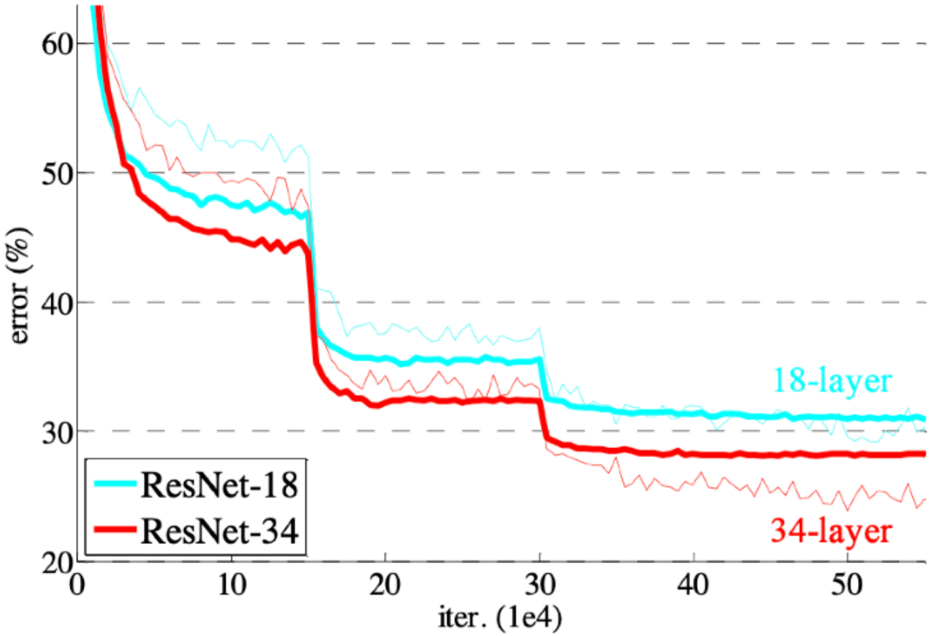


skip connection:
$f(x) \to x + f(x)$

ResNet

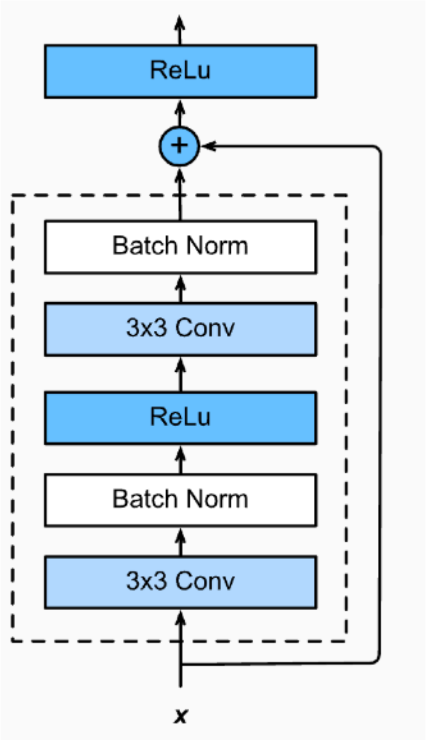
- Add an identity map (skip connection)



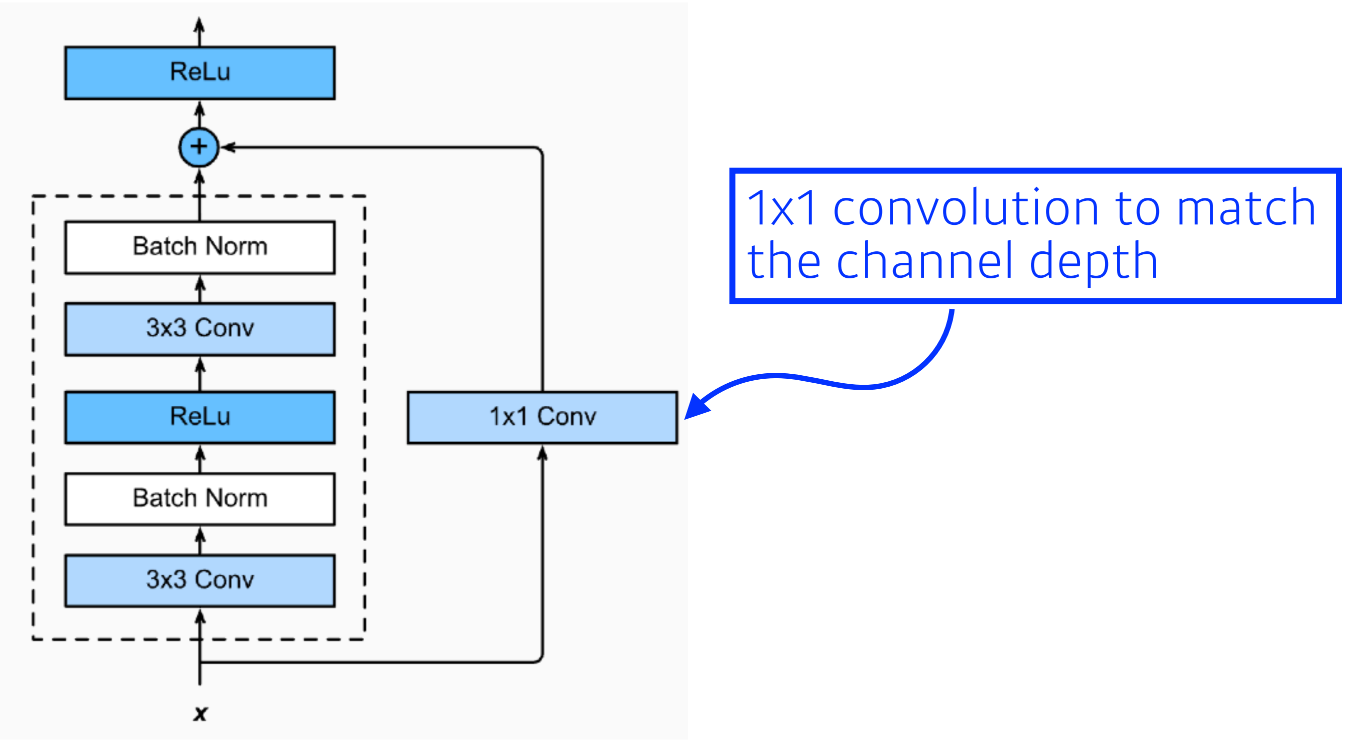


ResNet

- Add an identity map **after** nonlinear activations:



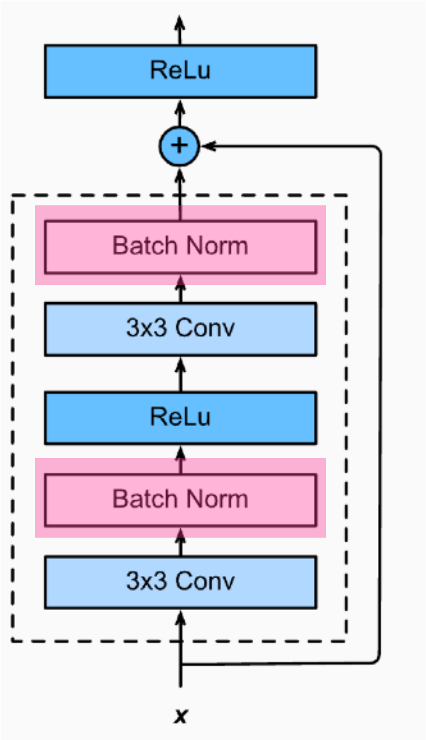
Simple Shortcut



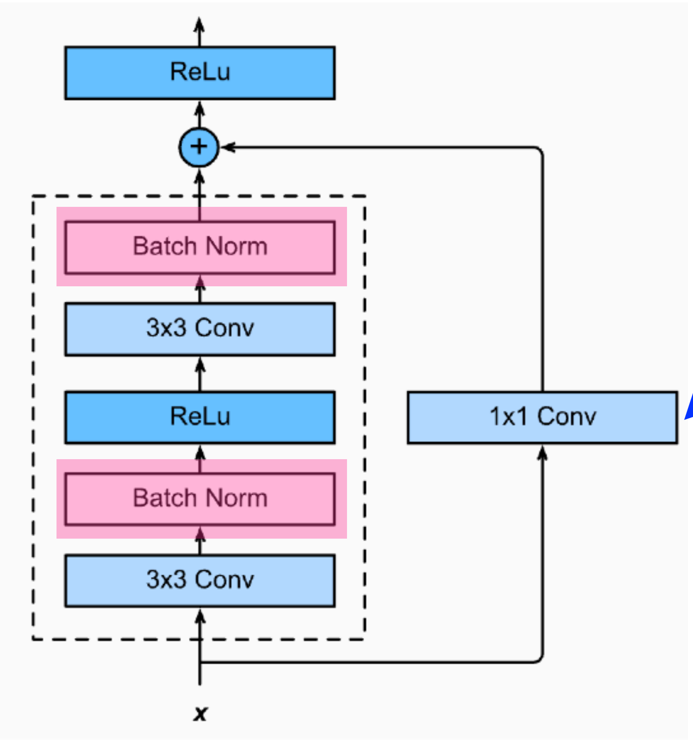
Projected Shortcut

ResNet

- Batch normalization **after** convolutions:



Simple Shortcut

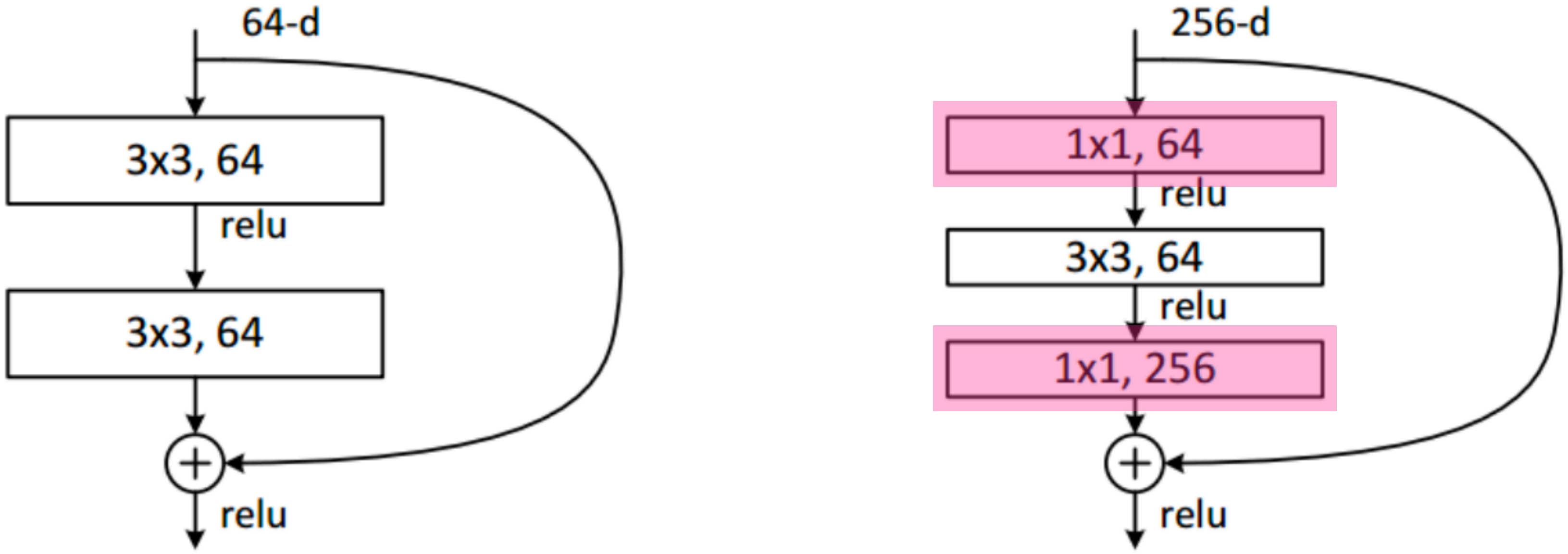


1x1 convolution to match the channel depth

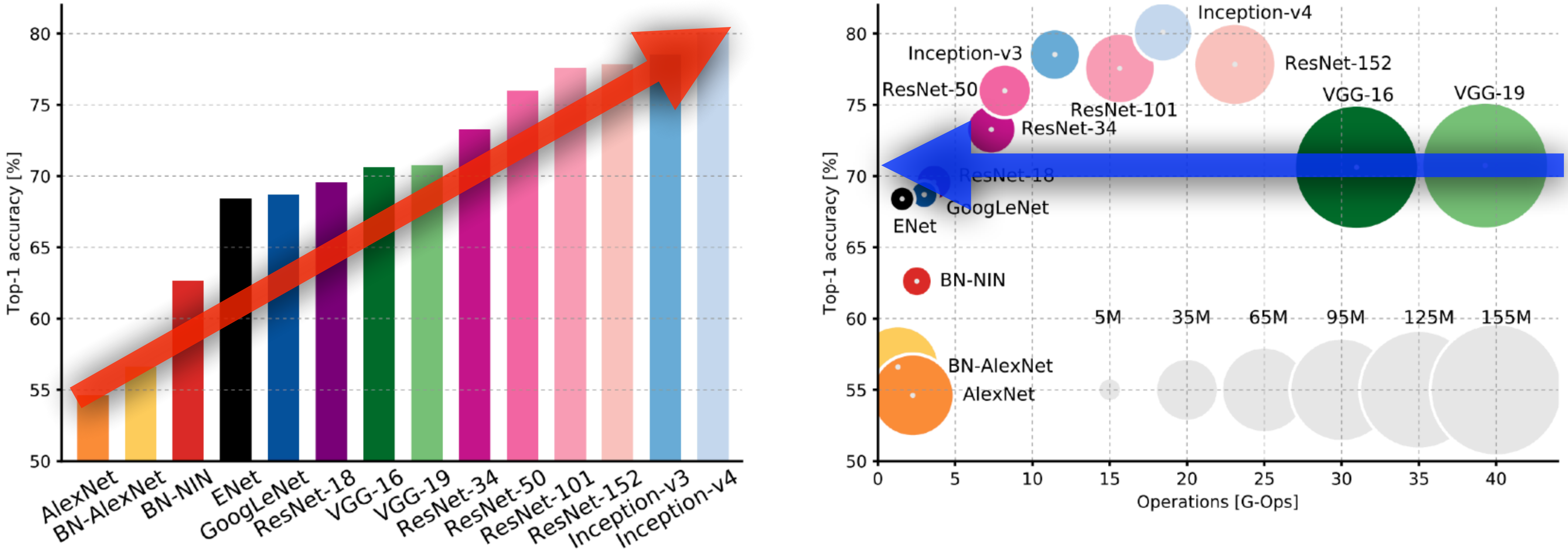
Projected Shortcut

ResNet

- Bottleneck architecture



ResNet



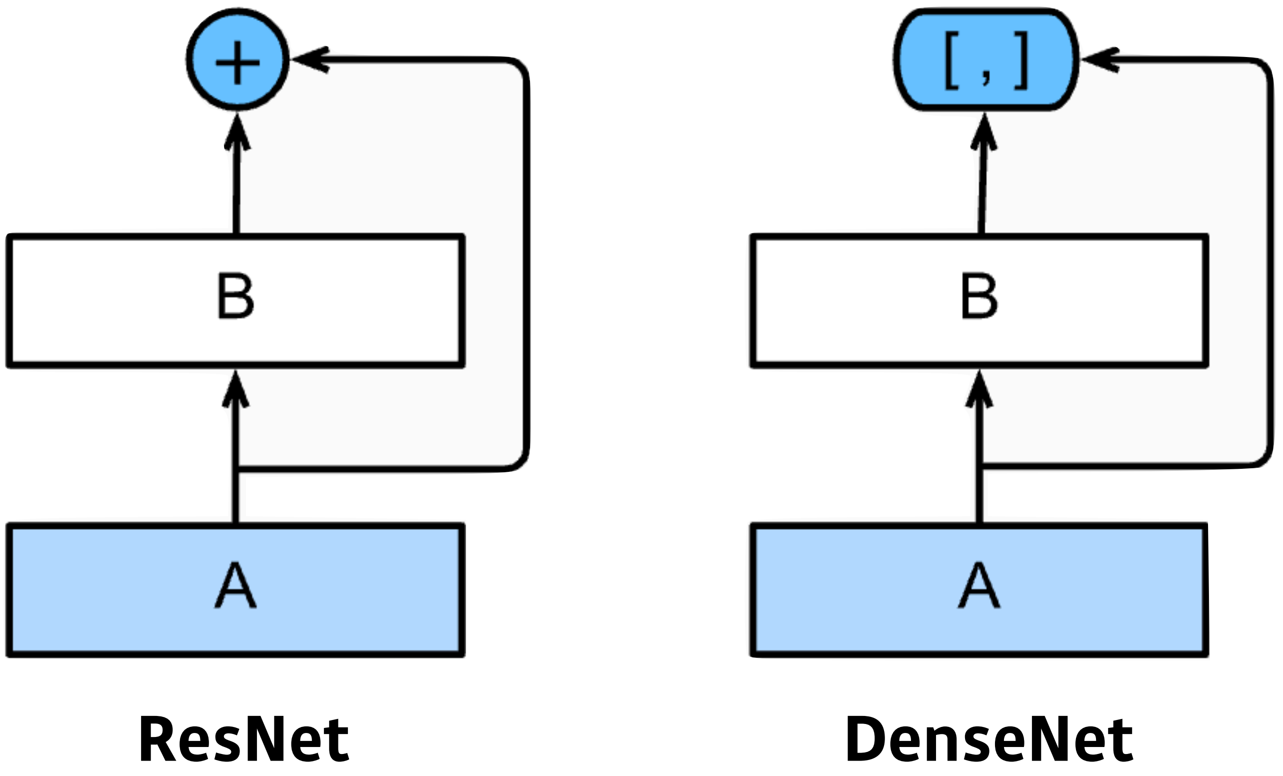
- **Performance** increases while **parameter size** decreases.

DenseNet

Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Weinberger,
“Densely Connected Convolutional Networks,” CVPR, 2017

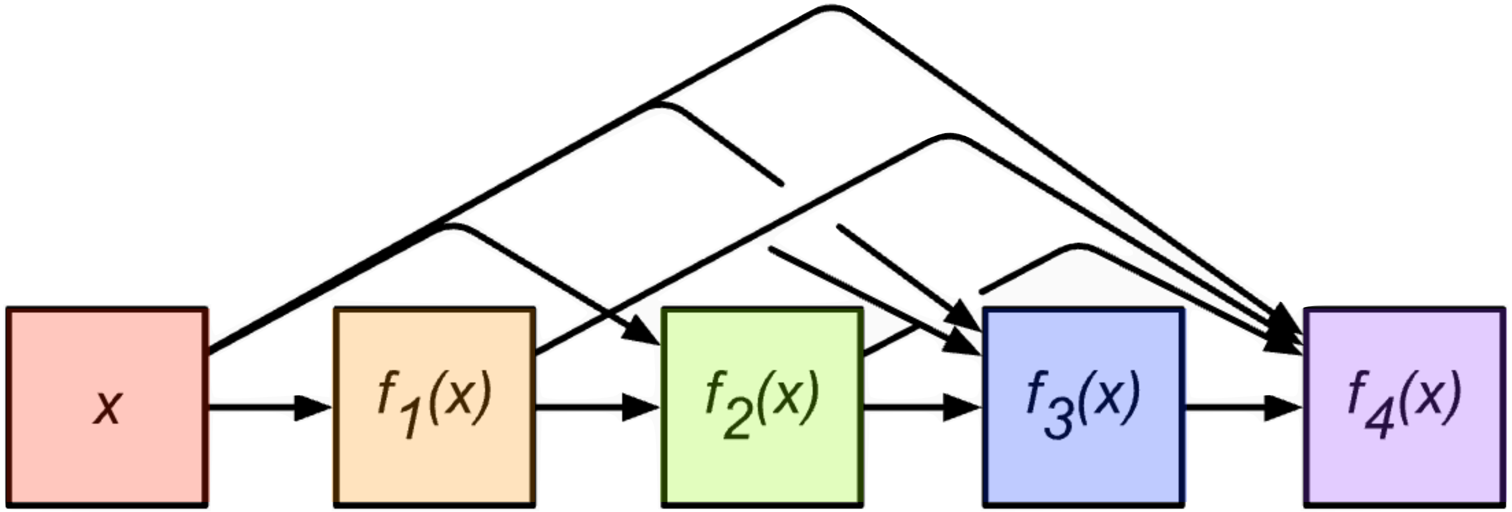
DenseNet

- DenseNet uses **concatenation** instead of **addition**.



DenseNet

- DenseNet uses **concatenation** instead of **addition**.

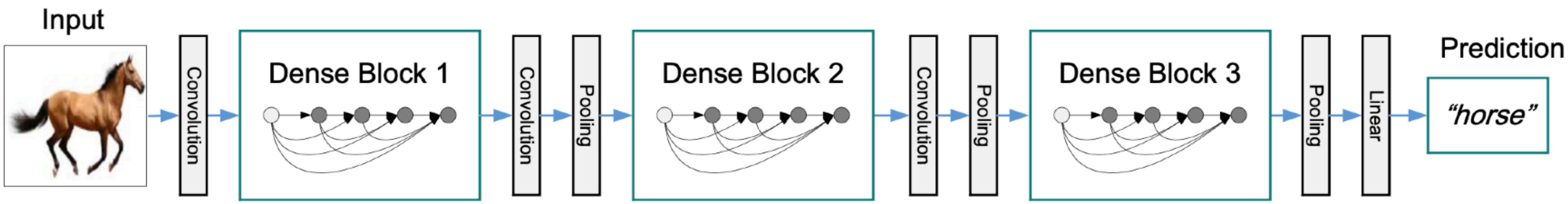


Dense Block

$$\mathbf{x} \mapsto [\mathbf{x}, f_1(\mathbf{x}), f_2(\mathbf{x}, f(\mathbf{x})), f_3(\mathbf{x}, f_1(\mathbf{x}), f_2(\mathbf{x}, f_1(\mathbf{x}))), f_4(\mathbf{x}, f_1(\mathbf{x}), f_2(\mathbf{x}, f_1(\mathbf{x})), f_3(\mathbf{x}, f_1(\mathbf{x}), f_2(\mathbf{x}, f_1(\mathbf{x}))))]$$

DenseNet

- Dense Block
  - Each layer concatenates the feature maps of all preceding layers.
  - The number of channels increases geometrically.
- Transition Block
  - BatchNorm -> 1x1 Conv -> 2x2 AvgPooling
  - Dimension reduction



Summary

- Key takeaways
  - **VGG:** repeated 3x3 blocks
  - **GoogLeNet:** 1x1 convolution
  - **ResNet:** skip-connection
  - **DenseNet:** concatenation

Thank you for listening