Ch06. Convolutional Neural Network (CNN)

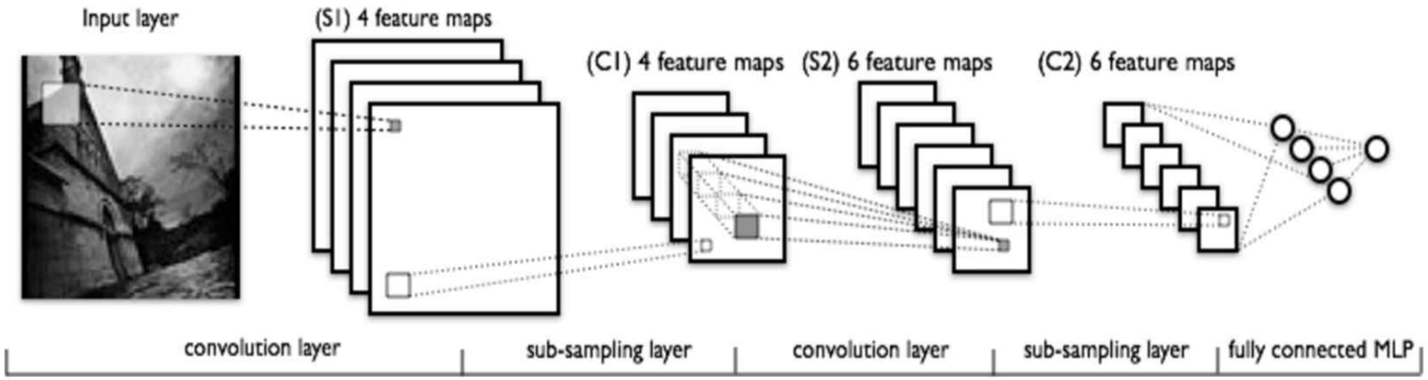
Hwanjo Yu

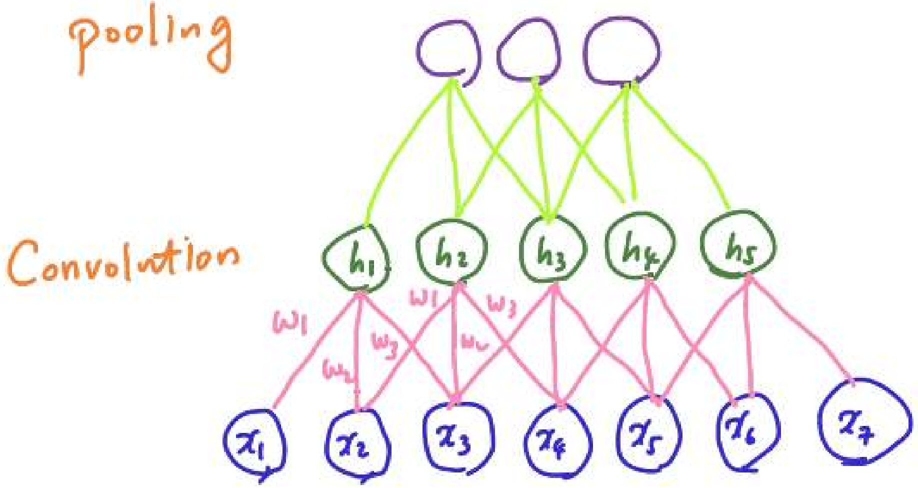
POSTECH

http://hwanjoyu.org

LeNet-5

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998), "Gradient-based learning applied to document recognition”, Proceedings of the IEEE.





ImageNet Challenge

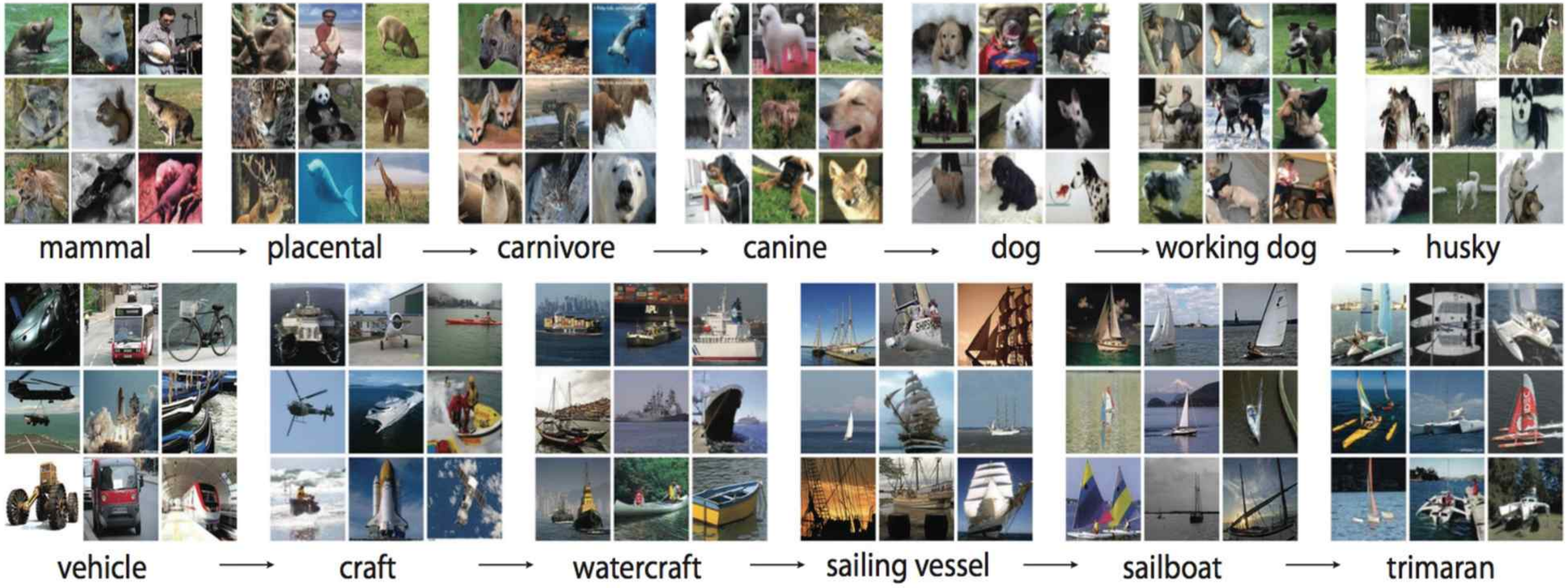
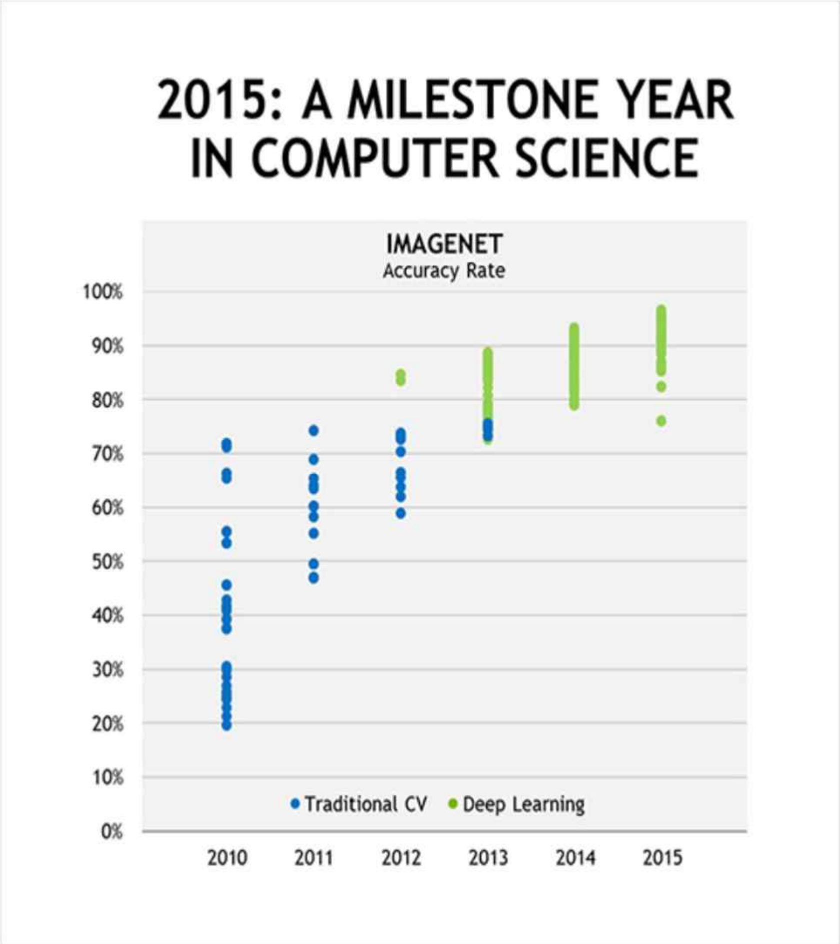


Figure 1: A snapshot of two root-to-leaf branches of ImageNet: the **top** row is from the mammal subtree; the **bottom** row is from the vehicle subtree. For each synset, 9 randomly sampled images are presented.

2015: A Milestone Year in Computer Science

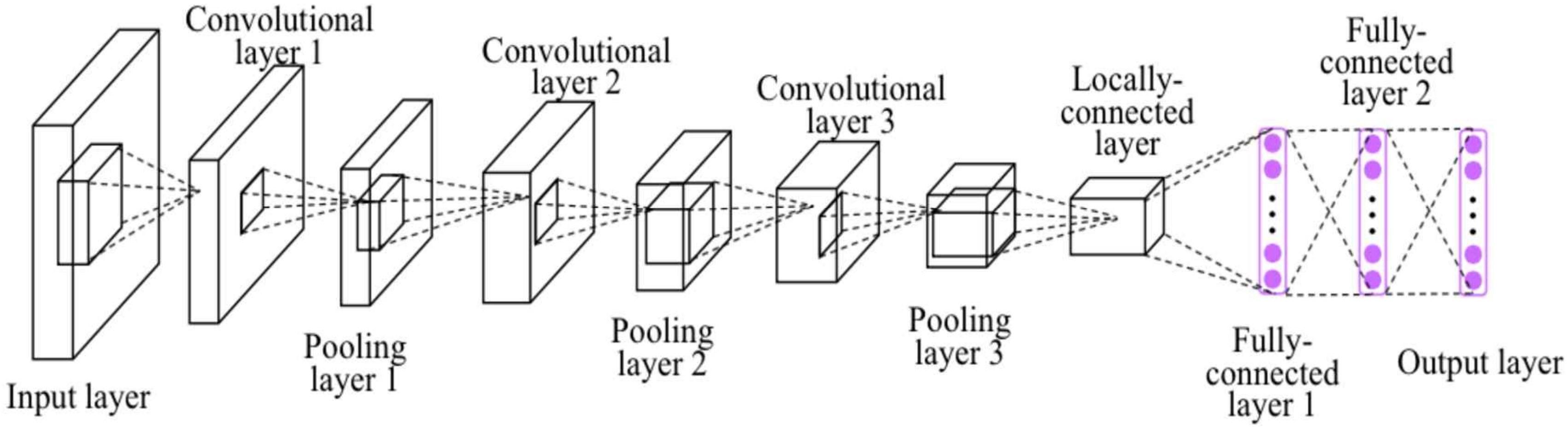


- AlexNet (5 convolutional layers + 3 fully connected layers), 2012
- VGG (very deep CNN, 16-19 weight layers), 2015
- GoogLeNet (22 layers), 2015
- Deep Residual Net (152 layers), 2015

https://blogs.nvidia.com/blog/2016/01/12/accelerating-ai-articial-intelligence-gpus/

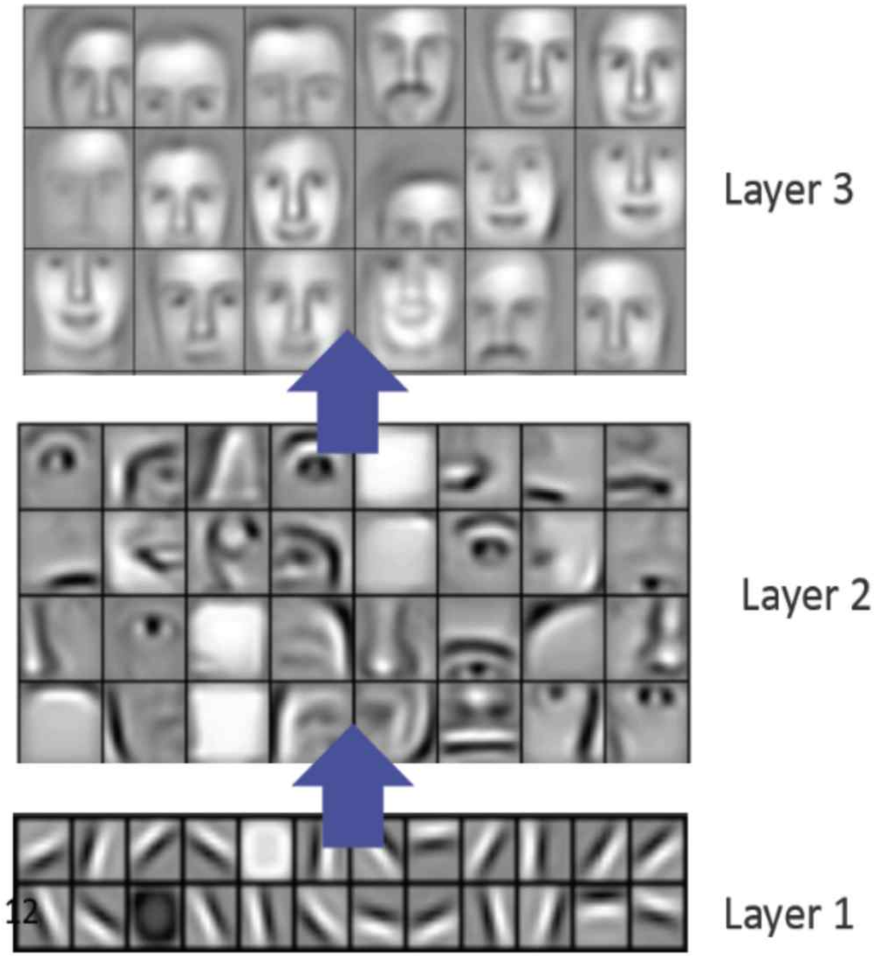
Why CNN so successful?

- Similar to simple and complex cells in V1 area of visual cortex
- Deep architecture
- Supervised representation learning: feature extraction + classification



CNN learns hierarchical representations

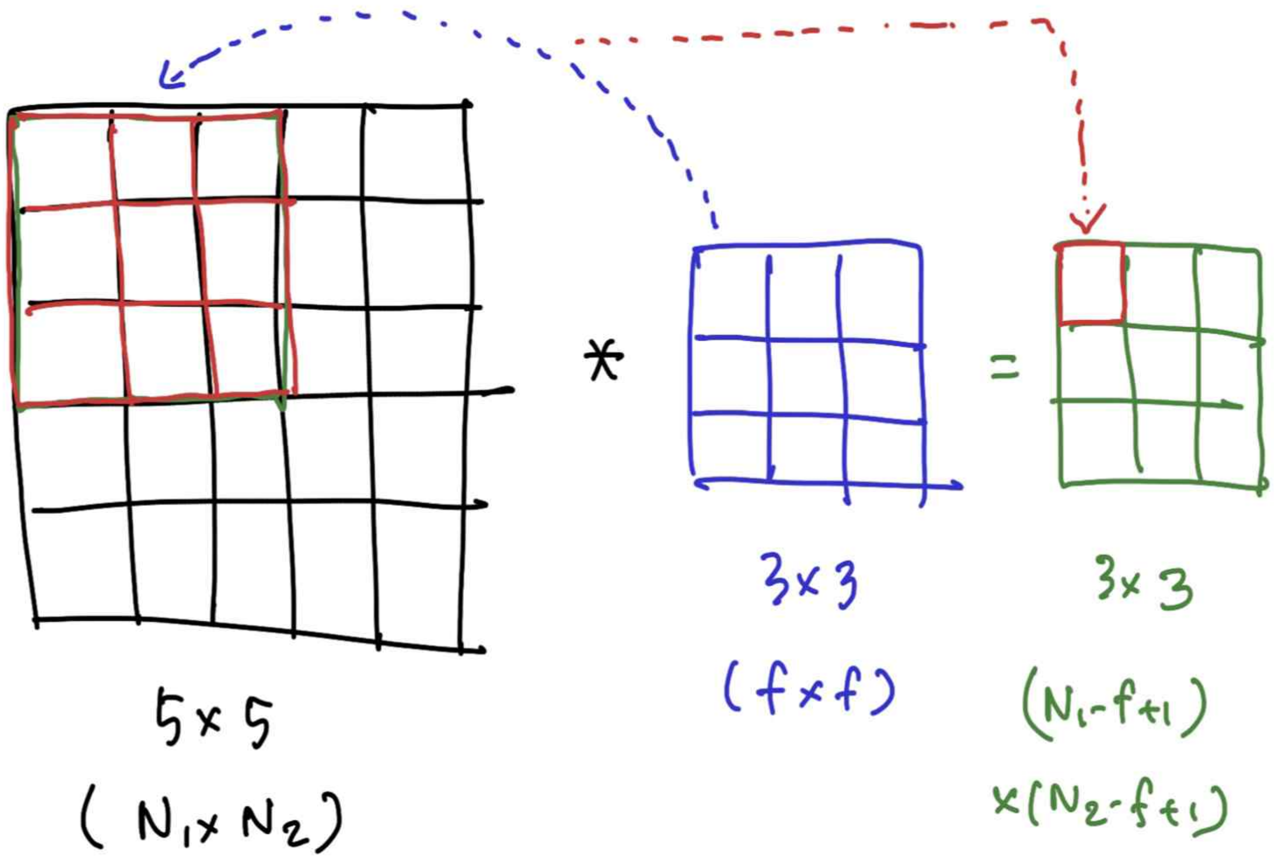
H. Lee et al. (2009), "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," ICML.



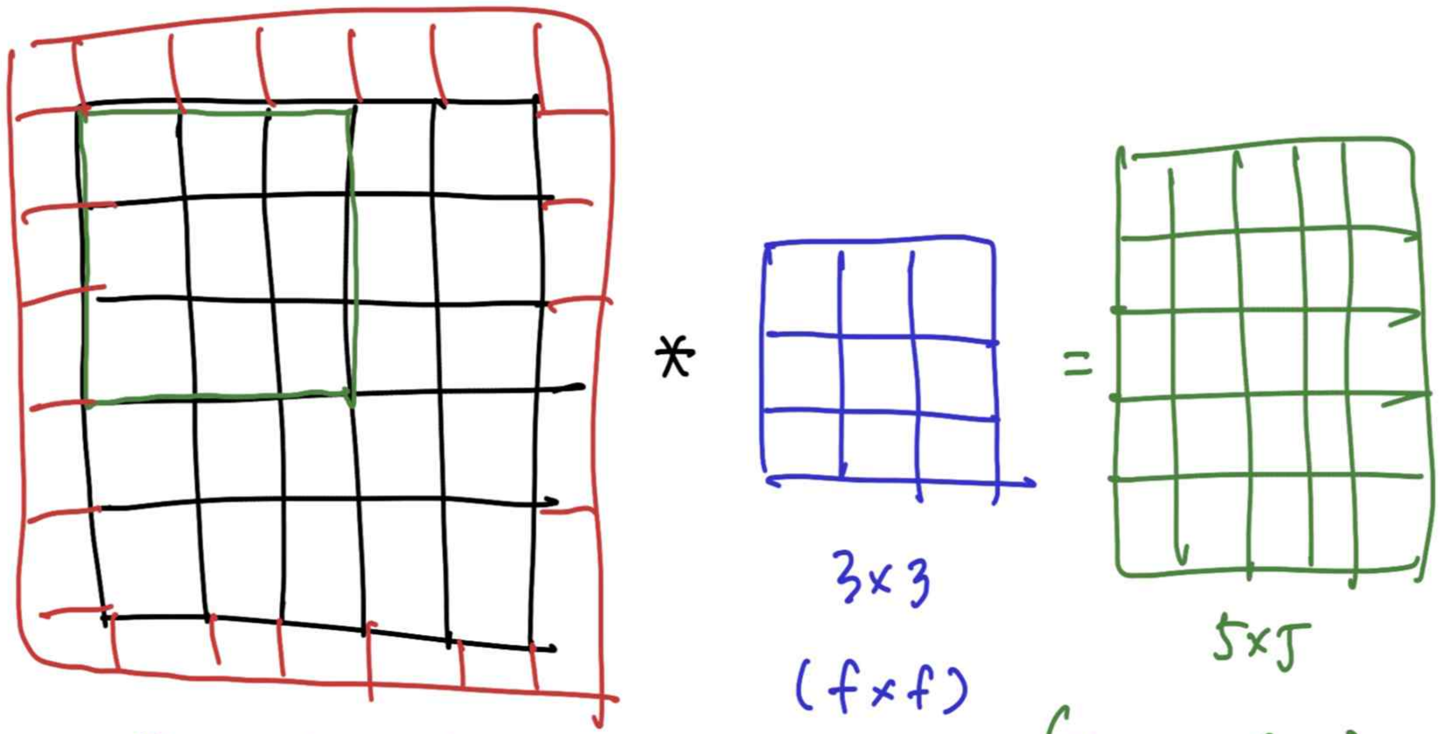
Pre-Trained CNNs

- AlexNet (5 convolutional layers + 3 fully connected layers): A. Krizhevsky, I. Sutskever, and G. E. Hinton (2012), "ImageNet classification with deep convolutional neural networks," NIPS.
- VGG (very deep CNN, 16-19 weight layers): K. Simonyan and A. Zisserman (2015), "Very deep convolutional networks for large-scale image recognition," ICLR.
- GoogLeNet (22 layers): C. Szegedy, W. Liu, Y. Jia, P. Sermanet, D. A. S. Reed, D. Erhan, V. Vanhoucke, and A. Rabinovich (2015), "Going deeper with convolutions," CVPR.
- Deep Residual Net (152 layers): K. He, X. Zhang, S. Ren, and J. Sun (2015), "Deep residual learning for Image recognition," arXiv:1512.03385.

Convolutions



Padding

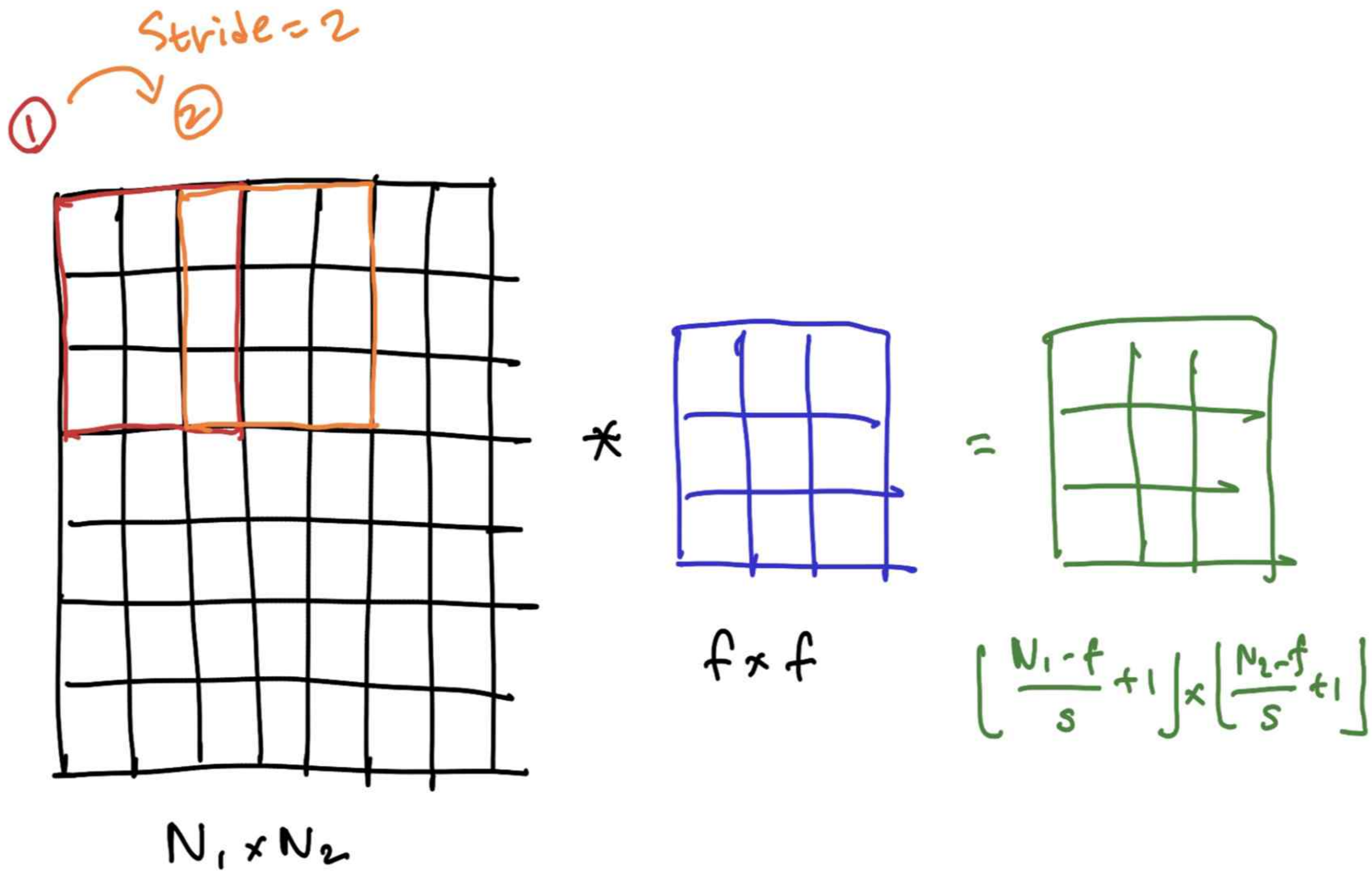


p : padding $p=1$

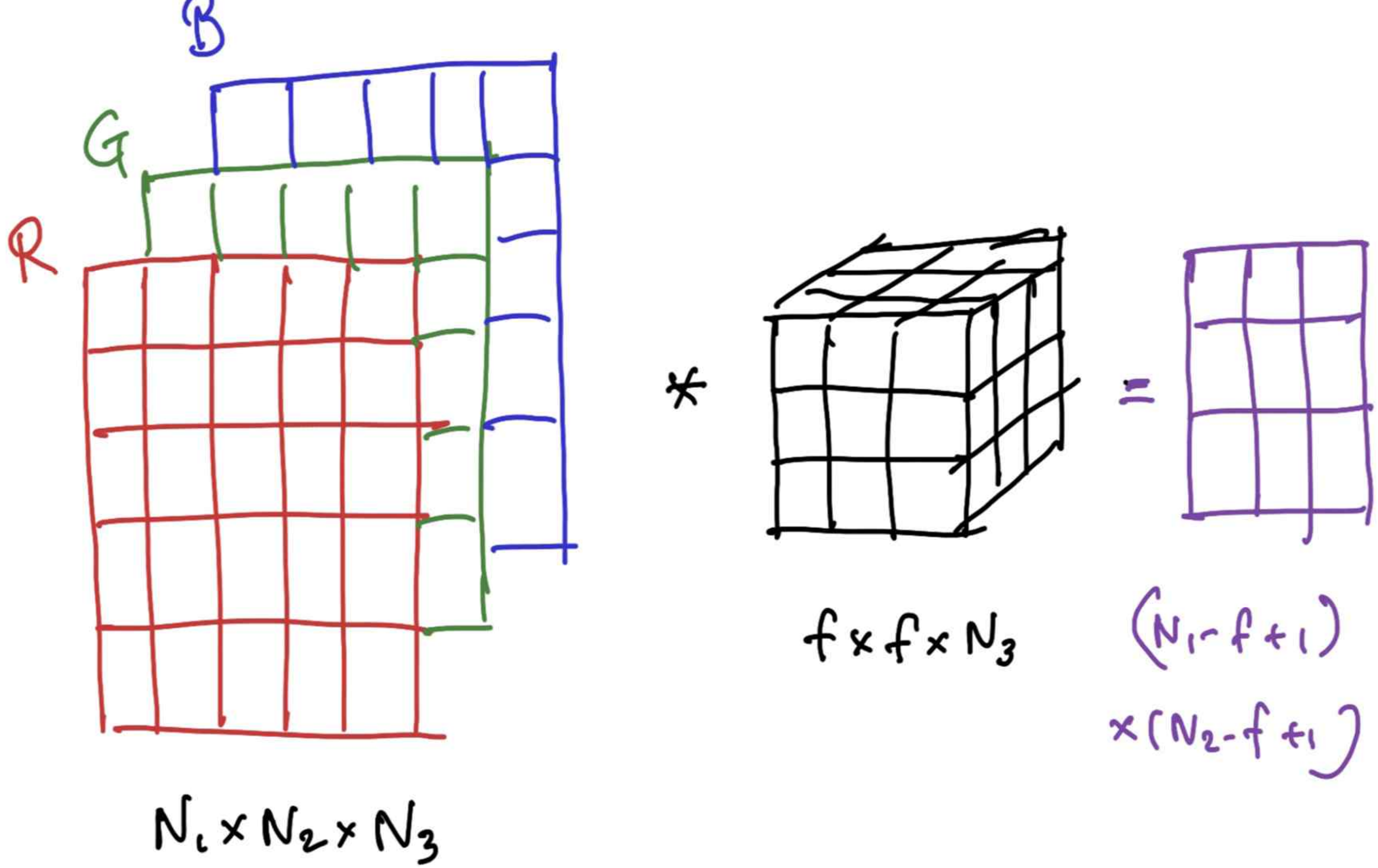
$(N_1+2p) \times (N_2+2p)$

$(N_1+2p-f+1) \times$
$(N_2+2p-f+1)$

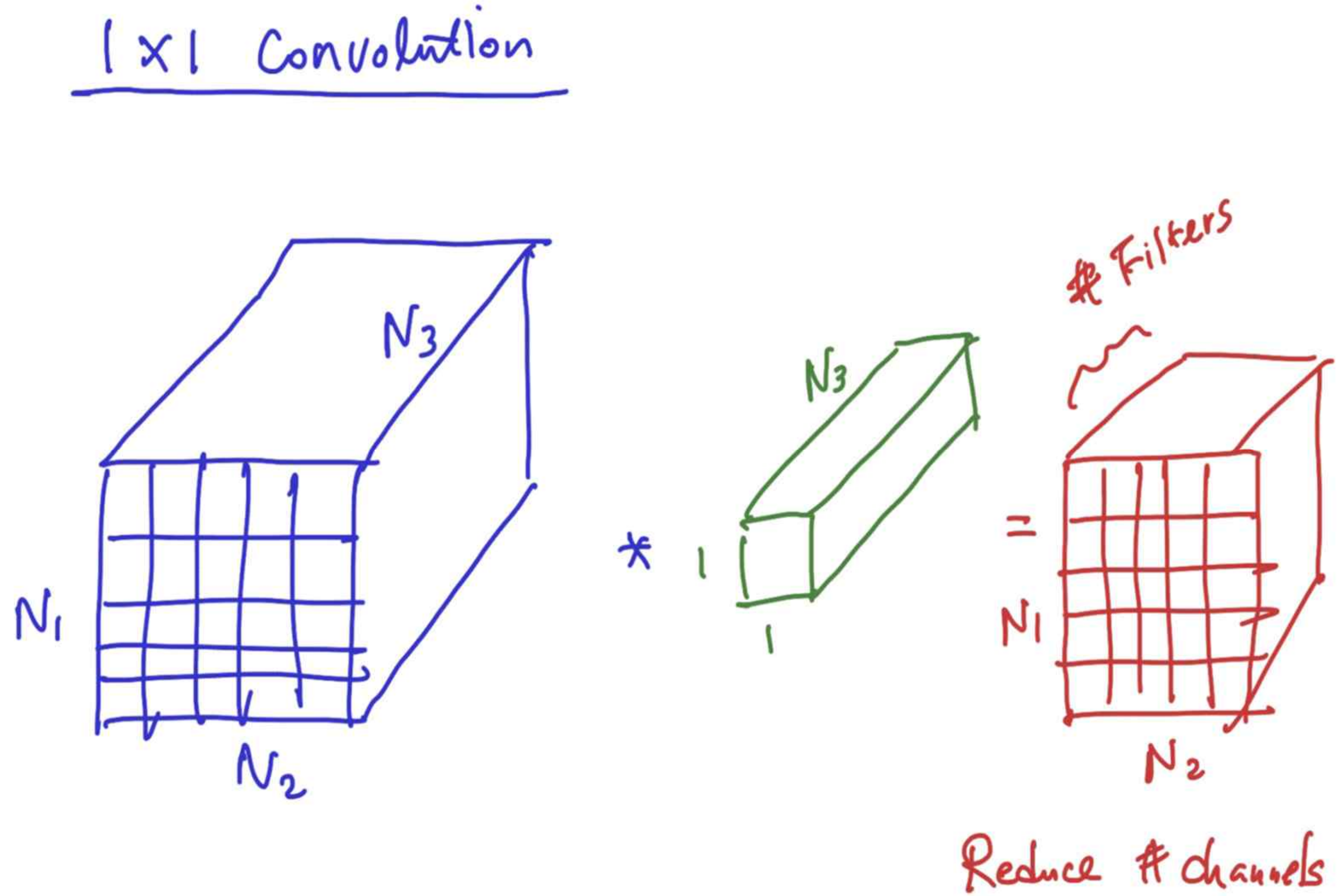
Strided Convolution



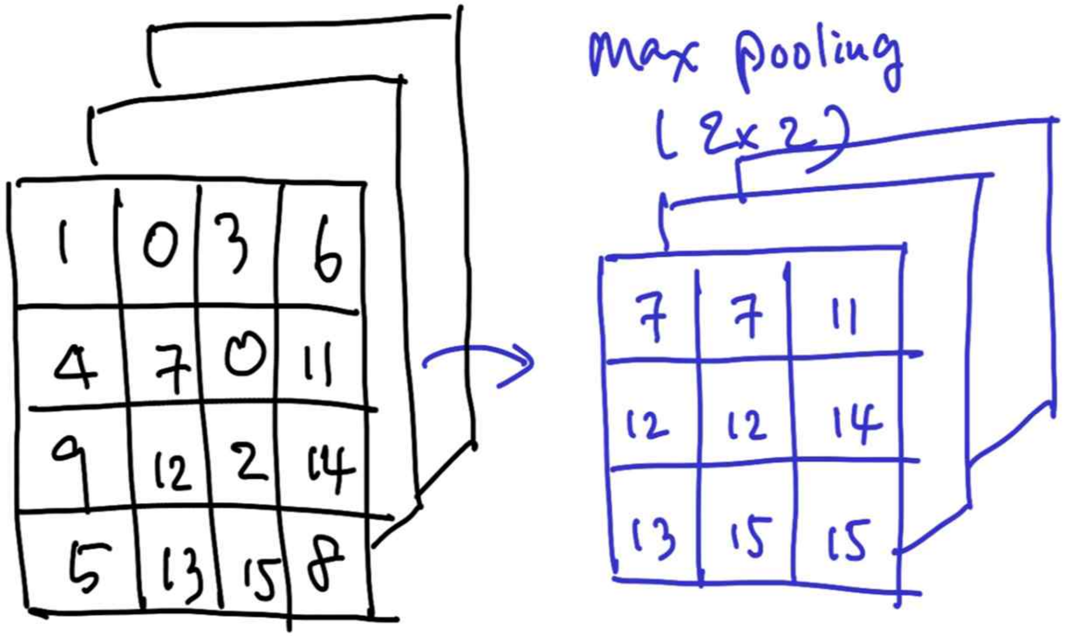
Convolution over Volume



1 x 1 Convolution

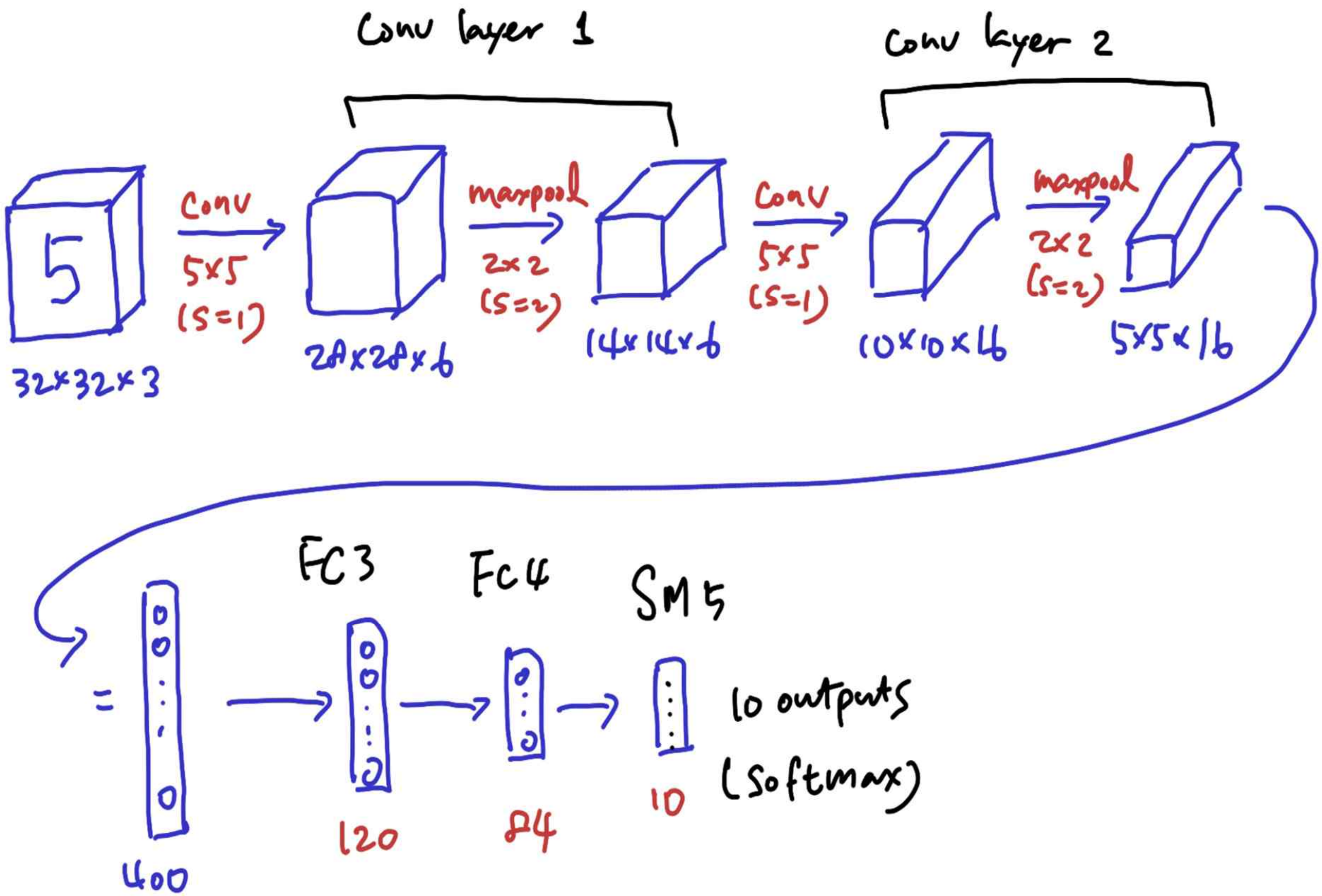


Max Pooling



Applied independently over Channels

LeNet-5

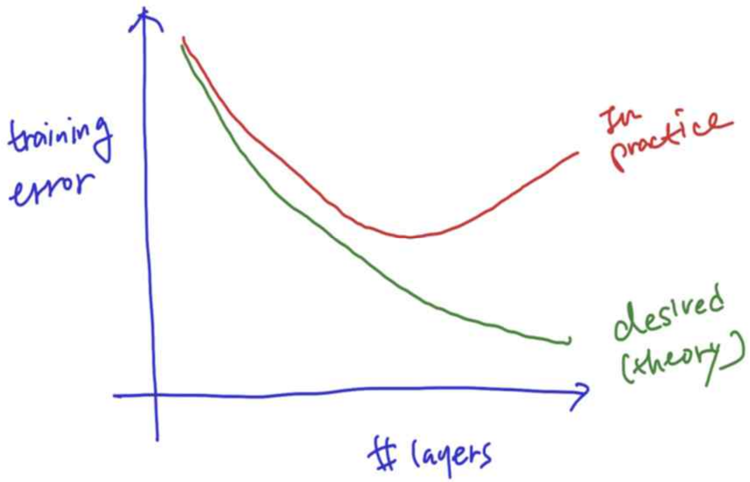


Conv layer 1
Conv layer 2
5
32×32×3
conv
5×5
(s=1)
28×28×6
maxpool
2×2
(s=2)
14×14×6
conv
5×5
(s=1)
10×10×16
maxpool
2×2
(s=2)
5×5×16
=
400
FC3
120
FC4
84
SM5
10
10 outputs
(softmax)

ResNet

K. He, X. Zhang, S. Ren, and J. Sun (2015), "Deep residual learning for Image recognition," Preprint arXiv:1512.03385.

- The deeper the better?



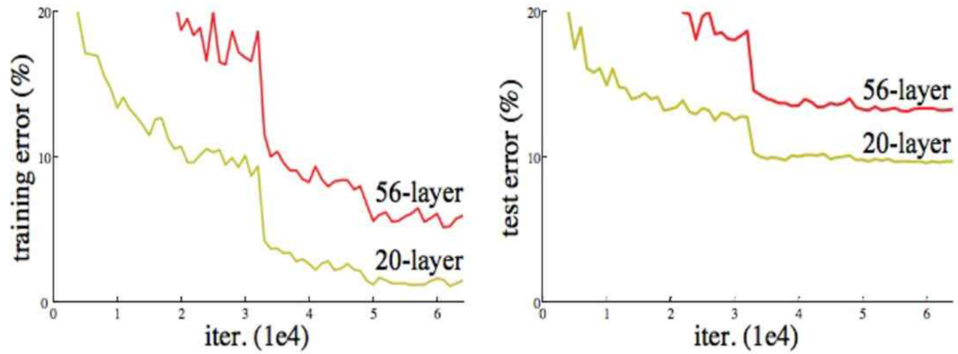
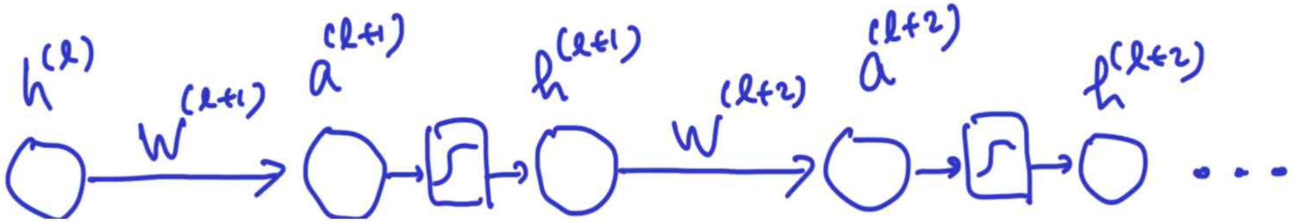


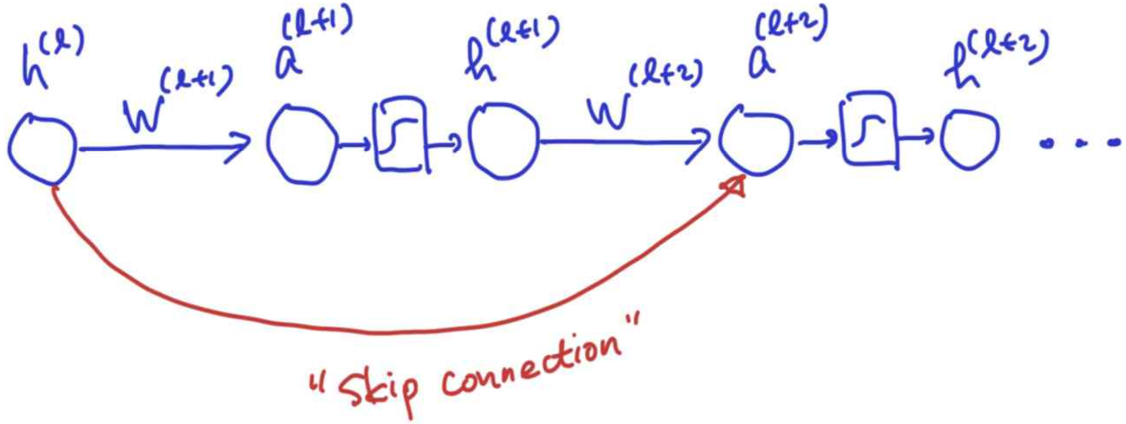
Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

ResNet

- Plain Net



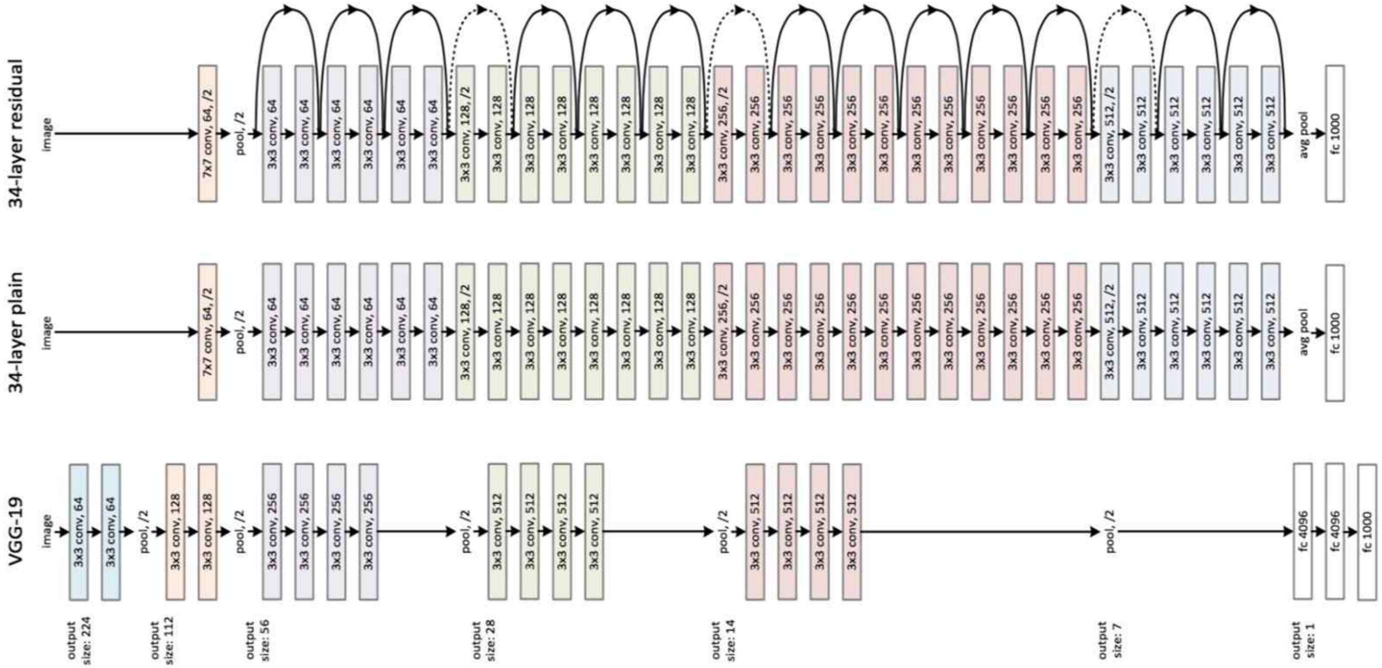
- ResNet

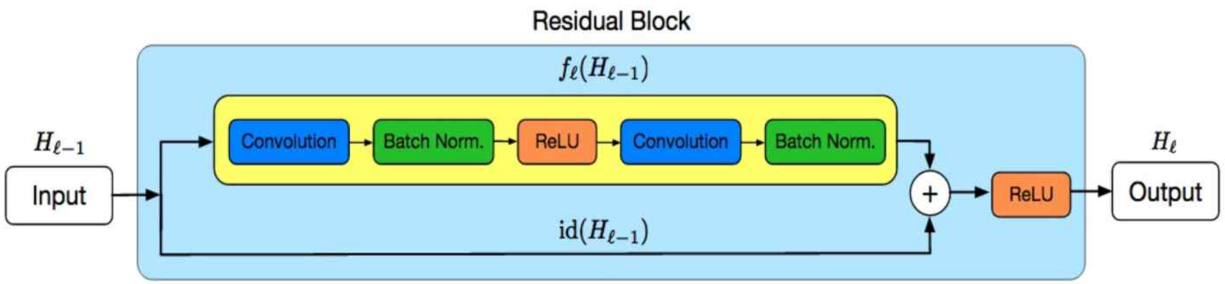


$$h^{(l+2)} = \varphi\left(W^{(l+2)} h^{(l+1)} + b^{(l+2)} + h^{(l)} \right)$$

ResNet

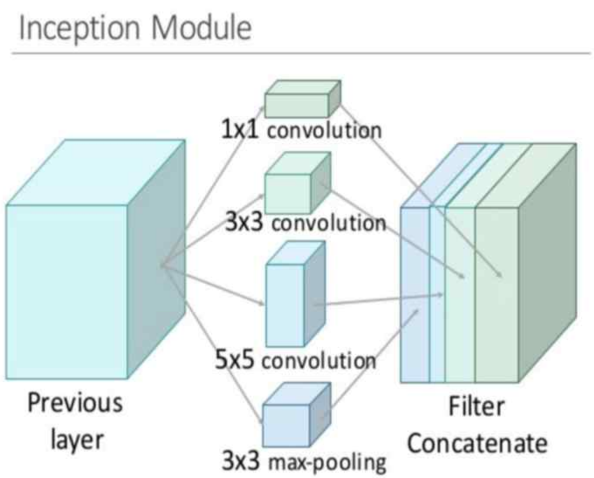
K. He, X. Zhang, S. Ren, and J. Sun (2015), "Deep residual learning for Image recognition," Preprint arXiv:1512.03385.

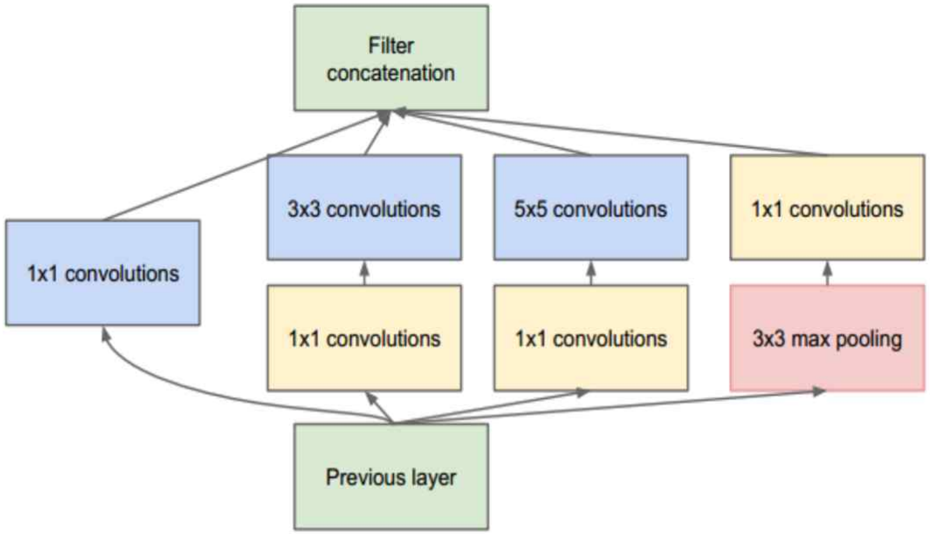




Inception Net

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, D. A. S. Reed, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In CVPR, 2015.





Inception Net