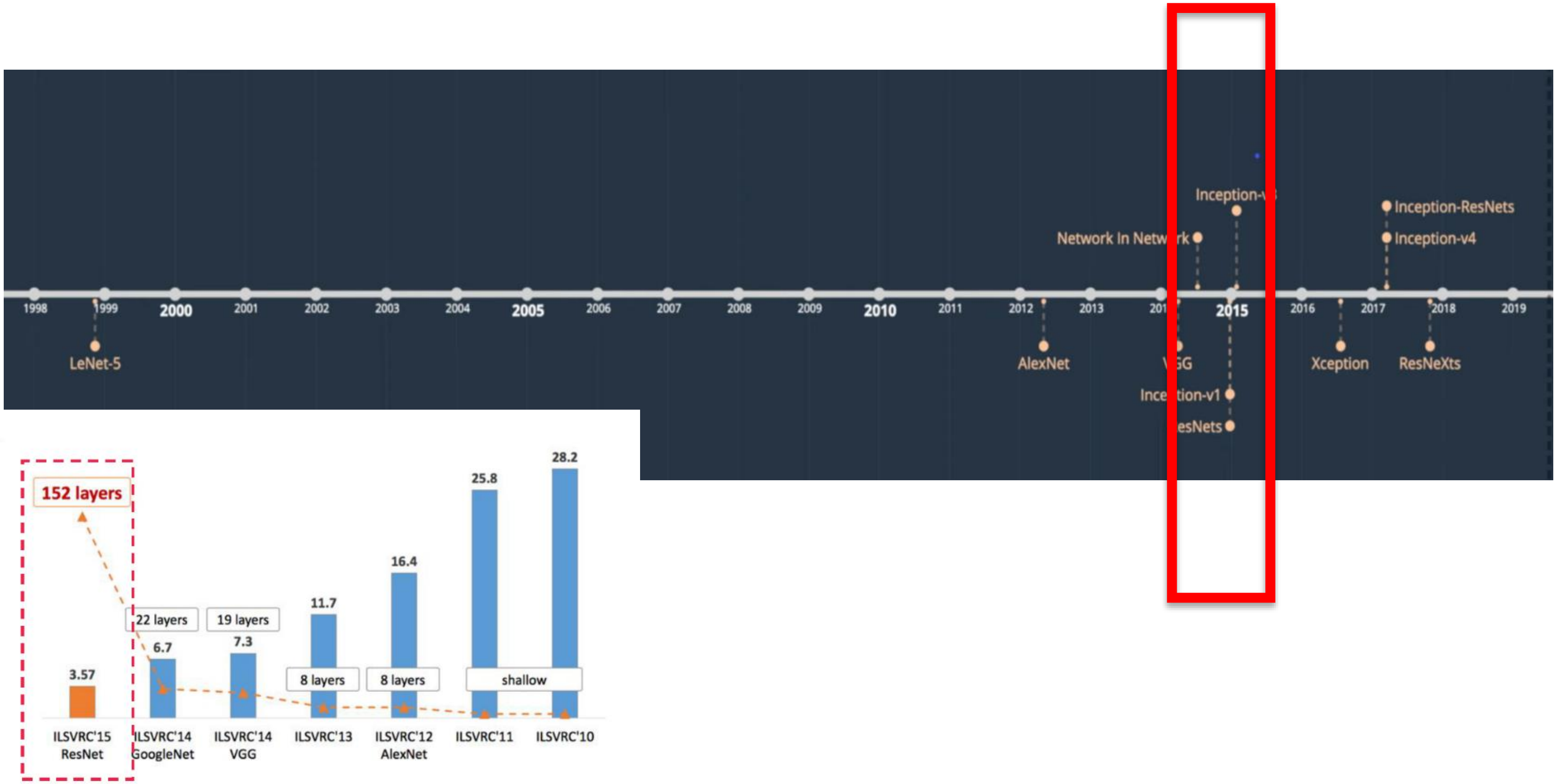
ResNet

Computer Vision & Augmented Reality 연구실
학부연구생 강 준 구

ResNet(2015)

- ILSVRC

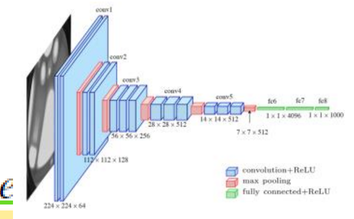


Contents

- 1. Overview paper
- 2. Experiments
- 3. ResNet Architecture
- 4. Future work

ResNet

- Introduction

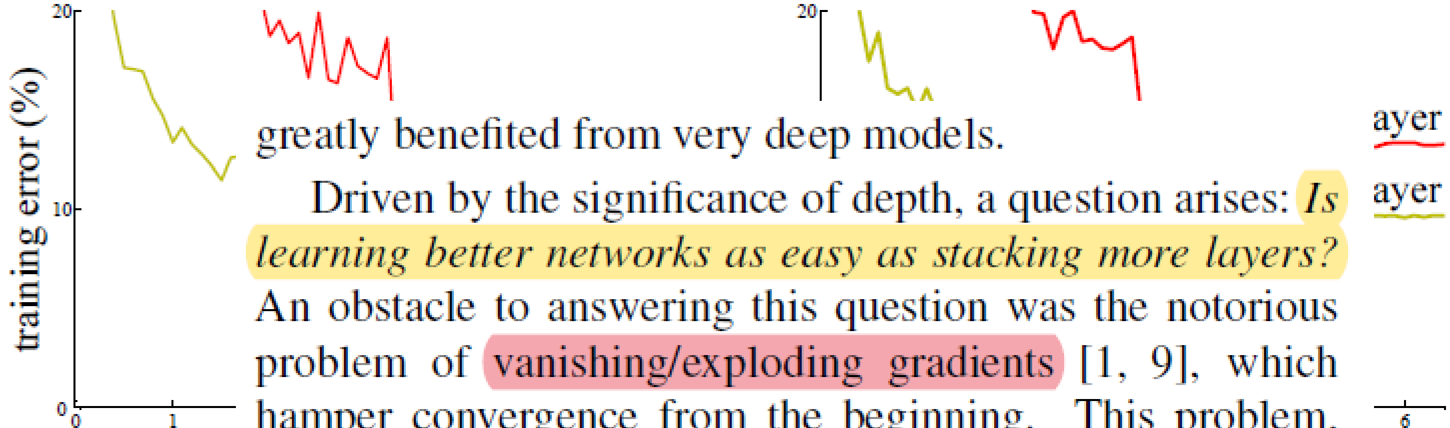


Deeper neural ne *ifficult to train.* We present a residual learning framework to ease the training considerably increased depth. On the ImageNet dataset we evaluate residual nets with a depth of up to 152 layers—8× deeper than VGG nets [41] but still having lower complexity. An ensemble of these residual nets achieves 3.57% error on the ImageNet test set. This result won the 1st place on the ILSVRC 2015 classification task. We also present analysis on CIFAR-10 with 100 and 1000 layers.

considerably increased depth. On the ImageNet dataset we
evaluate residual nets with a depth of up to 152 layers—8×

ResNet

- Introduction



greatly benefited from very deep models.

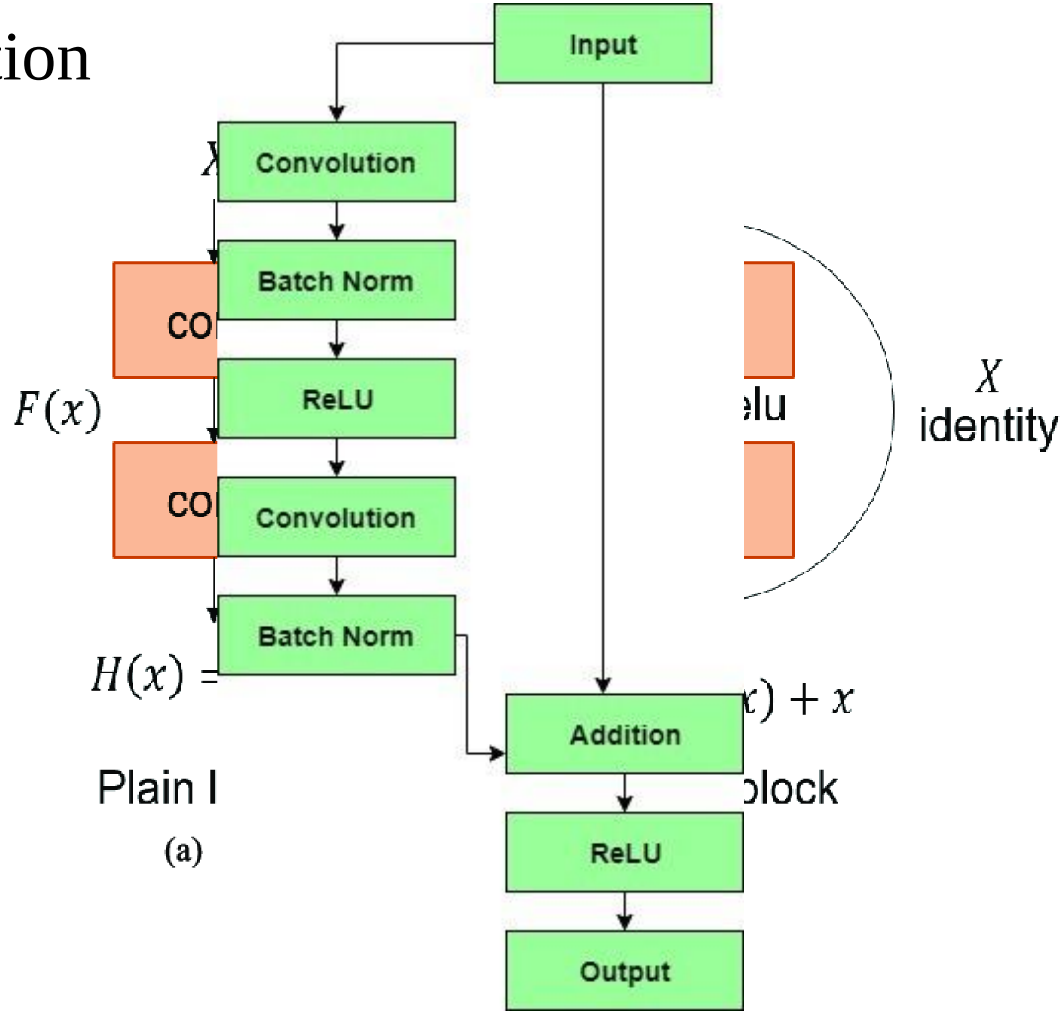
Driven by the significance of depth, a question arises: *Is learning better networks as easy as stacking more layers?* An obstacle to answering this question was the notorious problem of vanishing/exploding gradients [1, 9], which hamper convergence from the beginning. This problem, however, has been largely addressed by normalized initialization [23, 9, 37, 13] and intermediate normalization layers [16], which enable networks with tens of layers to start converging for stochastic gradient descent (SGD) with backpropagation [22].

Figure 1. ... -10 with 20-la... ork has higher ... ena on ImageN...

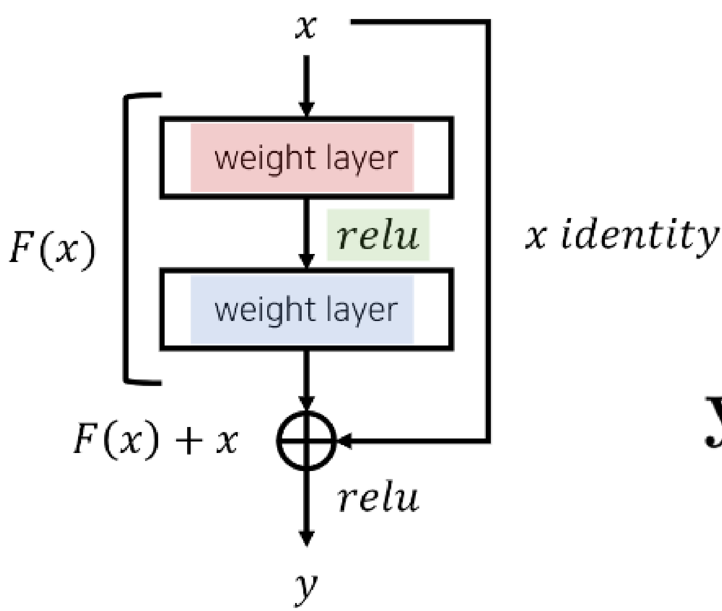
degradation

ResNet

- Introduction



ResNet



$$\mathcal{F} = W_2\sigma(W_1\mathbf{x})$$

$$\mathbf{y} = \underbrace{\mathcal{F}(\mathbf{x}, \{W_i\})}_{\text{multiple convolutional layers}} + \underbrace{W_s\mathbf{x}}_{\text{shortcut}}$$

ResNet

- Experiments

101-layer and 152-layer ResNets: We construct 101-layer and 152-layer ResNets by using more 3-layer blocks (Table 1). Remarkably, although the depth is significantly increased, the 152-layer ResNet (11.3 billion FLOPs) still has *lower complexity* than VGG-16/19 nets (15.3/19.6 billion FLOPs).

ResNet

- Experiments

Comparisons with State-of-the-art Methods. In Table 4 we compare with the previous best single-model results. Our baseline 34-layer ResNets have achieved very competitive accuracy. Our 152-layer ResNet has a single-model top-5 validation error of 4.49%. This single-model result outperforms all previous ensemble results (Table 5). We combine six models of different depth to form an ensemble (only with two 152-layer ones at the time of submitting). This leads to **3.57%** top-5 error on the test set (Table 5). *This entry won the 1st place in ILSVRC 2015.*

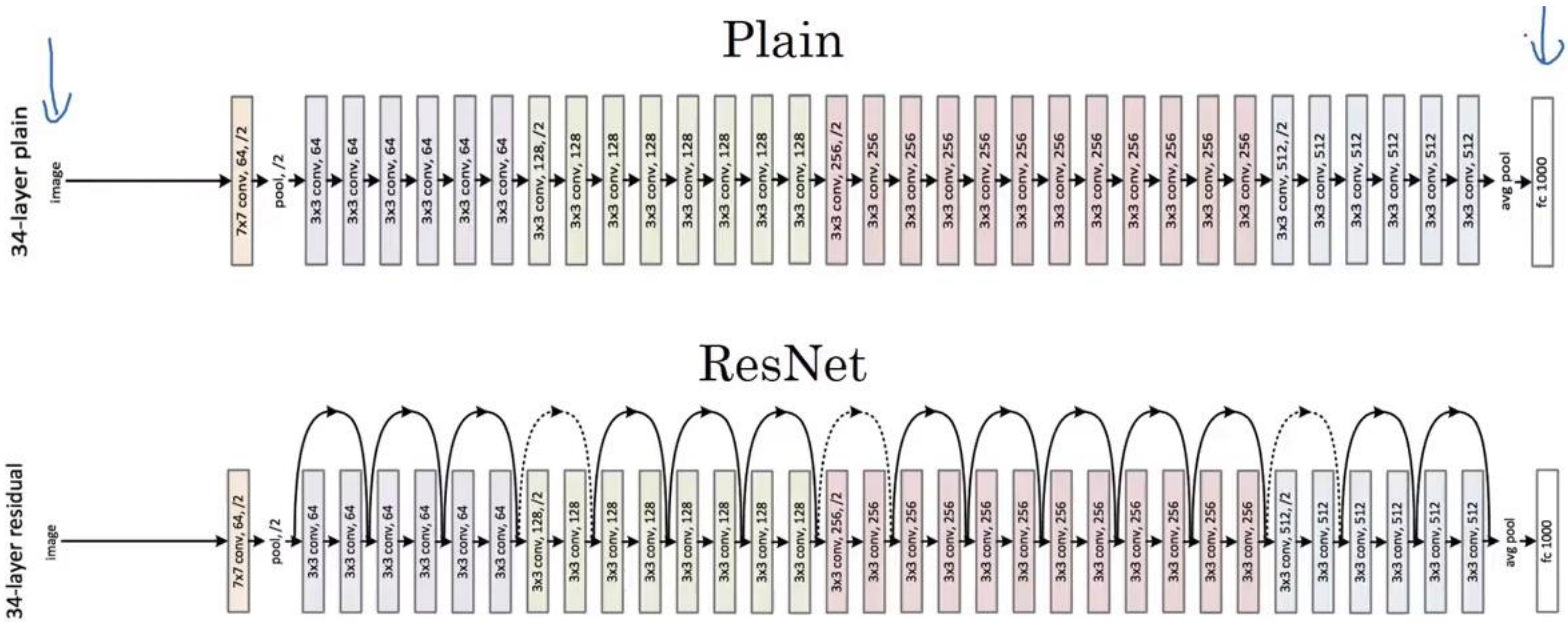
ResNet

- Experiments

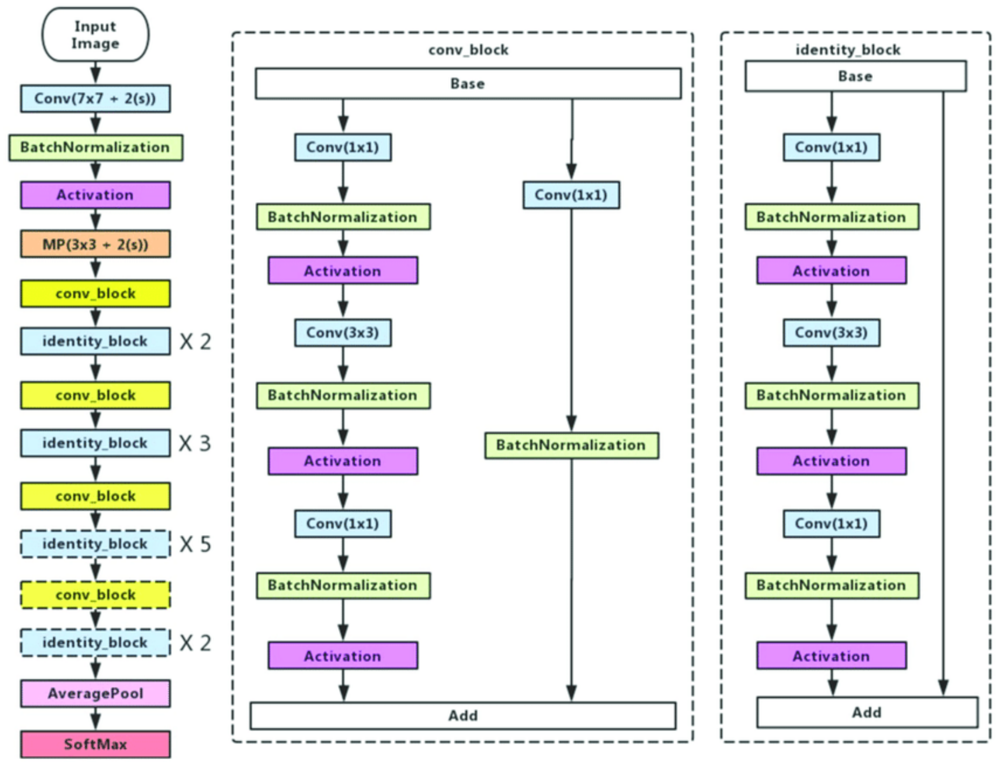
method			error (%)
Maxout [10]			9.38
NIN [25]			8.81
DSN [24]			8.22
	# layers	# params	
FitNet [35]	19	2.5M	8.39
Highway [42, 43]	19	2.3M	7.54 (7.72±0.16)
Highway [42, 43]	32	1.25M	8.80
ResNet	20	0.27M	8.75
ResNet	32	0.46M	7.51
ResNet	44	0.66M	7.17
ResNet	56	0.85M	6.97
ResNet	110	1.7M	**6.43** (6.61±0.16)
ResNet	1202	19.4M	7.93

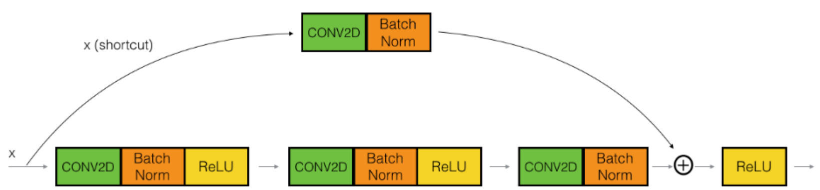
ResNet

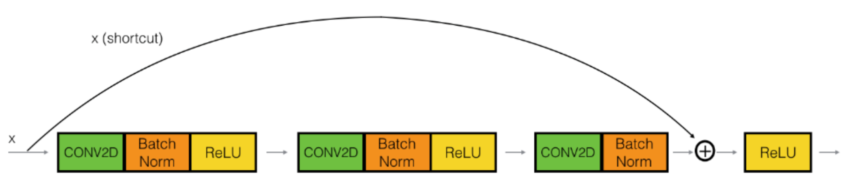
- ResNet Architecture



ResNet


Input Image
Conv(7x7 + 2(s))
BatchNormalization
Activation
MP(3x3 + 2(s))
conv_block
identity_block X 2
conv_block
identity_block X 3
conv_block
identity_block X 5
conv_block
identity_block X 2
AveragePool
SoftMax
conv_block
Base
Conv(1x1)
BatchNormalization
Activation
Conv(3x3)
BatchNormalization
Activation
Conv(1x1)
BatchNormalization
Activation
Conv(1x1)
BatchNormalization
Add
identity_block
Base
Conv(1x1)
BatchNormalization
Activation
Conv(3x3)
BatchNormalization
Activation
Conv(1x1)
BatchNormalization
Activation
Add


x (shortcut)
CONV2D
Batch Norm
x
CONV2D
Batch Norm
ReLU
CONV2D
Batch Norm
ReLU
CONV2D
Batch Norm
ReLU


x (shortcut)
x
CONV2D
Batch Norm
ReLU
CONV2D
Batch Norm
ReLU
CONV2D
Batch Norm
ReLU

ResNet

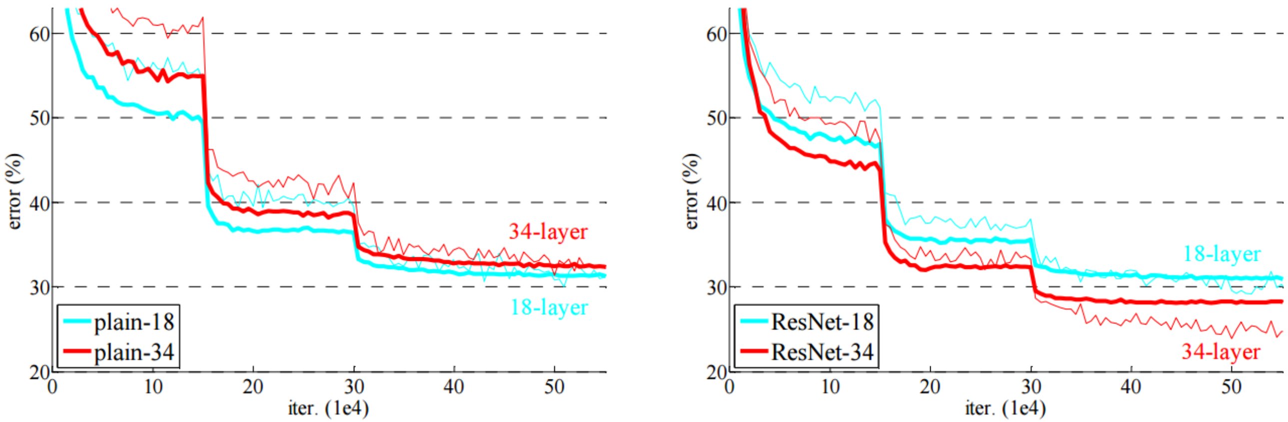


Figure 4. Training on **ImageNet**. Thin curves denote training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameter compared to their plain counterparts.