Paper Review

Deep Residual Learning for Image Recognition (2016 CVPR)

JooYoung Song

Department of Electronics and Computer Engineering
Hongik University

Contents

Abstract

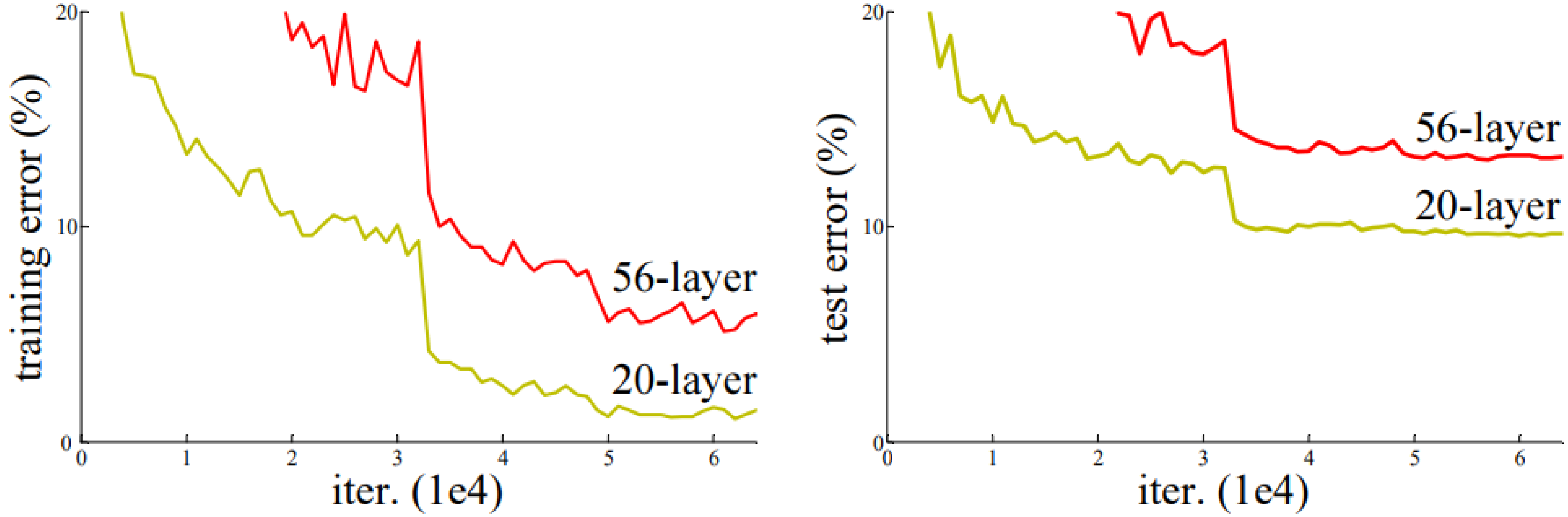


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

ResNet @ ILSVRC & COCO 2015 Competitions

1st places in all five main tracks

- ImageNet Classification : “Ultra-deep” 152-layer nets
- ImageNet Detection : 16% better than 2nd
- ImageNet Localization : 27% better than 2nd
- COCO Detection : 11% better than 2^{nd}
- COCO Segmentation : 12% better than 2nd

Is learning better networks
as simple as stacking more layers?

Is learning better networks as simple as stacking more layers?

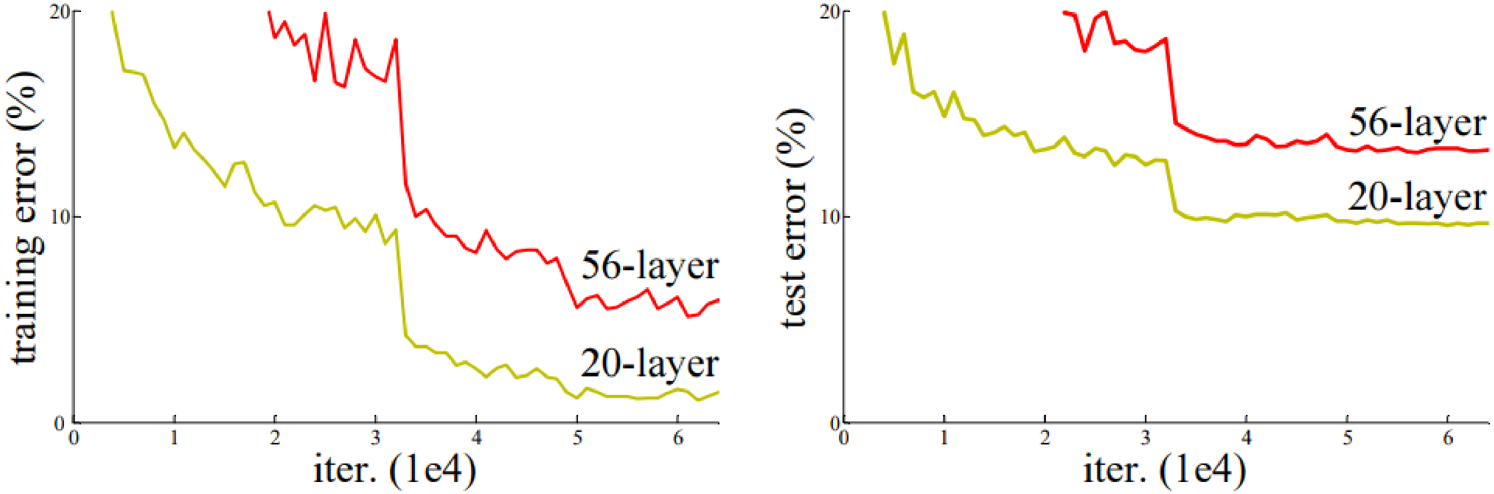
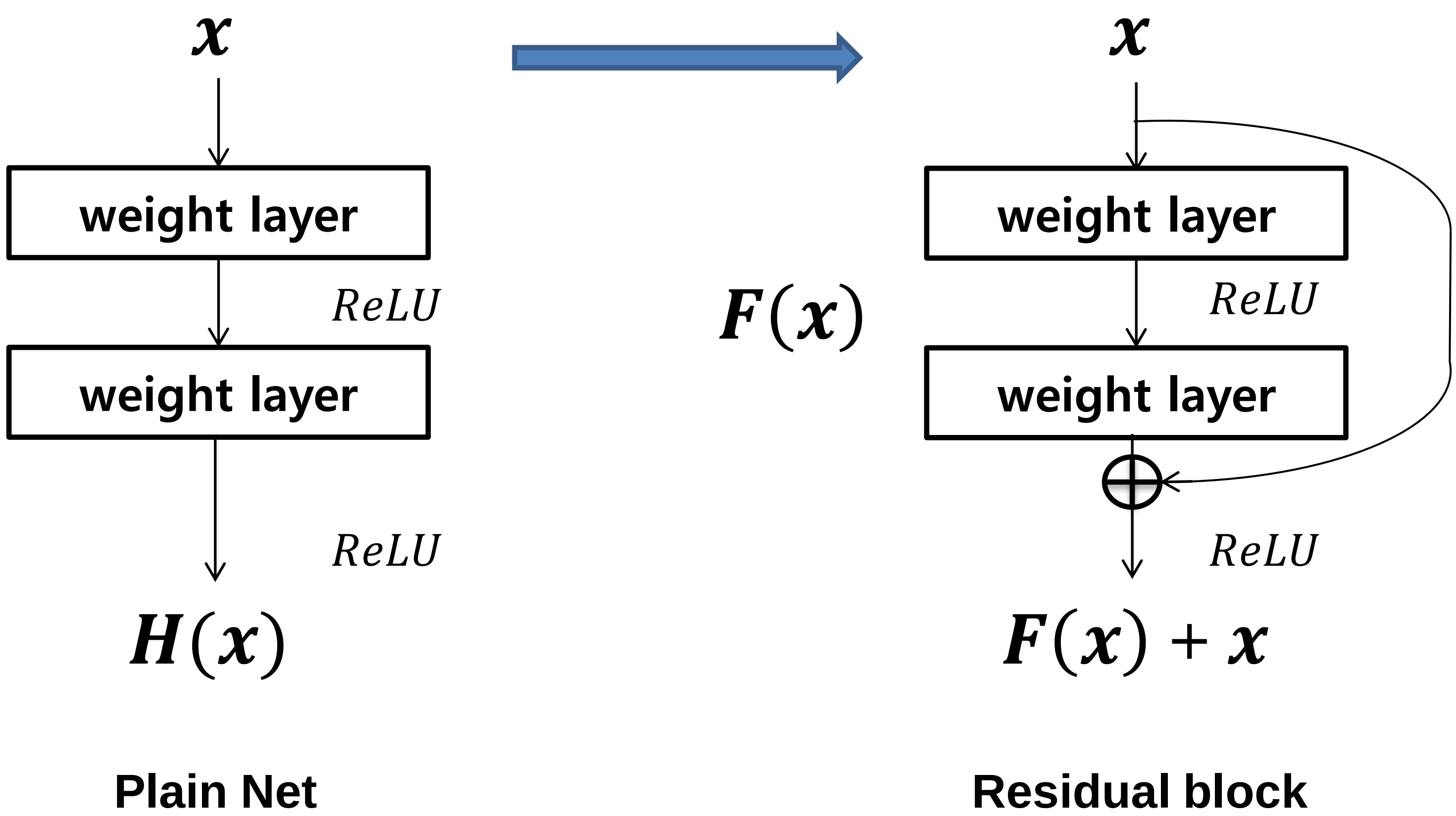


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

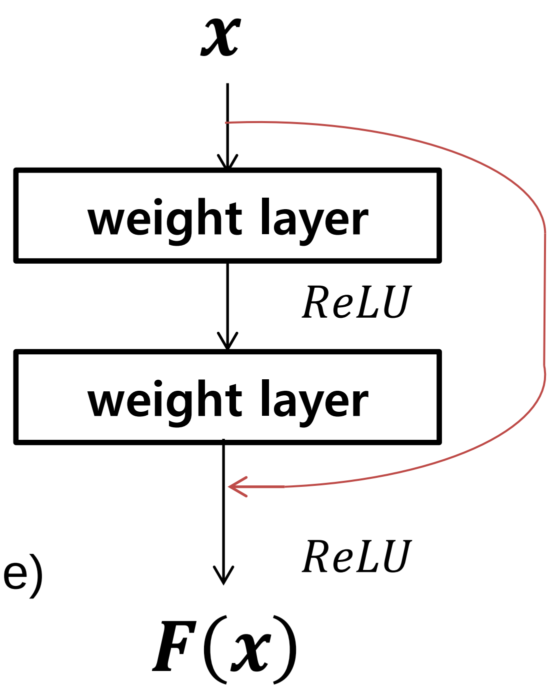
Introduction

$$F(x) := H(x) - x$$



Related Work

Shortcut Connection (Skip Connection)



GoogleNet vs ResNet

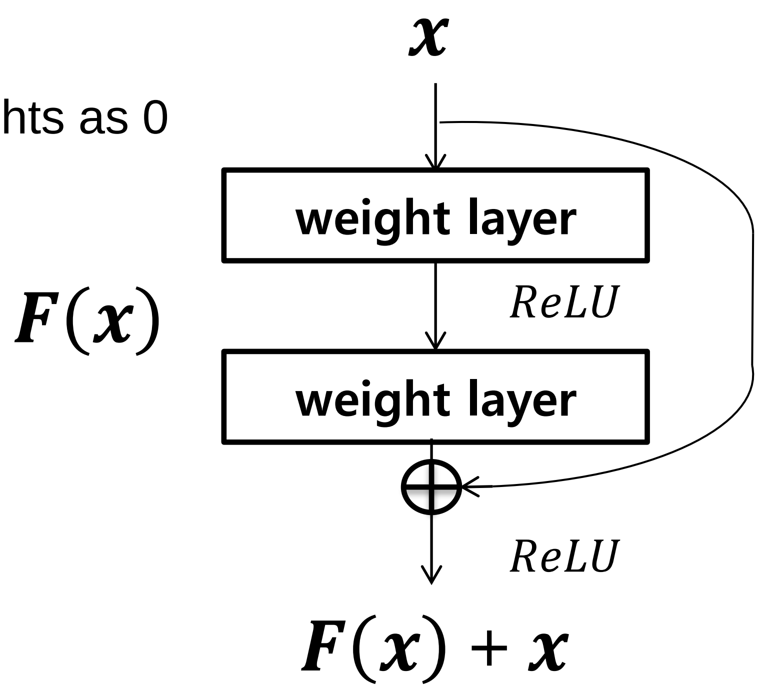
- Low parameter
- Not required Auxillary Classifier separately.

Highway Networks vs ResNet

- Low parameter (identity shortcut – parameter free)
- Not demonstrated accuracy gains with extremely increased depth (highway)

Residual Learning

- If identity were optimal, easy to set weights as 0
- If optimal mapping is closer to identify, easier to find small functions



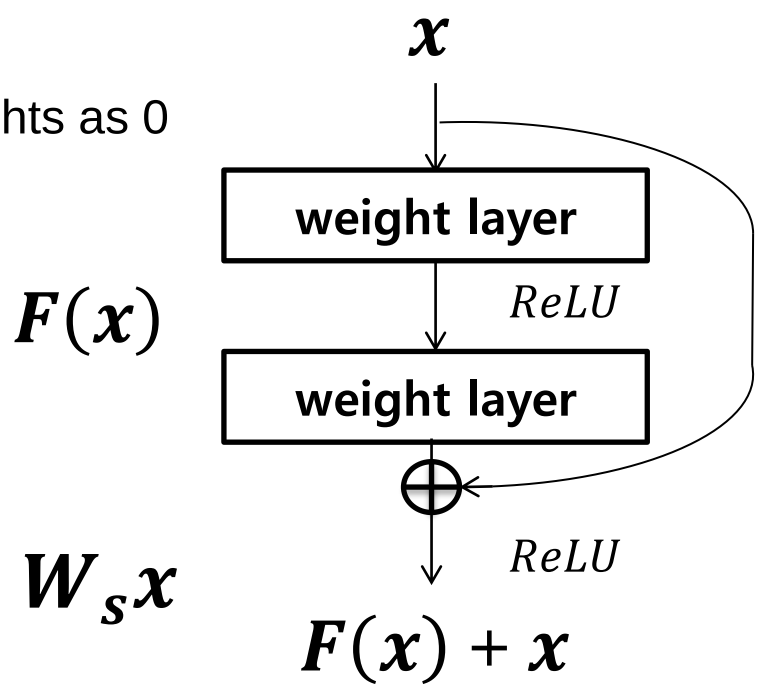
Residual block

Residual Learning

- If identity were optimal, easy to set weights as 0

- If optimal mapping is closer to identify, easier to find small functions

$$y = F(x, W_i) + W_s x$$



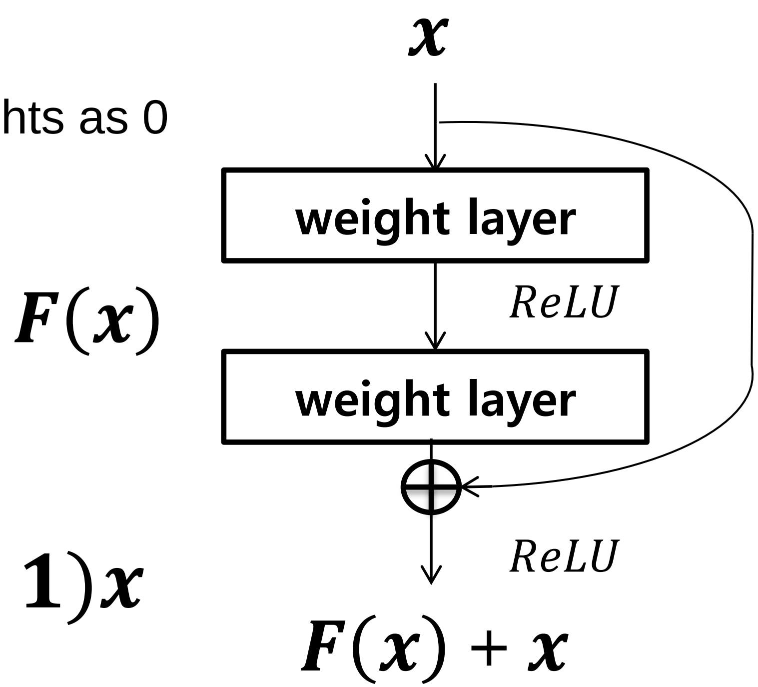
Residual block

Deep Residual Learning

Residual Learning

- If identity were optimal, easy to set weights as 0

- If optimal mapping is closer to identify, easier to find small functions



$$\boldsymbol{y = W_1 x + x = (W_1 + 1)x}$$

Residual block

Deep Residual Learning

Network Architectures

- **Keep it simple**
- **Basic design (VGG-style)**
  - All 3×3 conv (almost)
  - Spatial size / 2 => # filters × 2
  - Simple design; just deep!

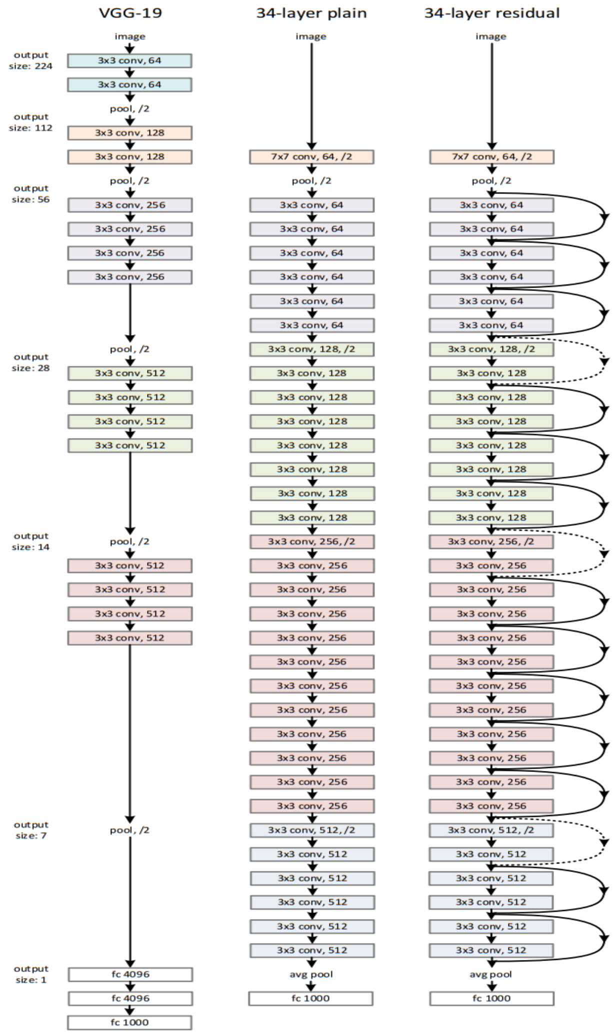
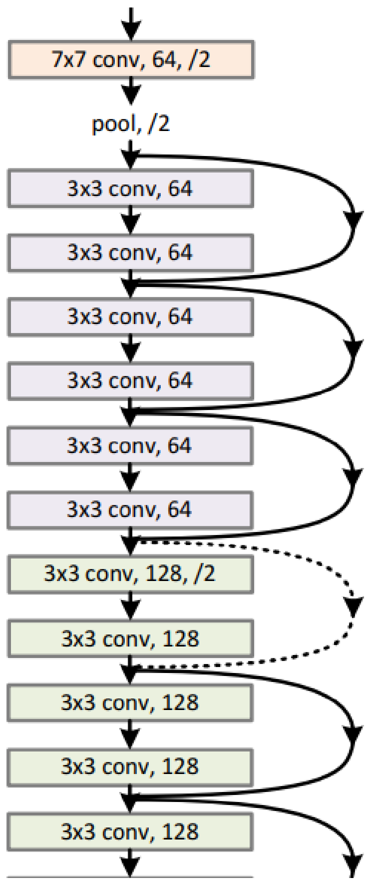


Figure 3. Example network architectures for ImageNet. **Left**: the VGG-19 model [41] (19.6 billion FLOPs) as a reference. **Middle**: a plain network with 34 parameter layers (3.6 billion FLOPs). **Right**: a residual network with 34 parameter layers (3.6 billion FLOPs). The dotted shortcuts increase dimensions. **Table 1** shows more details and other variants.

Deep Residual Learning

Network Architectures



Just added

Additional techniques

- **Zero padding → identity mapping**
- **Projection Shortcut**

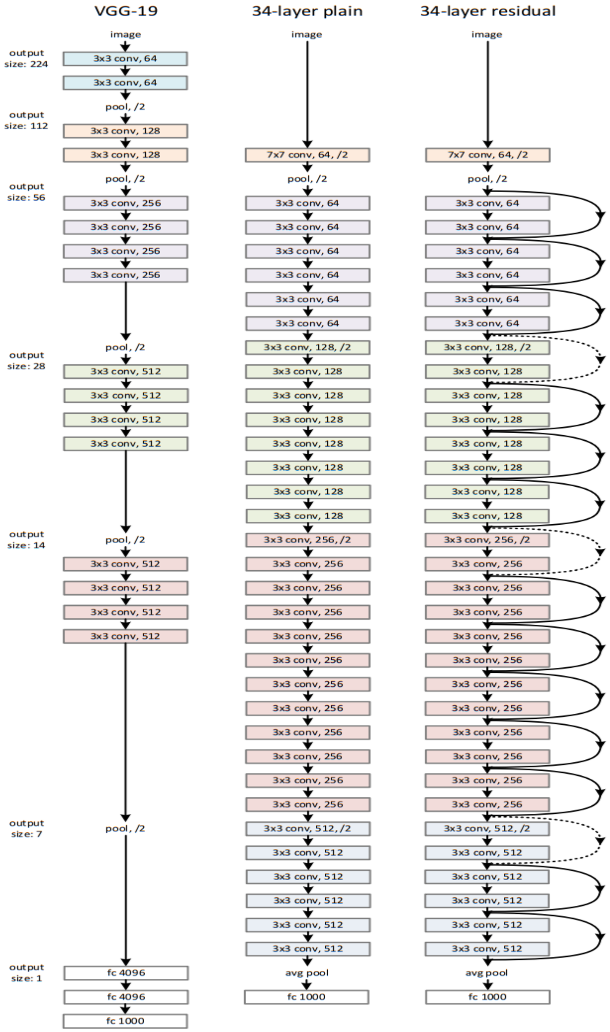
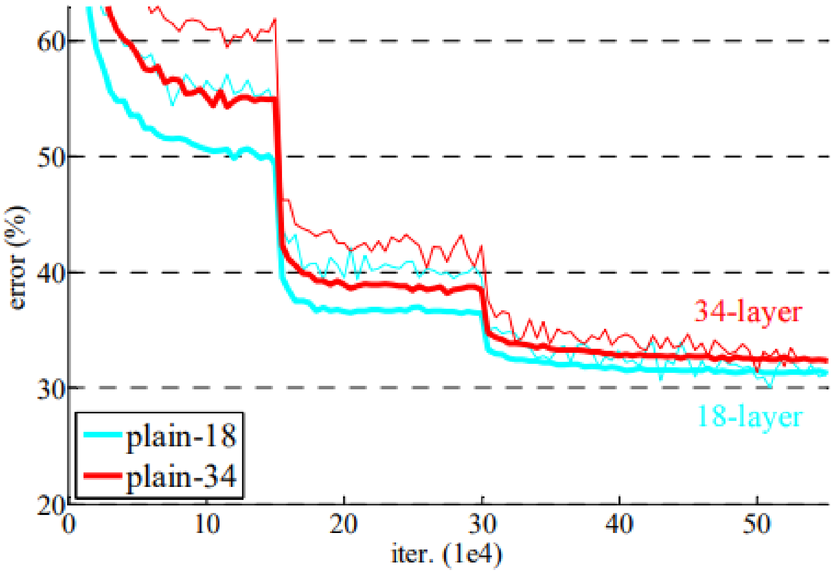
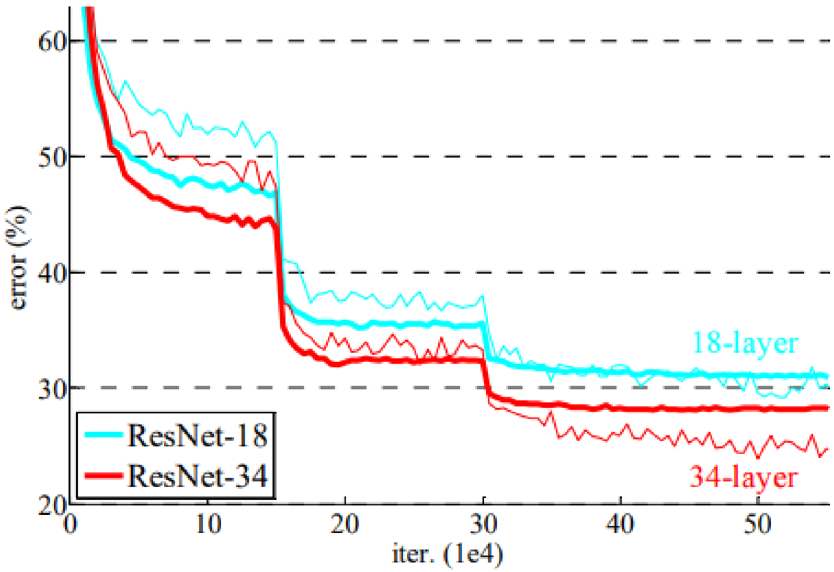


Figure 3. Example network architectures for ImageNet. **Left**: the VGG-19 model [41] (19.6 billion FLOPs) as a reference. **Middle**: a plain network with 34 parameter layers (3.6 billion FLOPs). **Right**: a residual network with 34 parameter layers (3.6 billion FLOPs). The dotted shortcuts increase dimensions. **Table 1** shows more details and other variants.

Experiments

ImageNet experiments



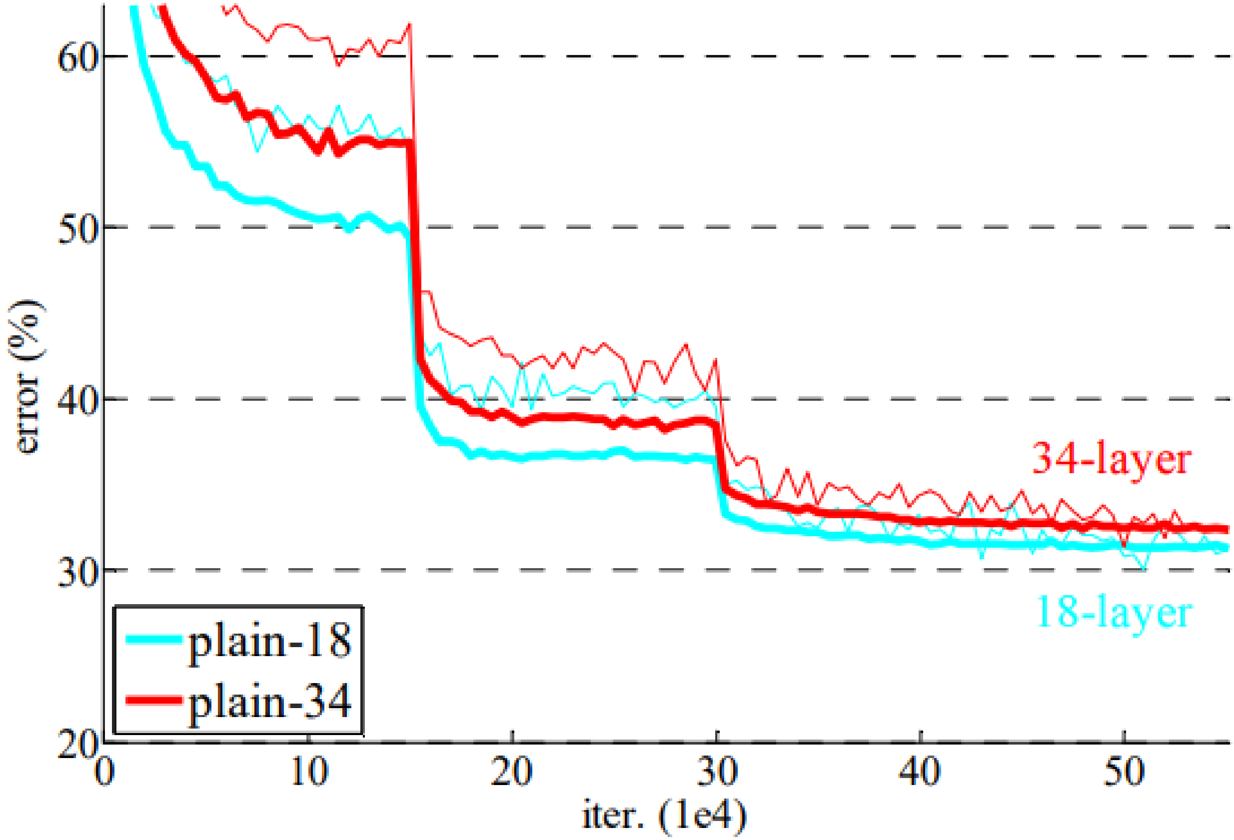


Plain networks of 18 and 34 layers.

ResNets of 18 and 34 layers.

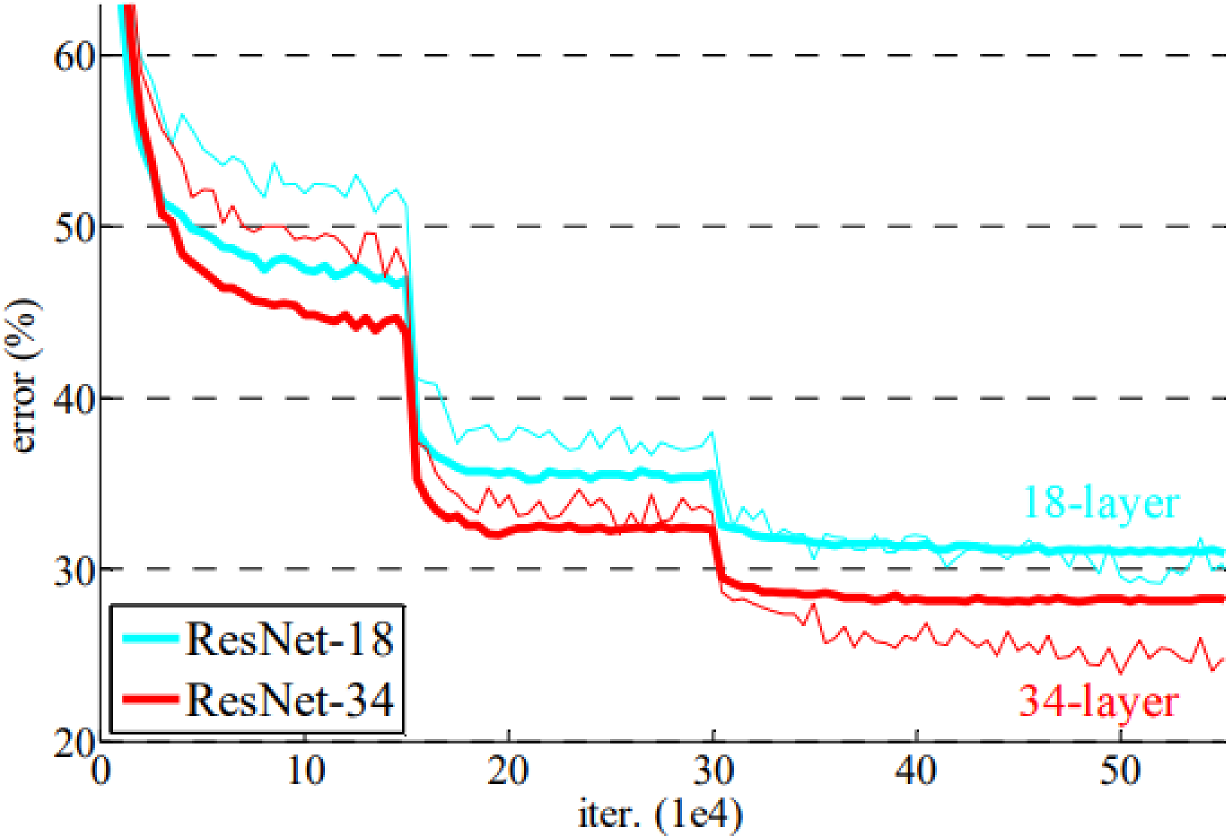
Experiments

ImageNet experiments



Plain networks of 18 and 34 layers.

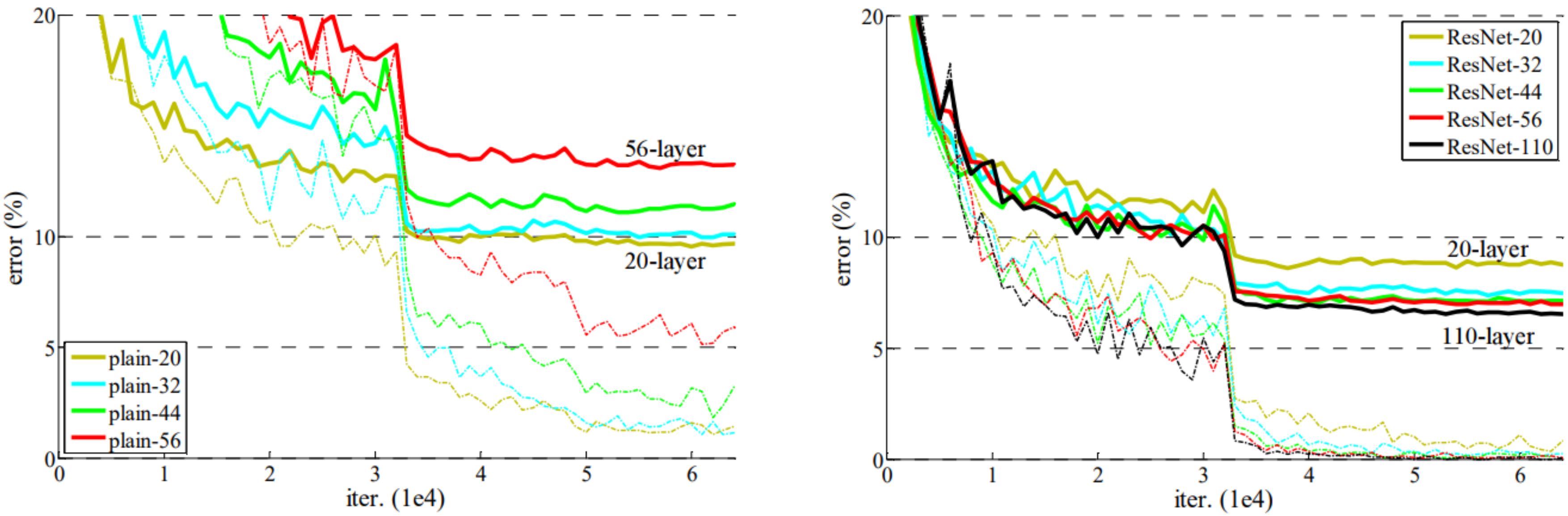
ImageNet experiments



ResNets of 18 and 34 layers.

Experiments

CIFAR-10 experiments



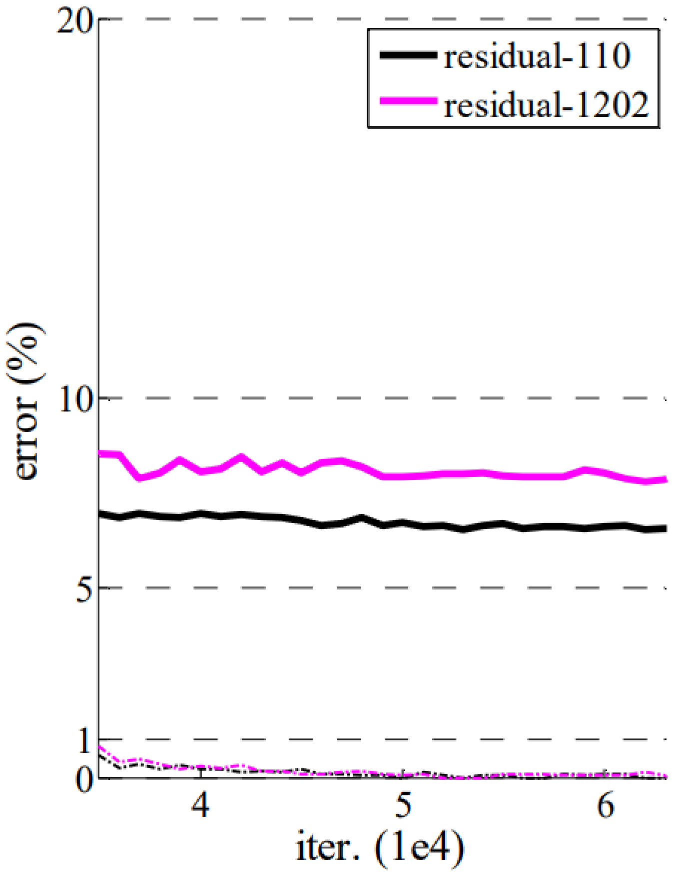
- Deep ResNets can be trained without difficulties
- Deeper ResNets have lower training error, and also lower test error

Experiments

Does the performance continue to increase
if you continue to increase the layers of the model?

Experiments

Does the performance continue to increase
if you continue to increase the layers of the model?



No → Because of **overfitting.**

But it could be studied by adding regularization**.**

Conclusions

- Deep Residual Networks:
  - Easy to train
  - Simply gain accuracy from depth
  - Well transferable