Deep Residual Learning for Image Recognition

Author : Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun

(accepted to CVPR 2016)

Presenter : Hyeongseok Son

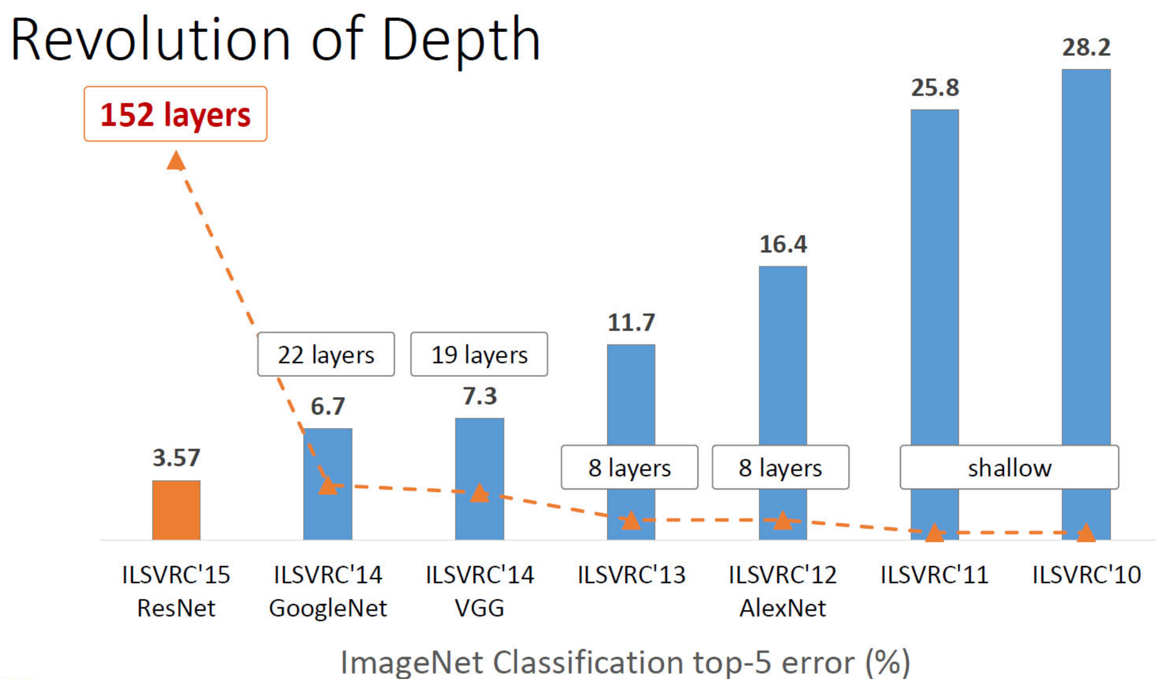
The deeper, the better

- The deeper network can cover more complex problems
  - Receptive field size ↑
  - Non-linearity ↑
- However, training the deeper network is more difficult because of vanishing/exploding gradients problem

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Deep Neural Network

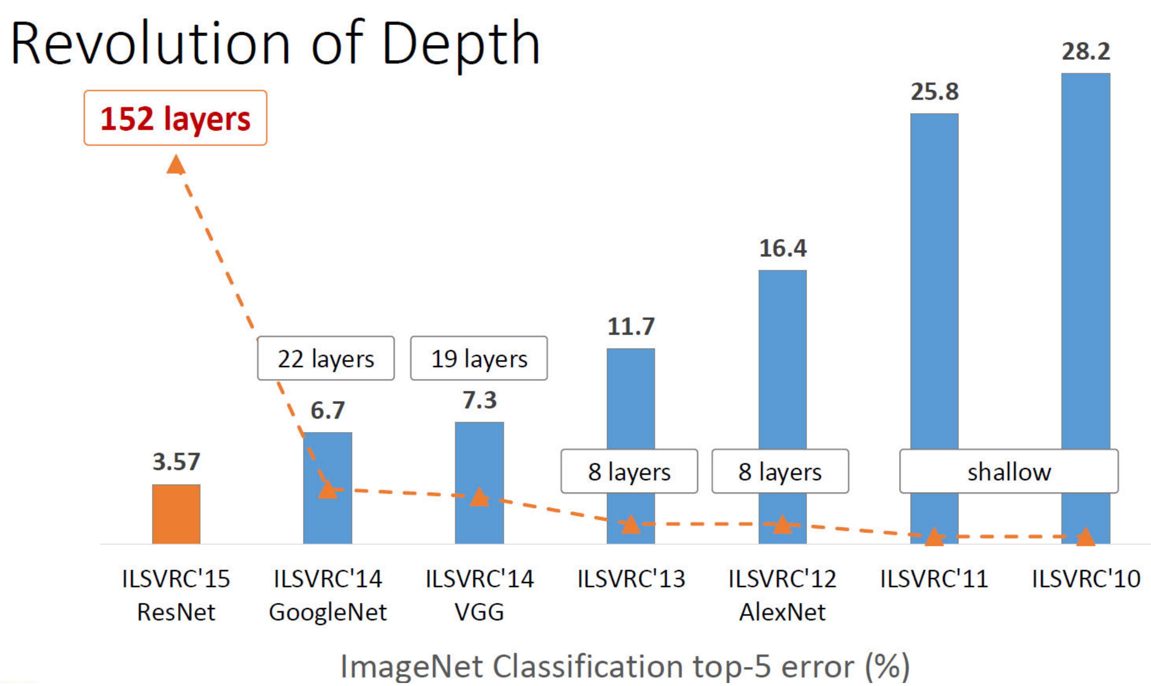
- Escape from few layers
  - ReLU for solving gradient vanishing problem
  - Dropout …



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Deep Neural Network

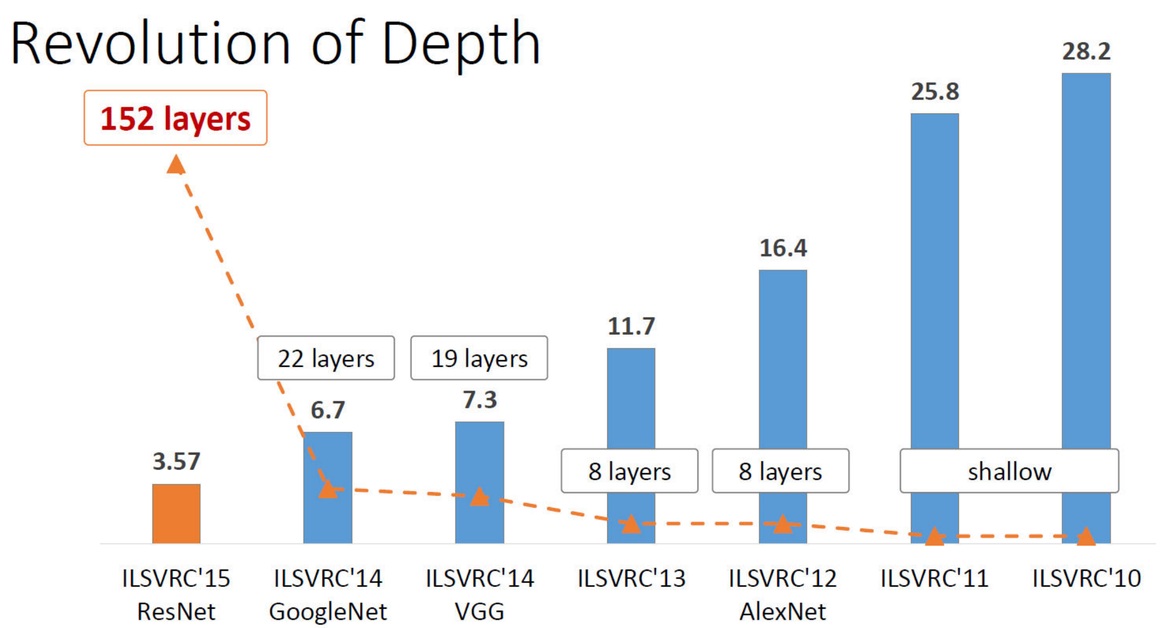
- Escape from 10 layers
  - Normalized initialization
  - Intermediate normalization layers



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

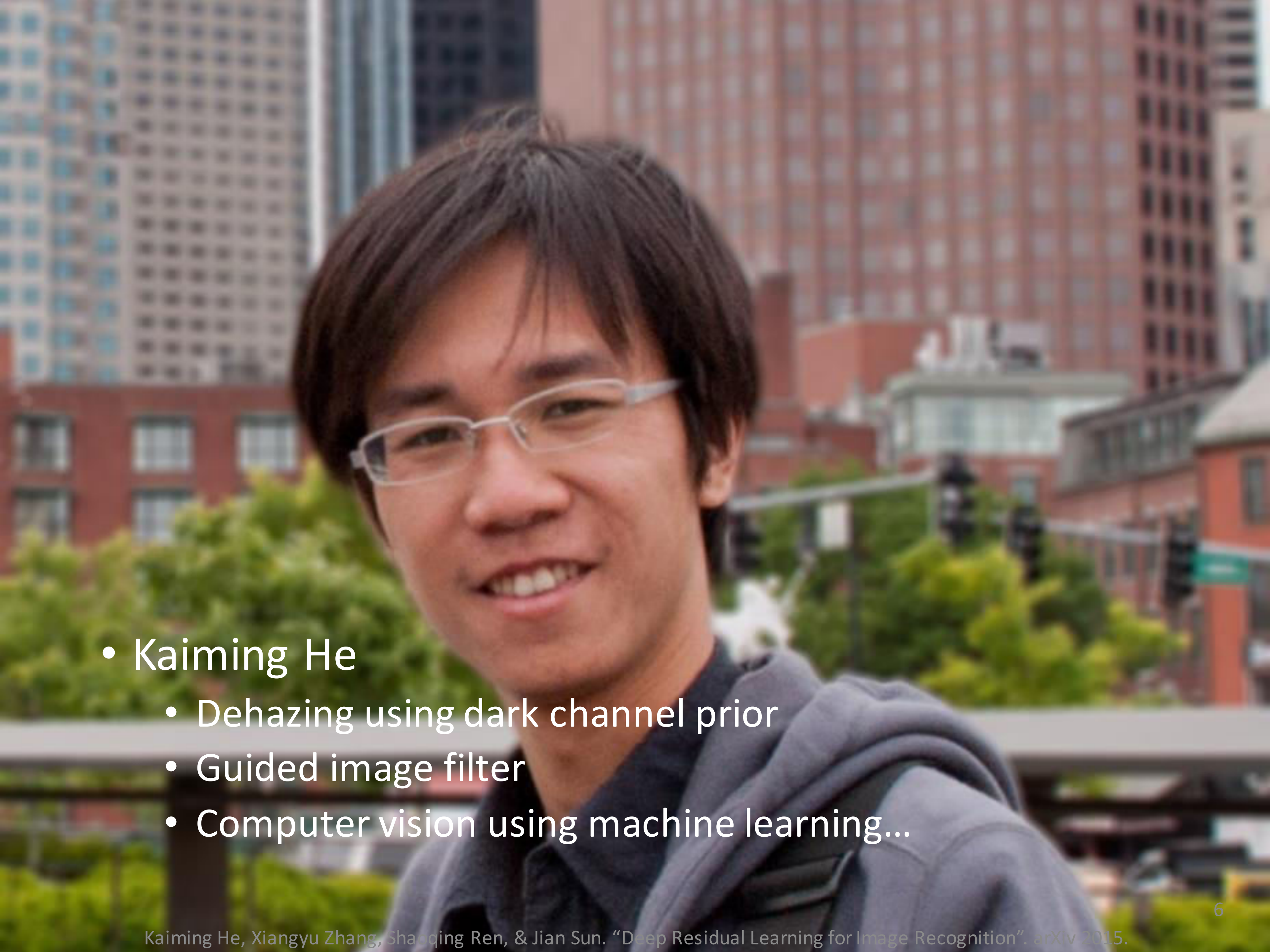
Deep Neural Network

- Escape from 100 layers
  - Residual network



ImageNet Classification top-5 error (%)

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

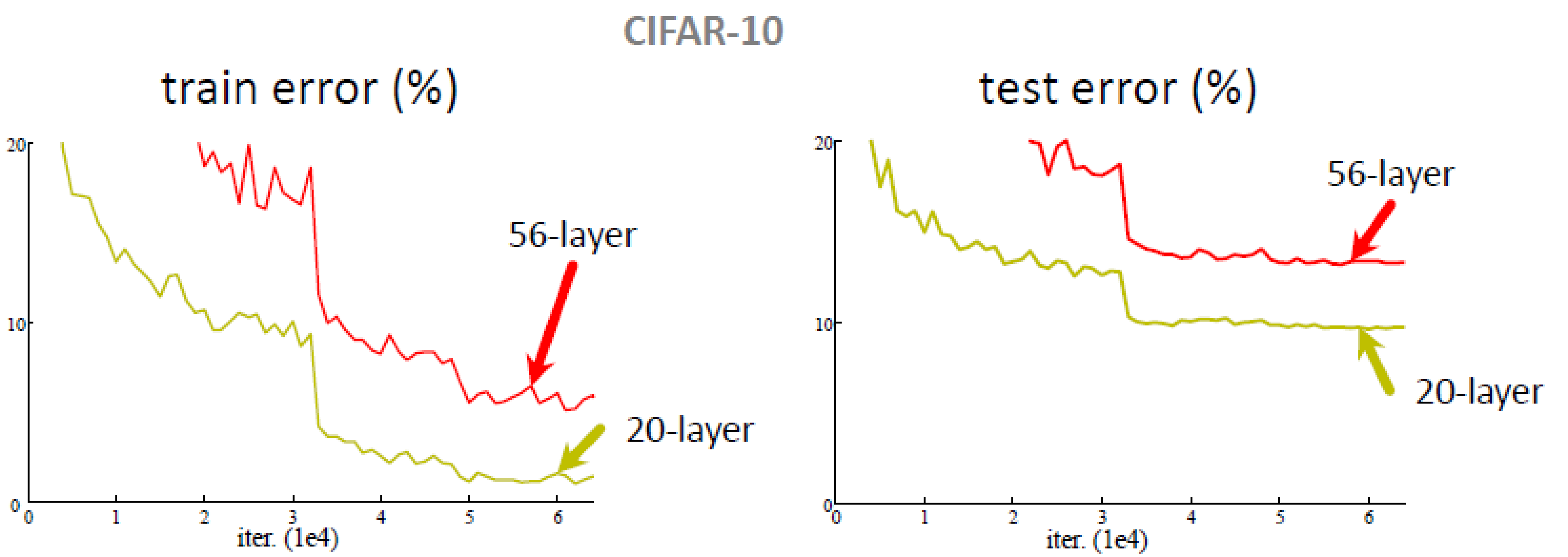


- Kaiming He
  - Dehazing using dark channel prior
  - Guided image filter
  - Computer vision using machine learning…

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Plain Network

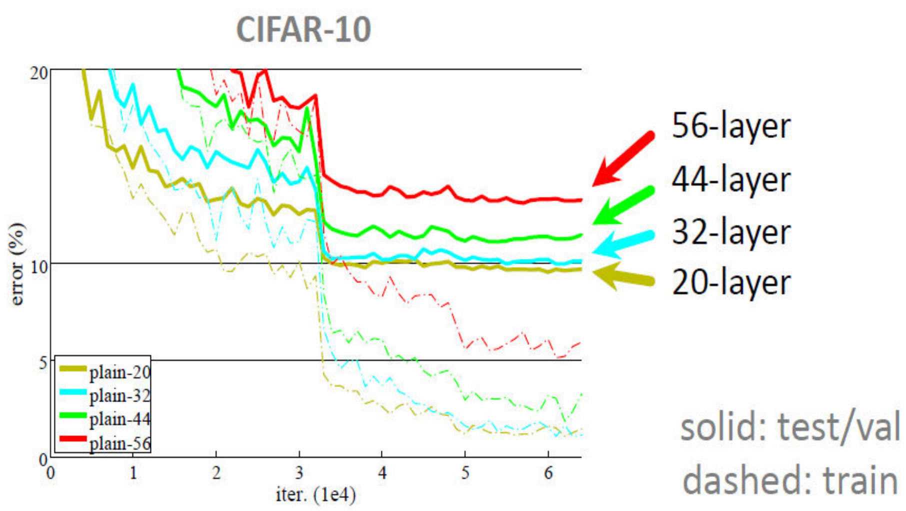
- Plain nets: stacking 3x3 conv layers
- 56-layer net has higher training error and test error than 20-layers net

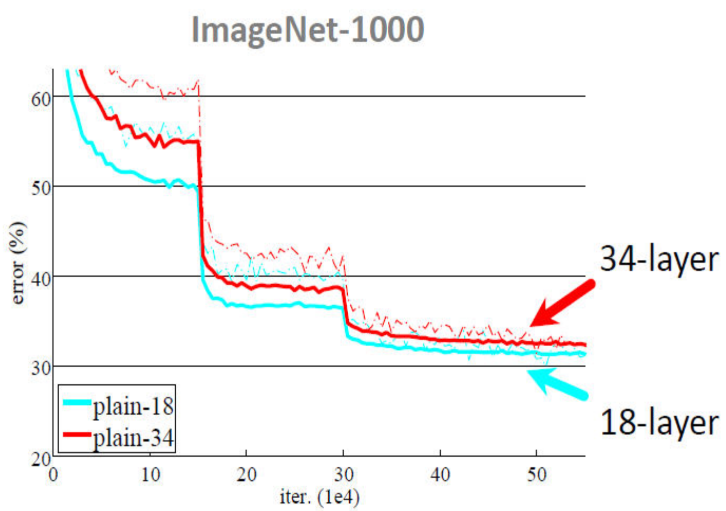


Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Plain Network

- "Overly deep" plain nets have higher training error
- A general phenomenon, observed in many datasets

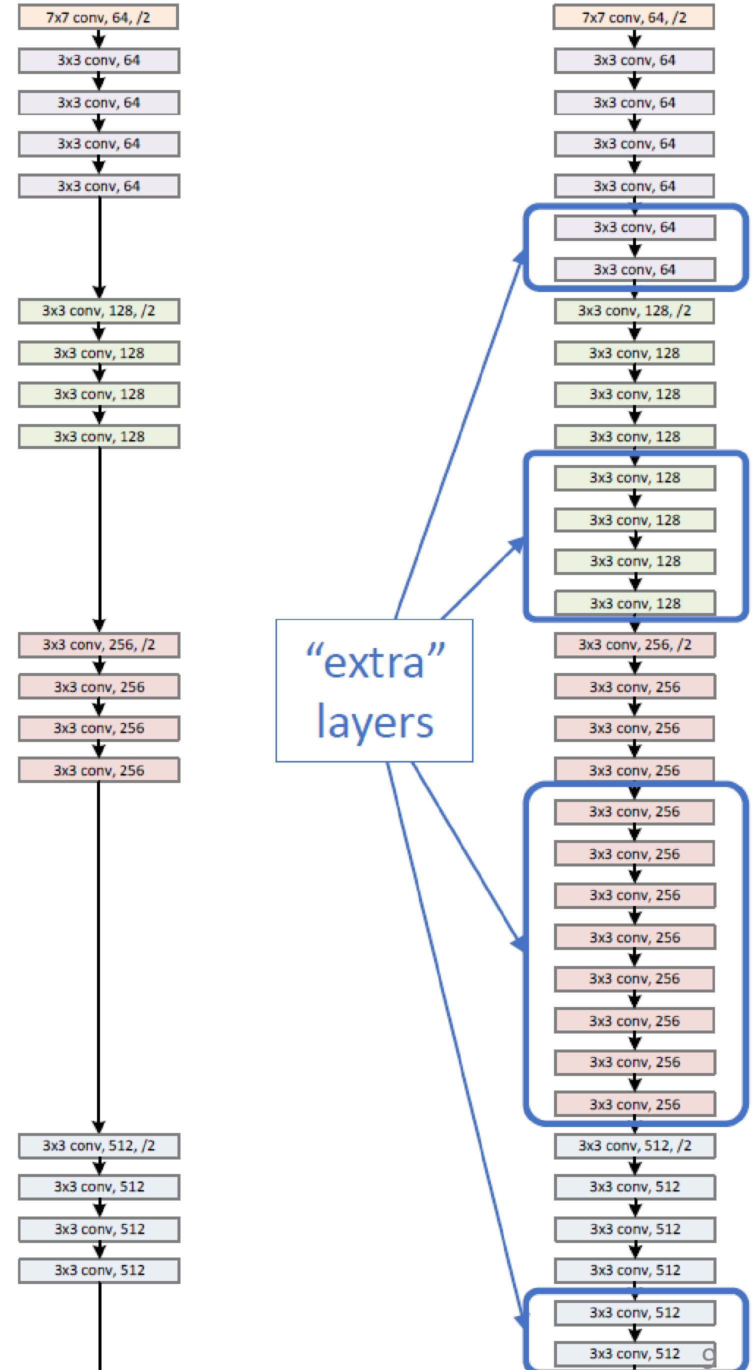




Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Residual Network

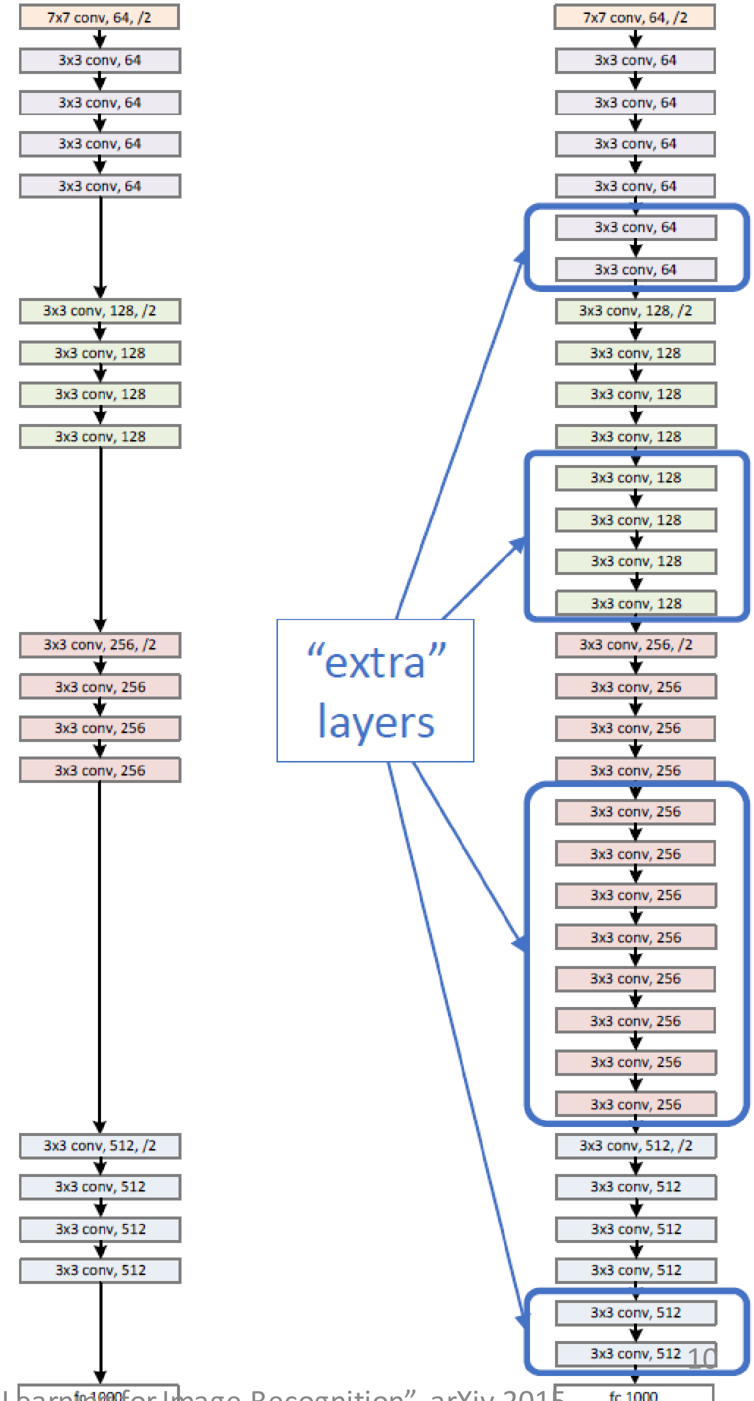
- Naïve solution
  - If extra layers are an identity mapping, then a training errors does not increase



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Residual Network

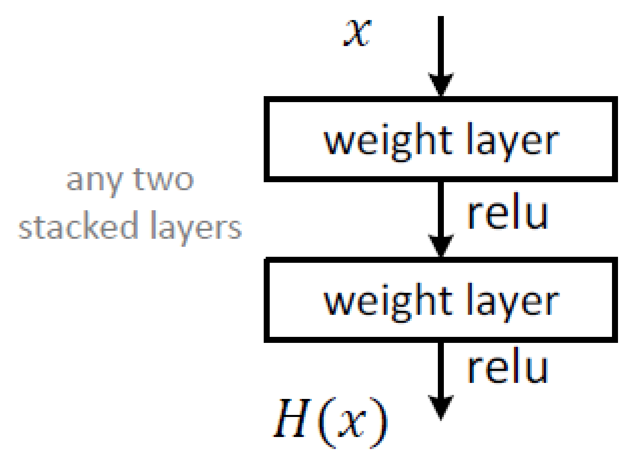
- Deeper networks also maintain the tendency of results
  - Features in same level will be almost same
  - An amount of changes is fixed
  - Adding layers makes smaller differences
  - Optimal mappings are closer to an identity



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Residual Network

- Plain block
  - Difficult to make identity mapping because of multiple non-linear layers

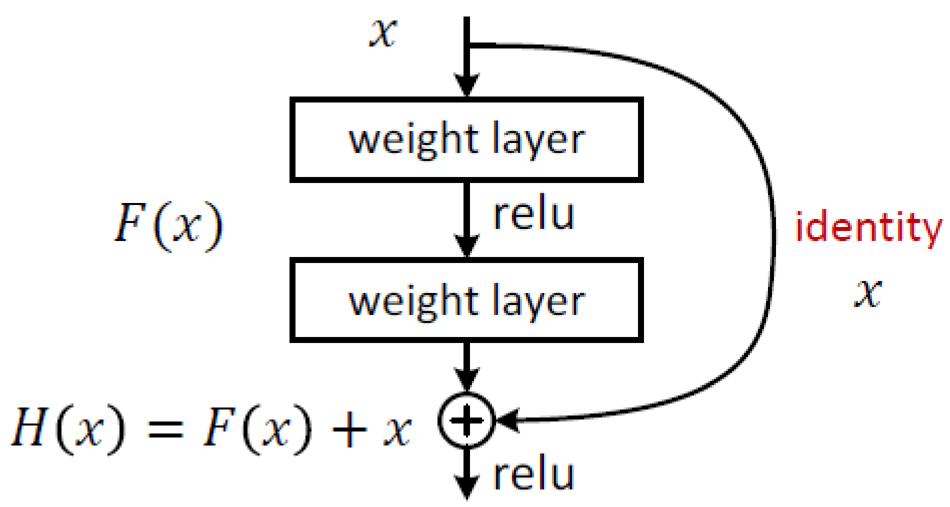


Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Residual Network

- Residual block
  - If identity were optimal, easy to set weights as 0
  - If optimal mapping is closer to identity, easier to find small fluctuations

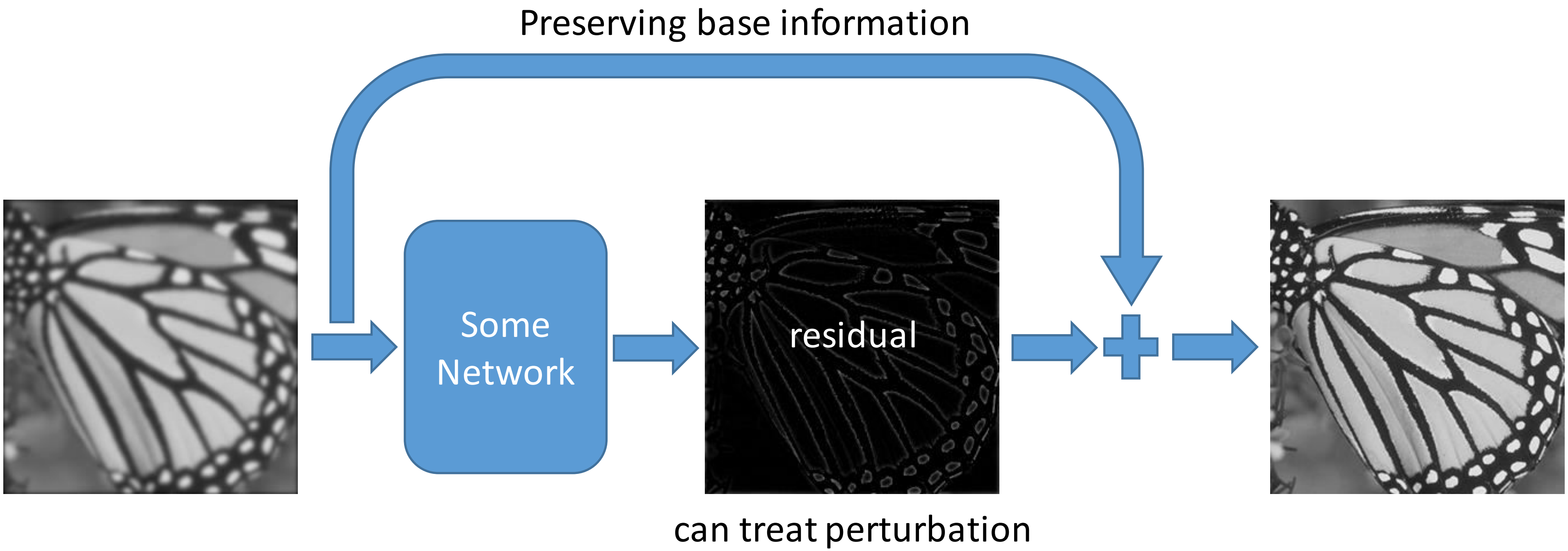
-> Appropriate for treating perturbation as keeping a base information



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

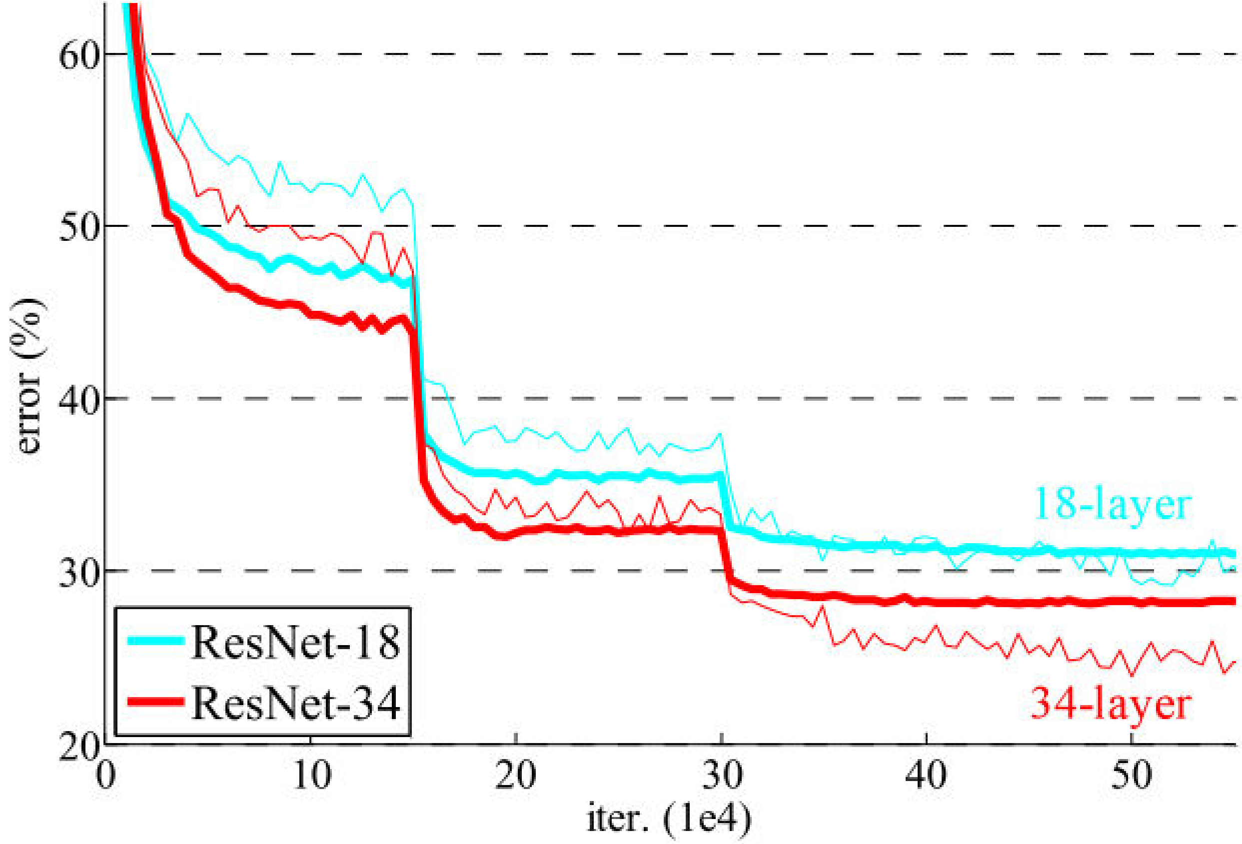
Residual Network

- Difference between an original image and a changed image



Residual Network

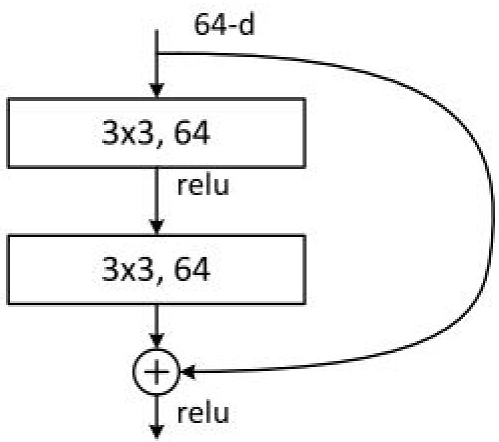
- Deeper ResNets have lower training error



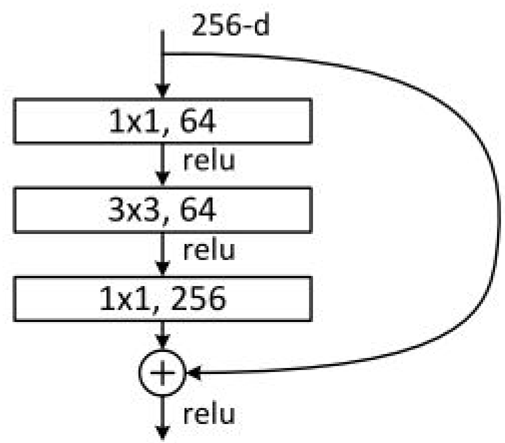
Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Residual Network

- Residual block
  - Very simple
  - Parameter-free



A naïve residual block



"bottleneck" residual block
(for ResNet-50/101/152)

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Residual Network

- Shortcuts connections
  - Identity shortcuts

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}.$$

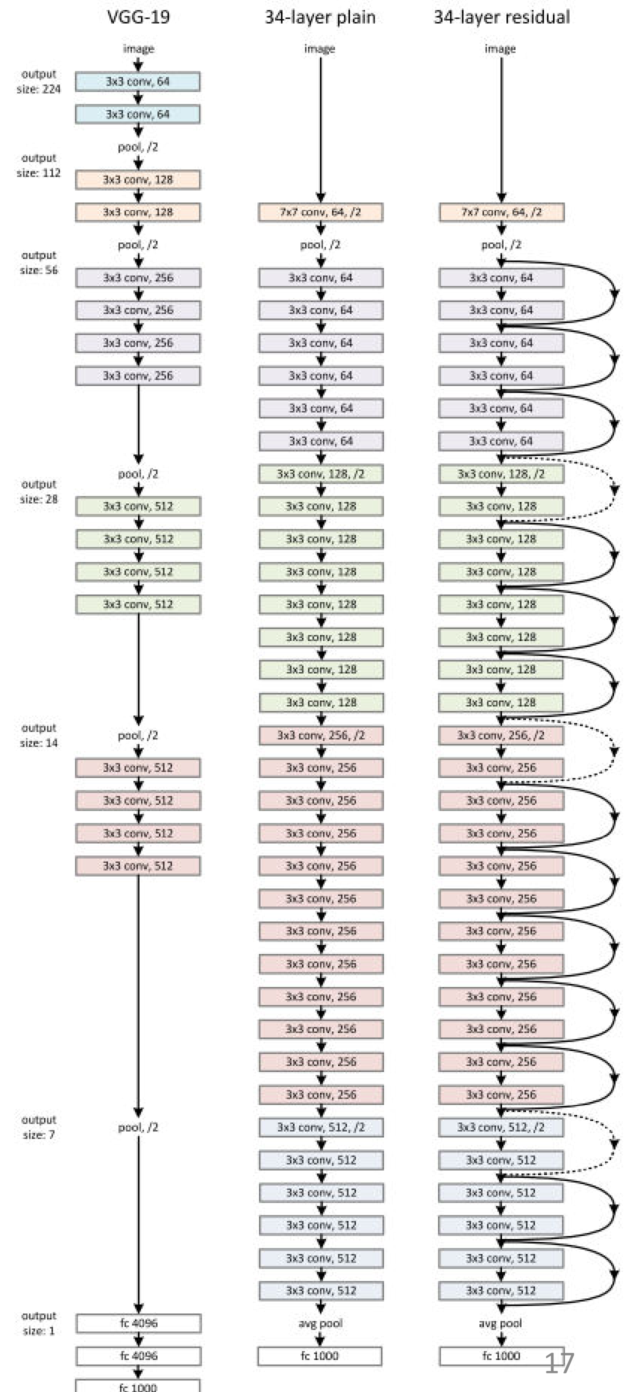
  - Projection shortcuts

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + W_s\mathbf{x}.$$

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Network Design

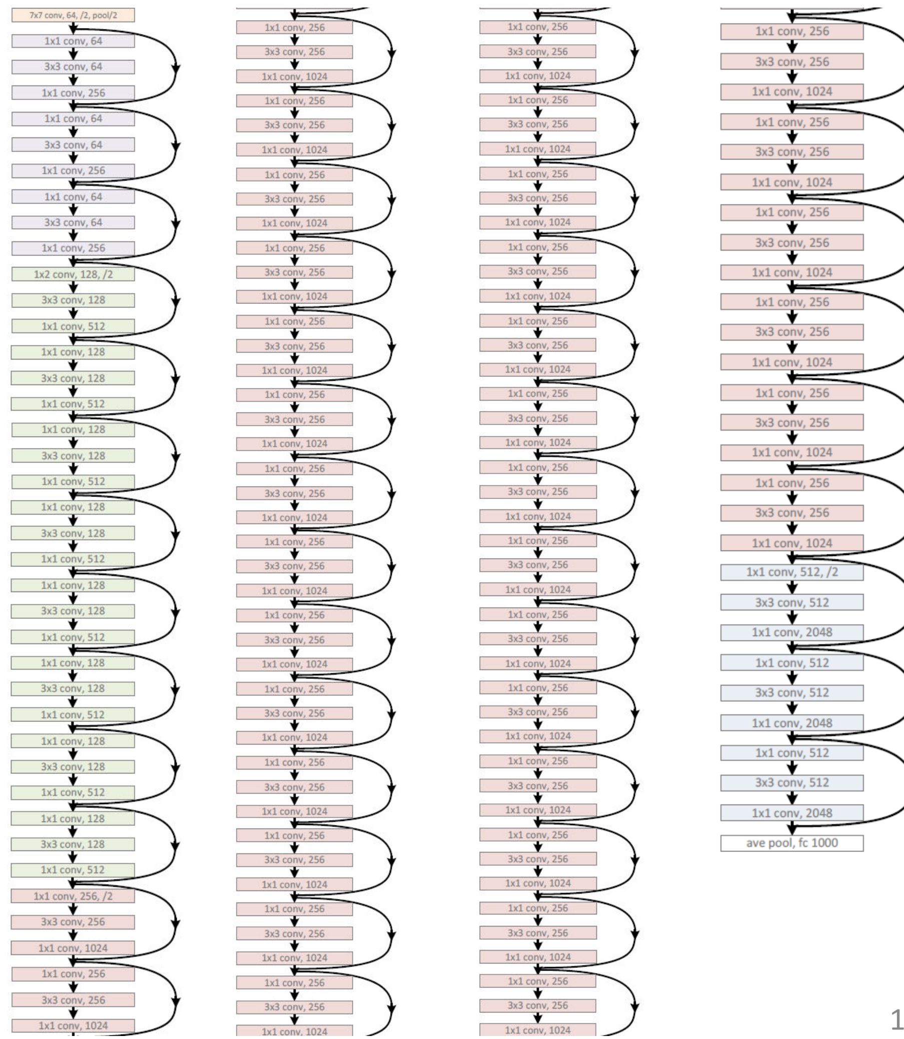
- Basic design (VGG-style)
  - All 3x3 conv (almost)
  - Spatial size/2 => #filters x2
  - Batch normalization
  - Simple design, just deep
- Other remarks
  - No max pooling (almost)
  - No hidden fc
  - No dropout



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Network Design

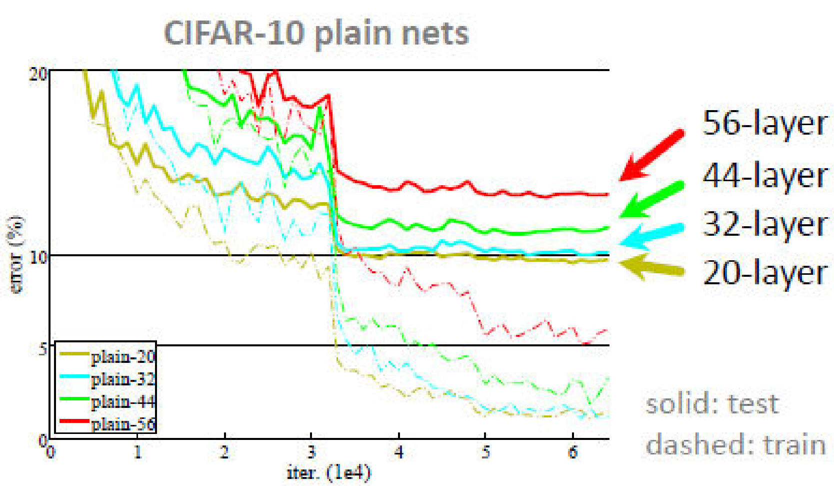
- ResNet-152
  - Use bottlenecks
  - ResNet-152(11.3 billion FLOPs) has lower complexity than VGG-16/19 nets (15.3/19.6 billion FLOPs)

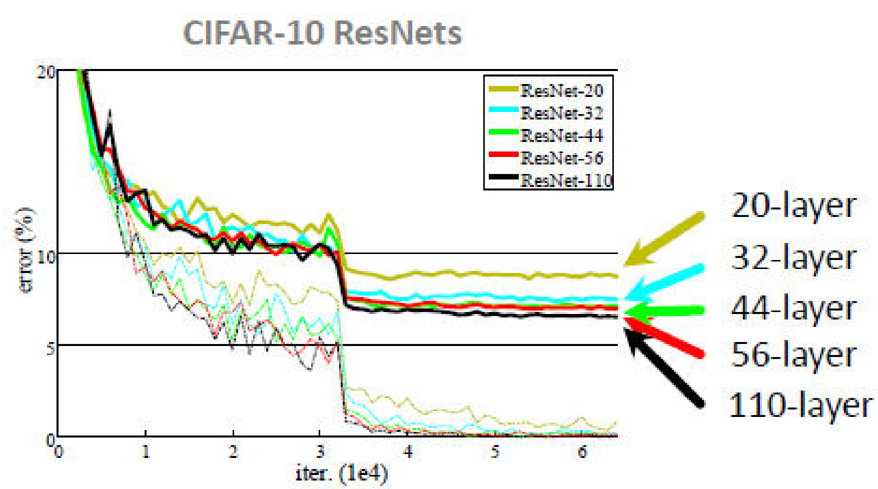


Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Results

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

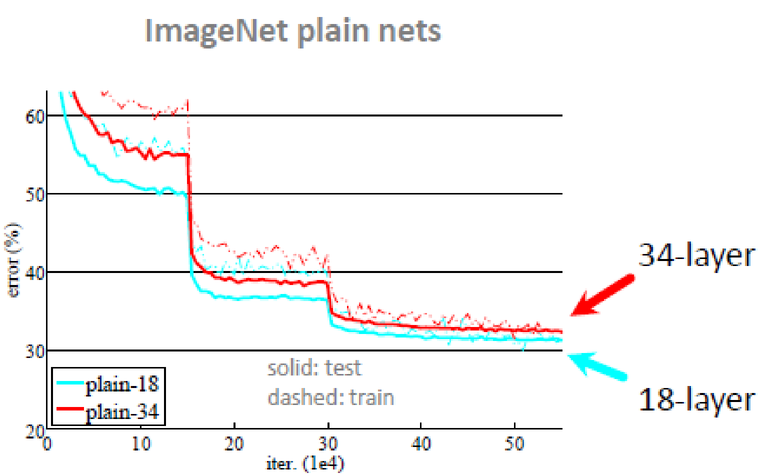


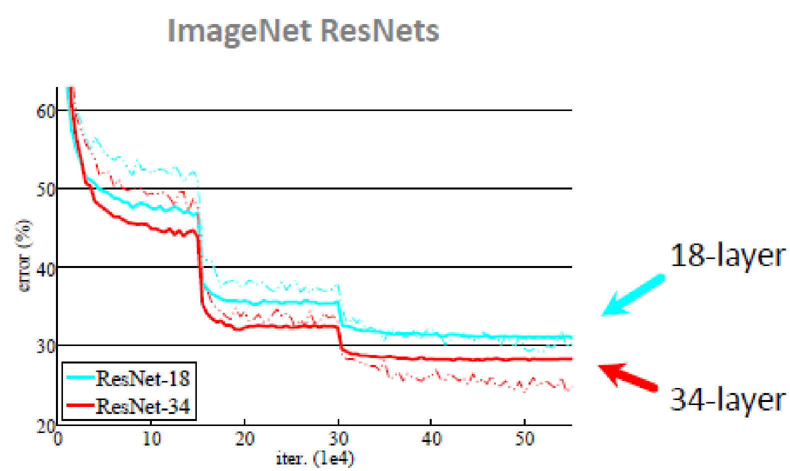


Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Results

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





Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

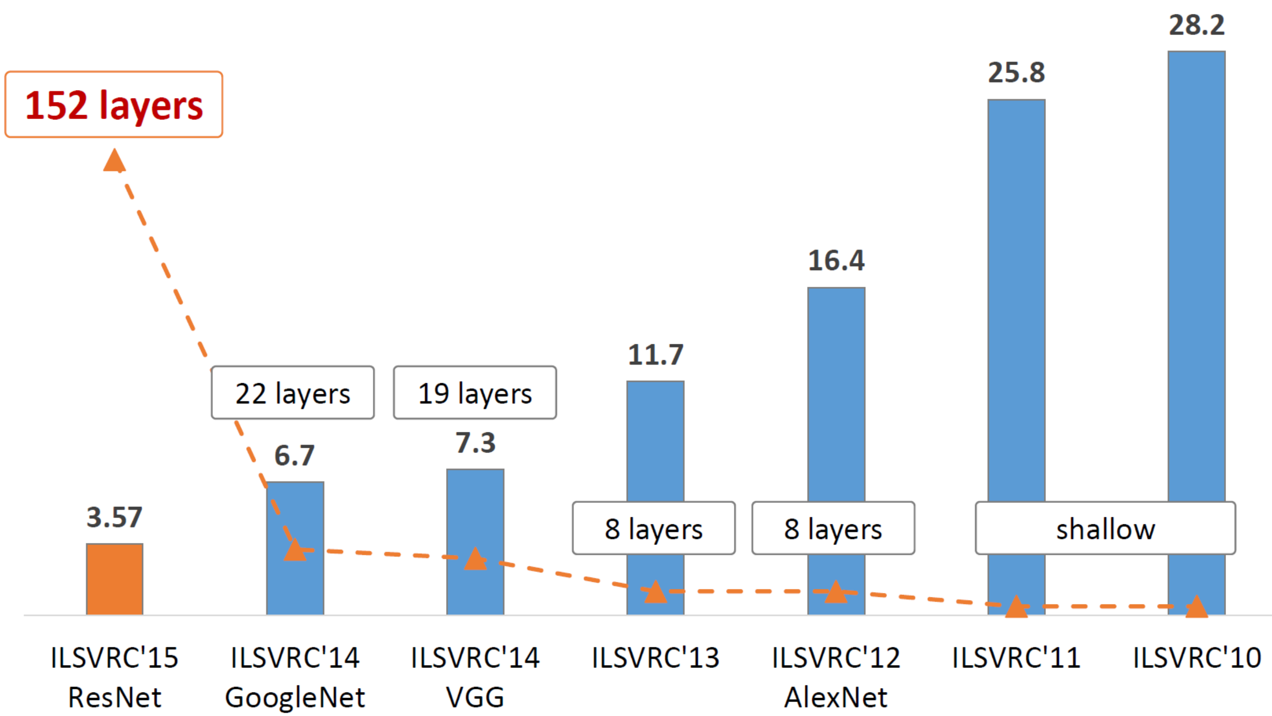
Results

- 1st places in all five main tracks in "ILSVRC & COCO 2015 Competitions"
  - ImageNet Classification
  - ImageNet Detection
  - ImageNet Localization
  - COCO Detection
  - COCO Segmentation

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Quantitative Results

- ImageNet Classification



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Result

- Performances increase absolutely

task	2nd-place winner	MSRA	margin (relative)
ImageNet Localization (top-5 error)	12.0	9.0	**27%**
ImageNet Detection (mAP@.5)	53.6	62.1	**16%**
COCO Detection (mAP@.5:.95)	33.5	37.3	**11%**
COCO Segmentation (mAP@.5:.95)	25.1	28.2	**12%**

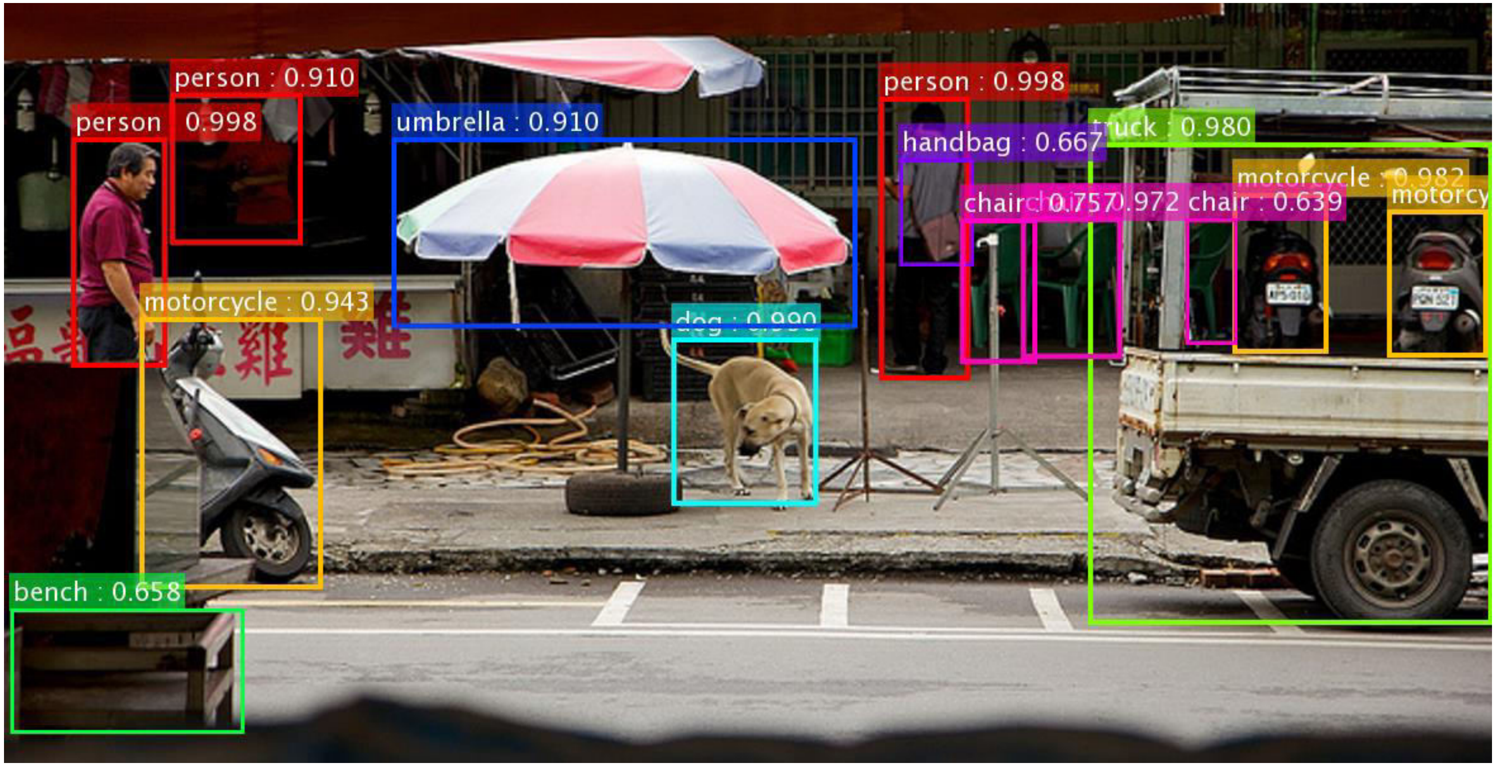
absolute 8.5% better!

- Based on ResNet-101
- Existing techniques can use residual networks or features from it

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Qualitative Result

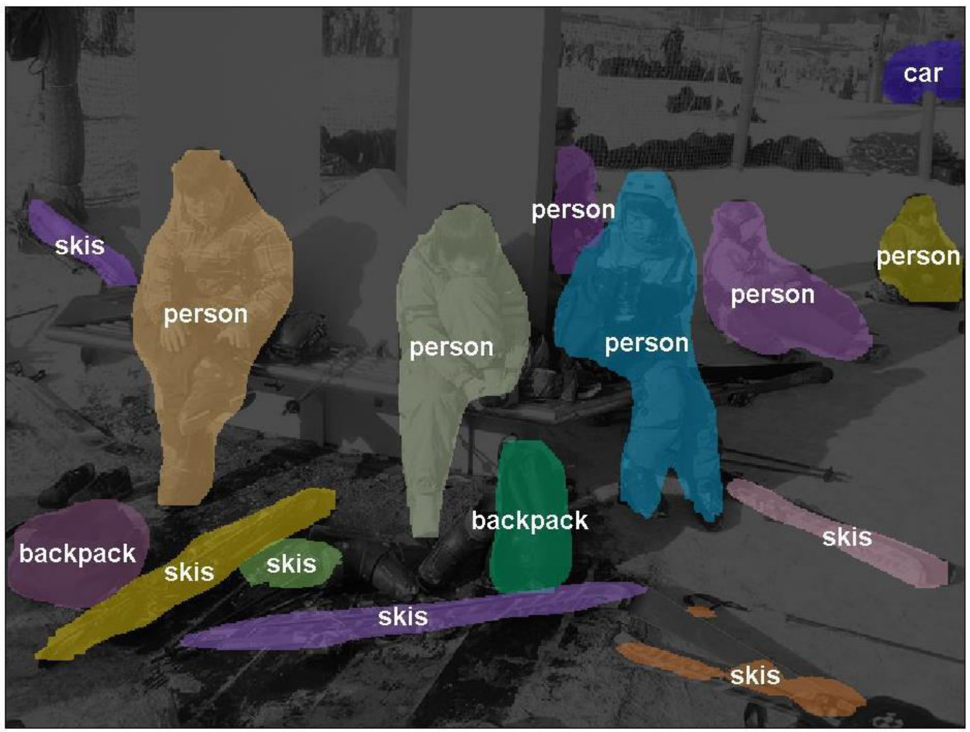
- Object detection
  - Faster R-CNN + ResNet



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.
Jifeng Dai, Kaiming He, & Jian Sun. “Instance-aware Semantic Segmentation via Multi-task Network Cascades”. arXiv 2015.

Qualitative Results

- Instance Segmentation



Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. "Deep Residual Learning for Image Recognition". arXiv 2015.

Exploring over 1000 layers

- Test 1202 layers
  - Training is finished
  - Training error is similar
  - Testing error is high because of over fitting

Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.

Conclusion

- Deeper networks are better expectably
- ResNet is very simple
- We should use it from now on

Reference

- Kaiming He, Xiangyu Zhang, Shaoqing Ren, & Jian Sun. “Deep Residual Learning for Image Recognition”. arXiv 2015.
- Slides of Deep Residual Learning @ ILSVRC & COCO 2015 competitions

Thank You