Microsoft

Benchmark Developments in Convolutional Neural Networks

Mithun Prasad, PhD
miprasad@Microsoft.com

Microsoft

AlexNet (2012)

ImageNet

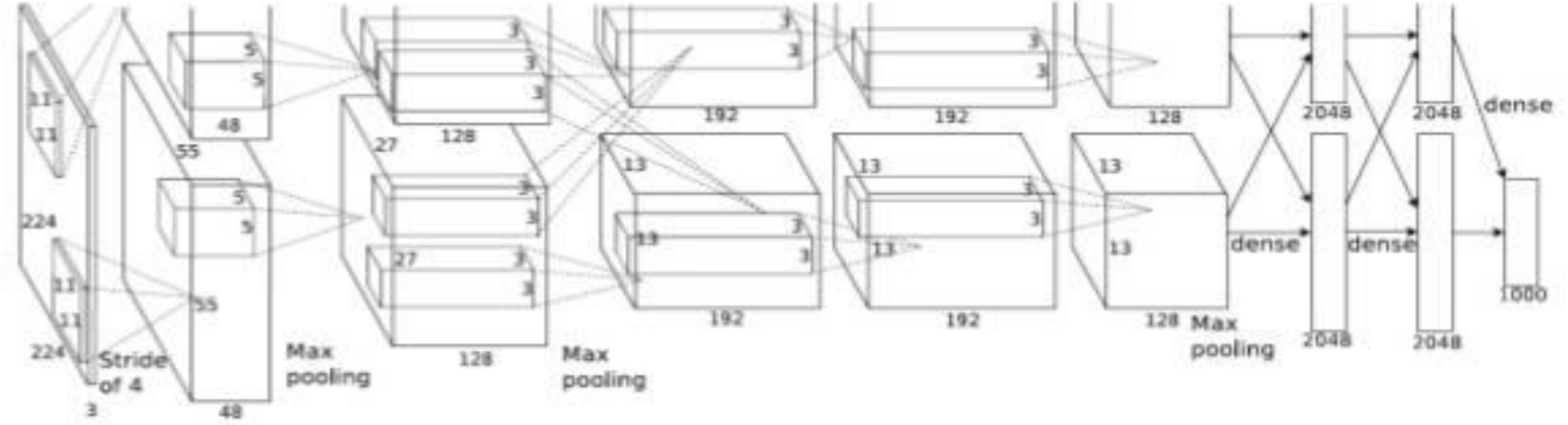
- Over 15M labeled high resolution images
- Roughly 22K categories
- Collected from web and labeled by Amazon Mechanical Turk

Microsoft

AlexNet Design

Microsoft

- Winner of ILSVRC 2012



A. Kryzhevsky, I, Sutskever, G.E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25 (NIPS2012)

Design Detail

- Trained the network on ImageNet data
  - >15 million annotated images from a total of over 22,000 categories
- ReLU
- Data augmentation techniques
  - Image translations, horizontal reflections, and patch extractions
- Dropout layers to combat overfitting
- Training using batch stochastic gradient descent
- Trained on two GPUs for five to six days

AlexNet Contributions

AlexNet was the first to put together several key advances:

1. ReLU
2. Dropout
3. Data Augmentation
4. Multiple GPUs

While not all invented by the AlexNet group, they were the first to put them all together and figure out how to train a deep neural network.

Microsoft

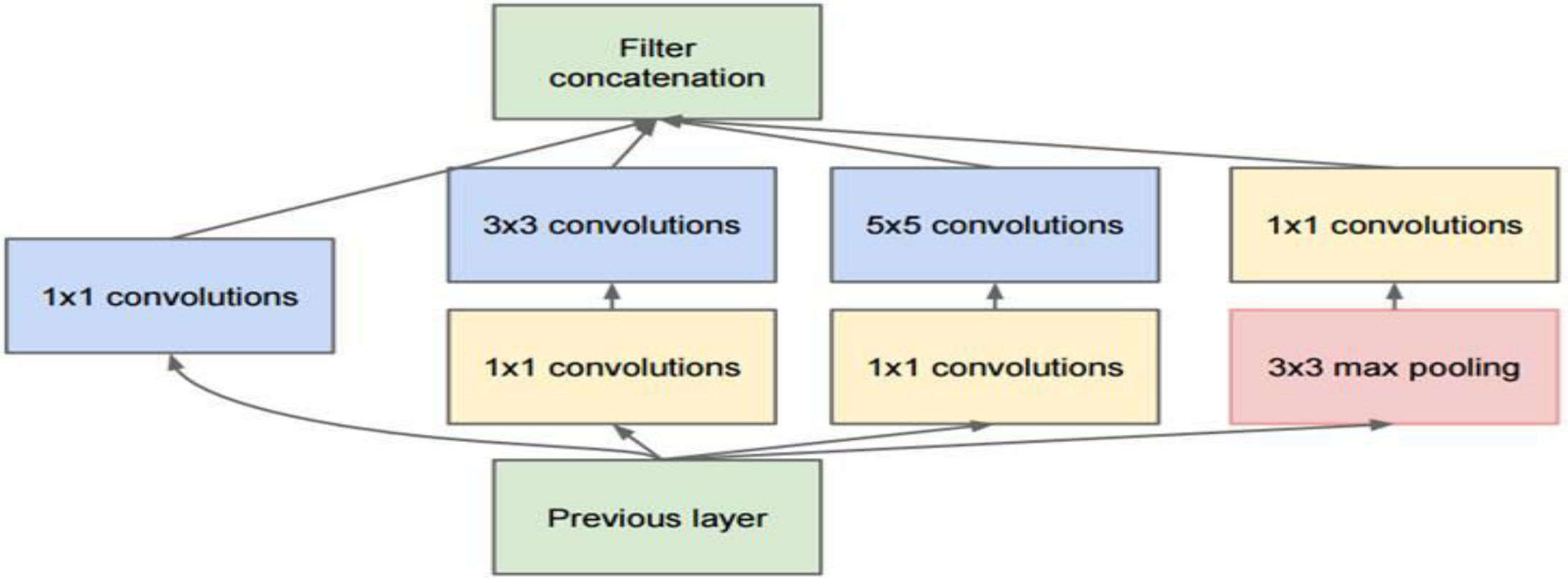
GoogleNet (2014)

GoogleNet Design

- Winner of ILSVRC 2014
- CNN layers didn't always have to be stacked up sequentially
- Computational and memory cost

GoogleNet (Inception Module)

- Ensemble model performs better than if you had a simple convolution



Microsoft

GoogleNet Contributions

- 9 Inception modules in the network, with over 100 layers in total
- Uses 12 times fewer parameters than AlexNet
- During testing, multiple crops were created and fed to the network. Softmax probabilities were averaged to give us the final solution
- Updated versions to the Inception module
- Trained on “a few high-end GPUs within a week”

Microsoft

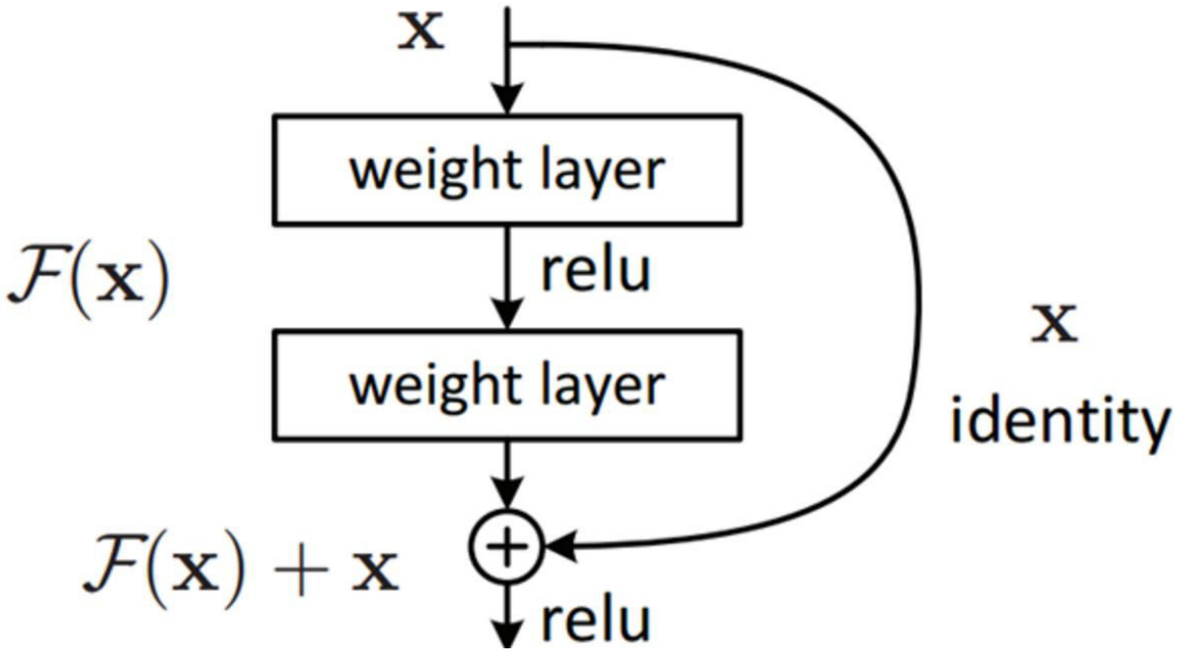
ResNet (2015)

ResNet Design

- Winner of ILSVRC 2015
- Microsoft Research Asia Innovation
- 152 layer architecture
- Residual learning

ResNet – Residual Block

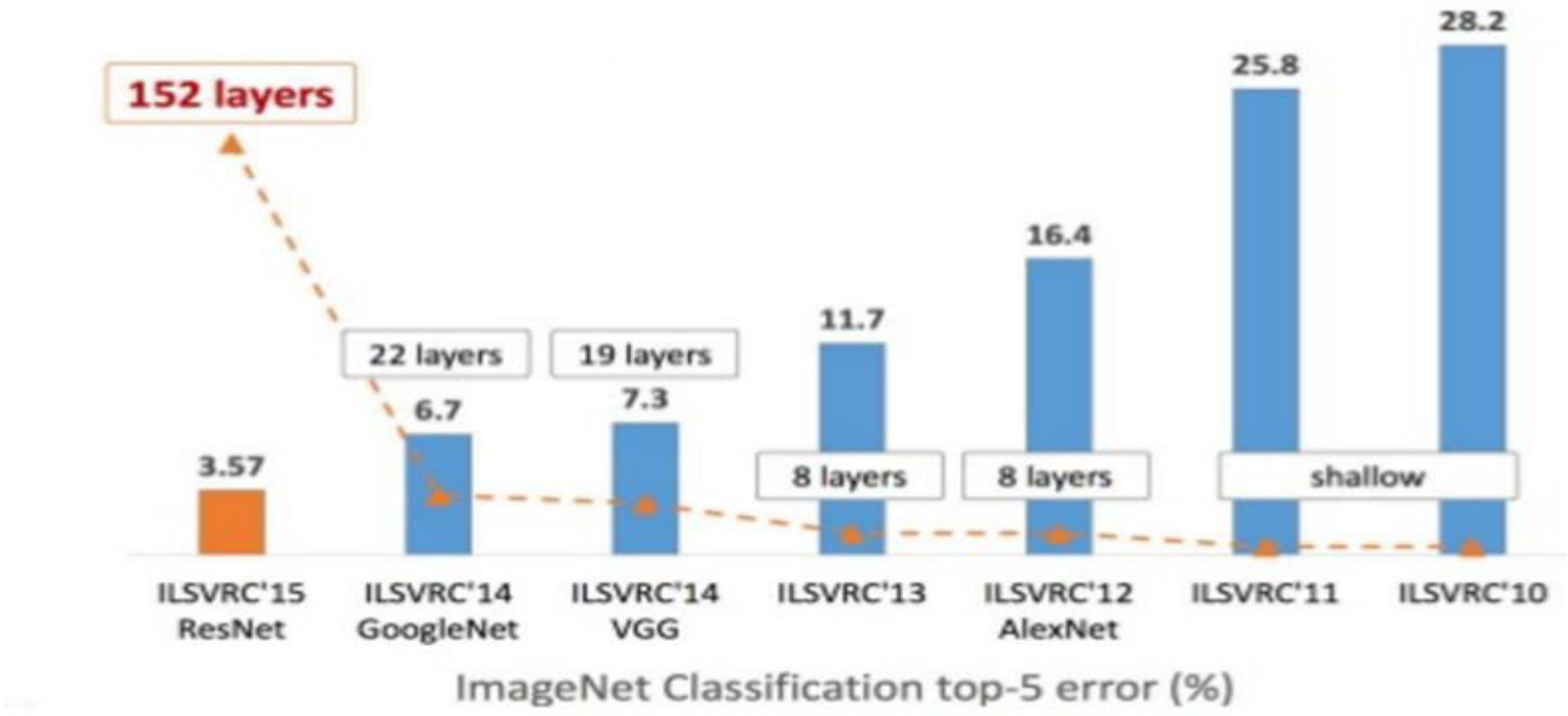
- With traditional CNNs, x maps to $f(x)$
- Traditional CNNs doesn't keep information about the original x
- Residual block computes a slight change to the original input x (altered representation)



ResNet Contributions

- Ultra-deep - 152 layers
- After only the first 2 layers, the spatial size gets compressed from an input volume of 224x224 to a 56x56 volume.
- Residual Block
- The group tried a 1202-layer network, but got a lower test accuracy, presumably due to overfitting.
- Trained on an 8 GPU machine for two to three weeks.
- Lowest error rate (3.6%)

Classification Trend on ImageNet



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

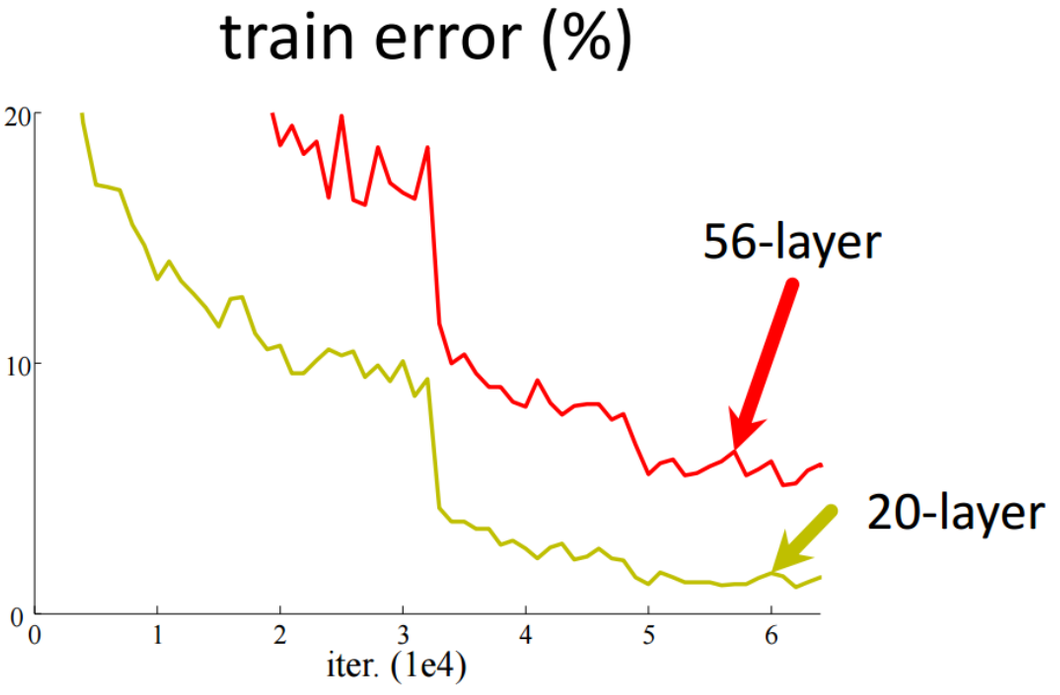
Microsoft

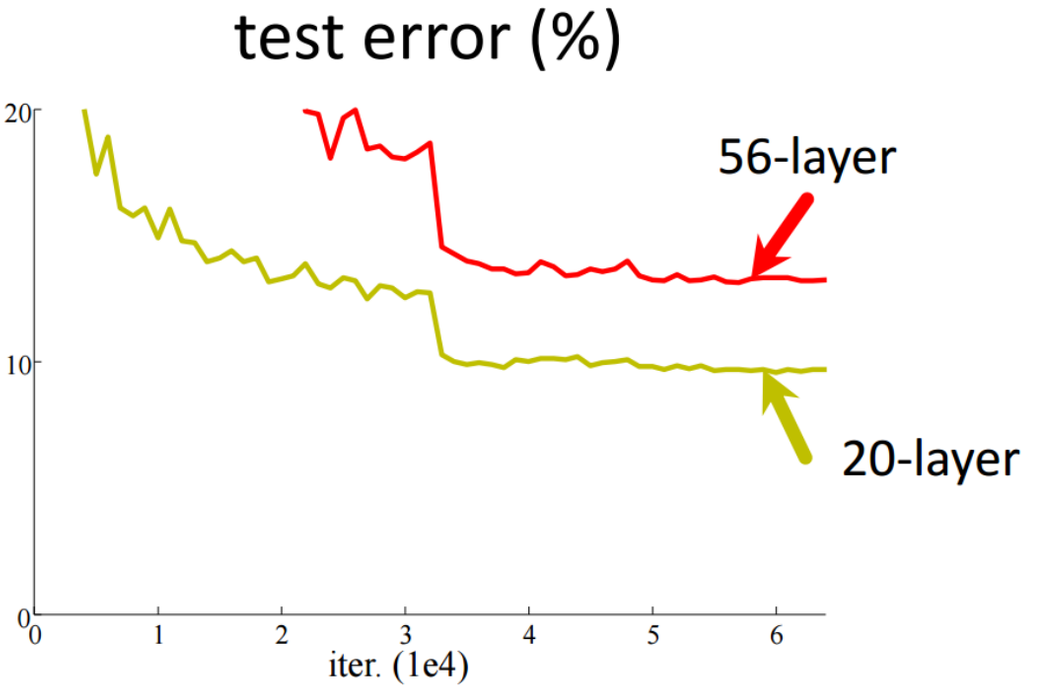
Microsoft

Is learning better networks as simple as stacking more layers?

Simply Stacking Layers?

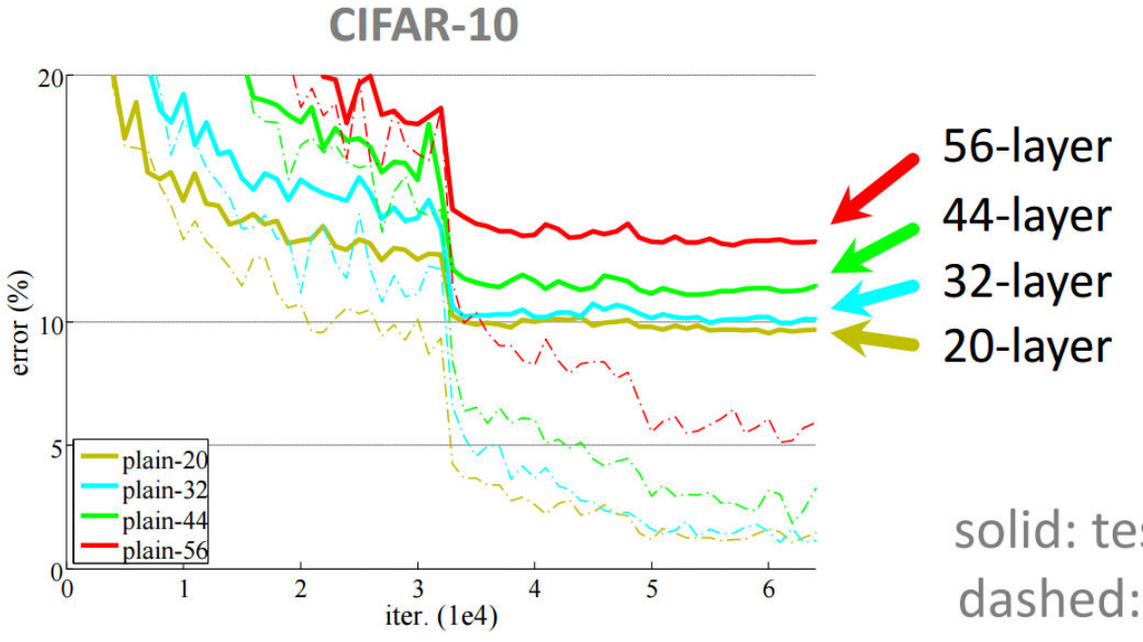
- CIFAR-10 (60000 32x32 colour images in 10 classes)
- Stacking 3x3 conv layers

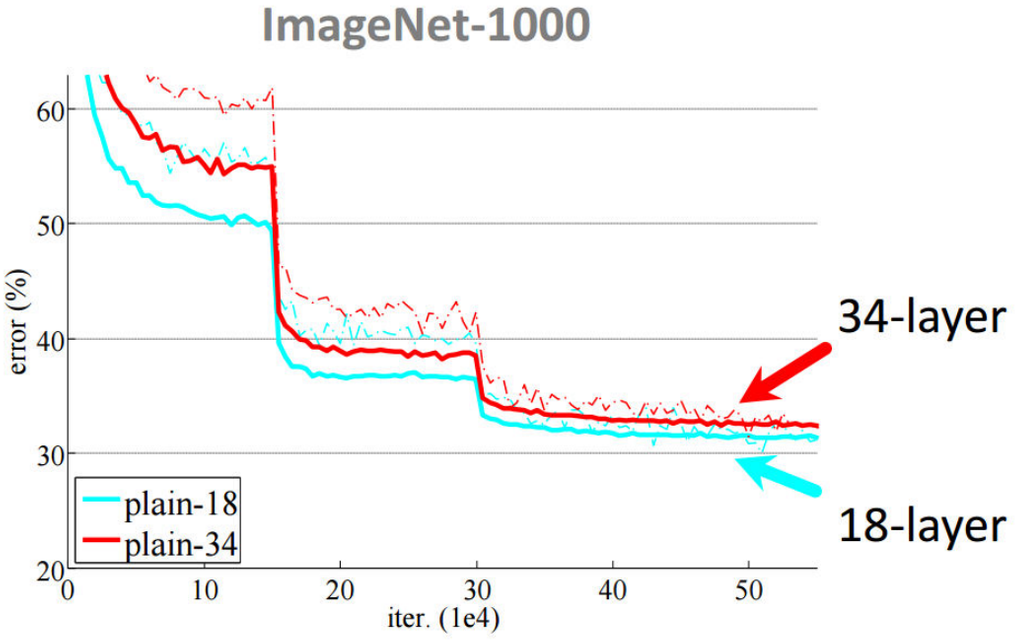




Simply Stacking Layers?

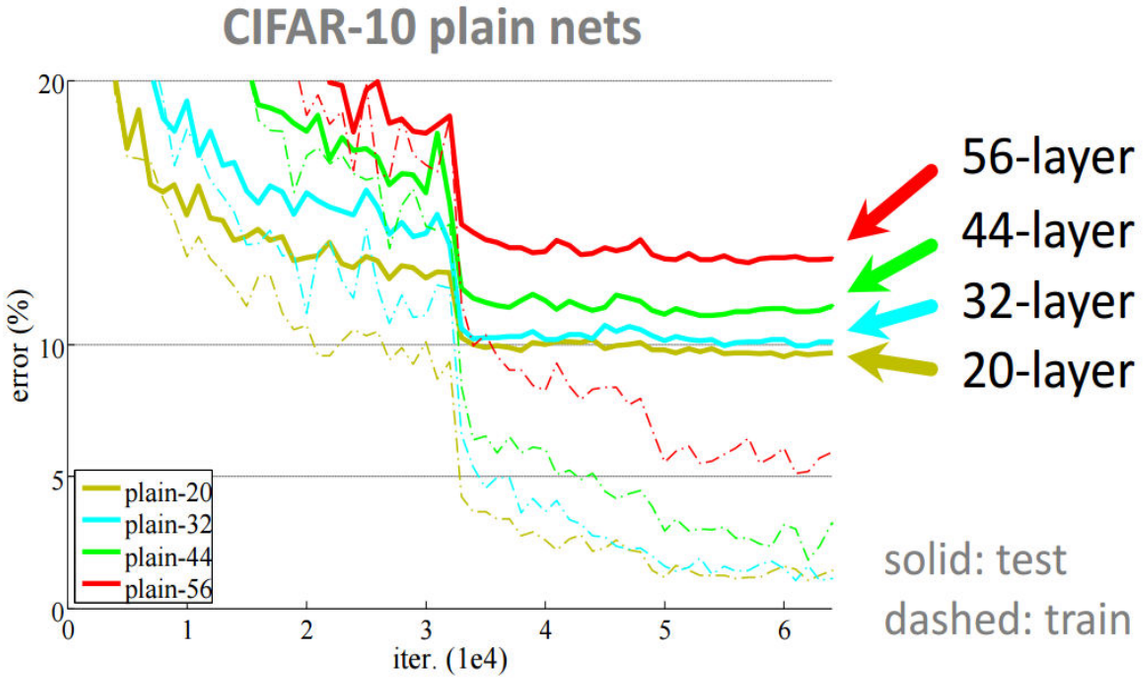
- "Overly deep" plain nets have higher errors
- Observed in many datasets

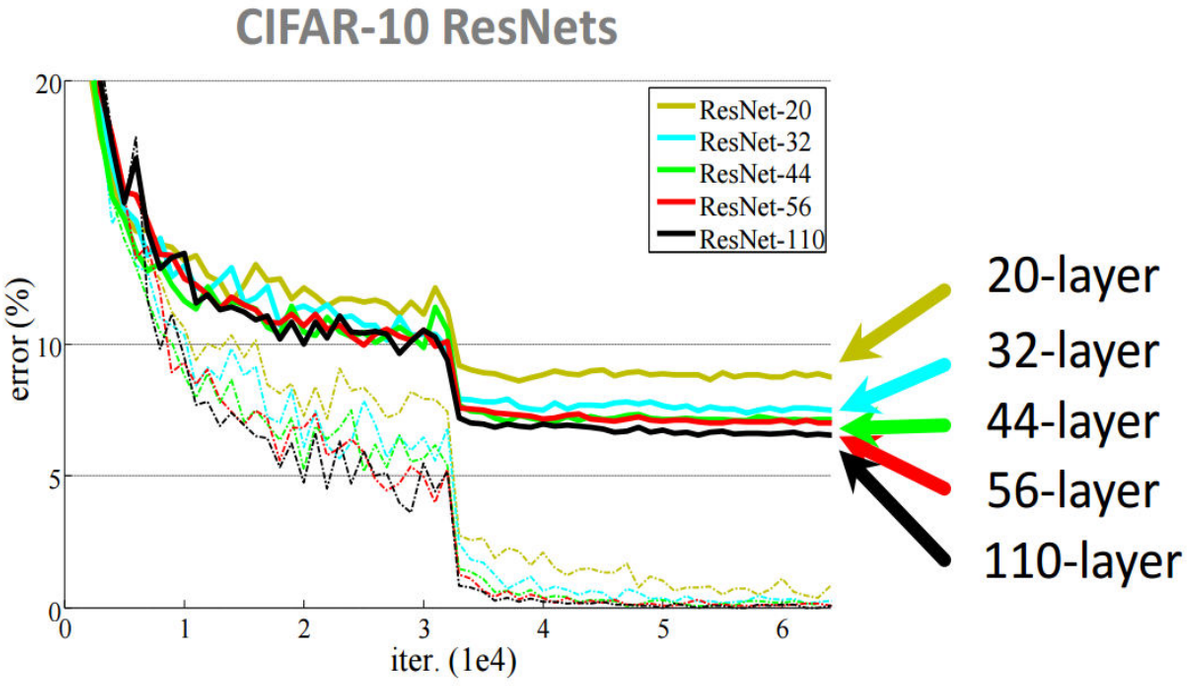




Deeper ResNets

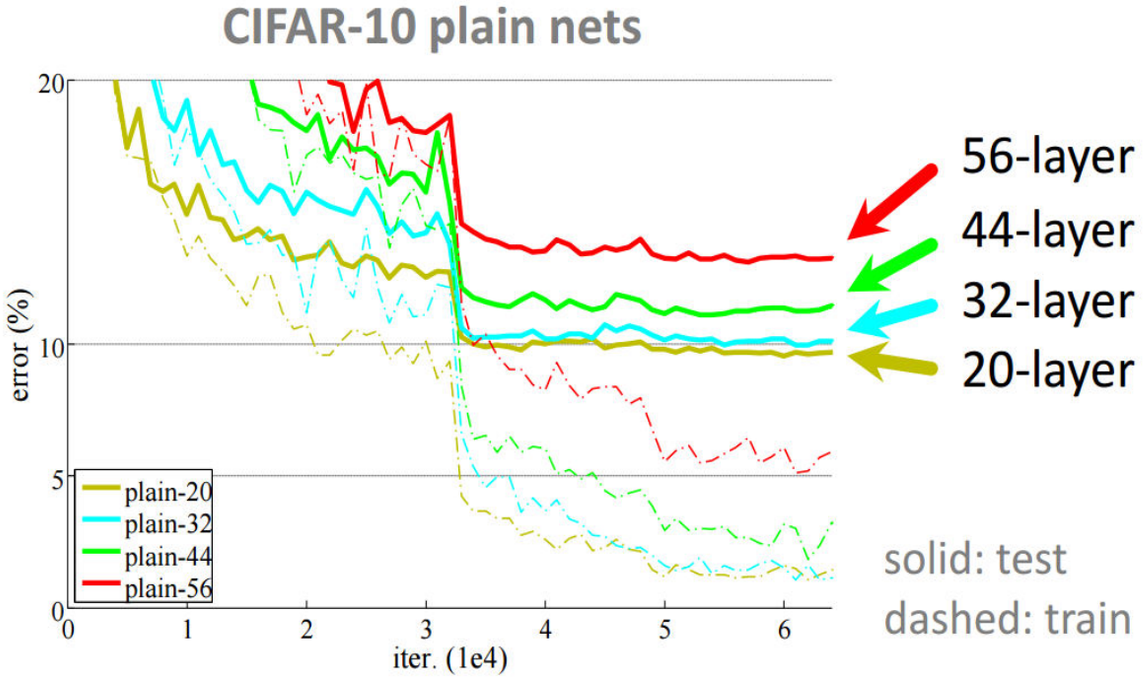
- CIFAR-10
- Deeper ResNets have lower errors

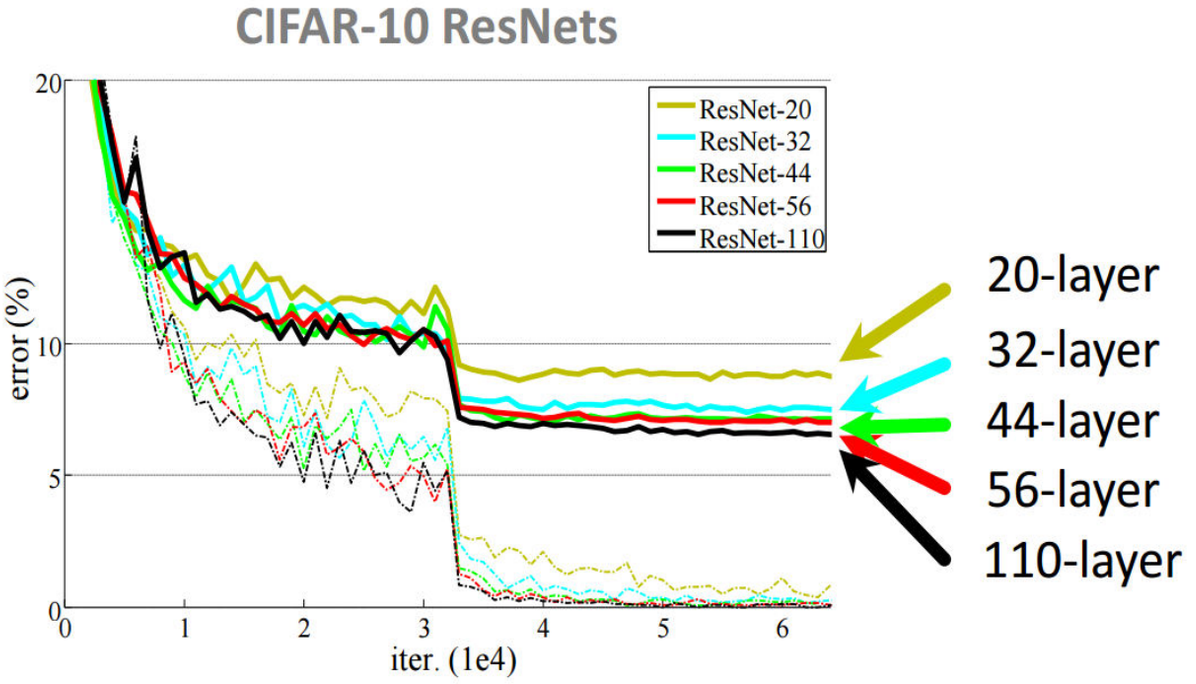




Deeper ResNets

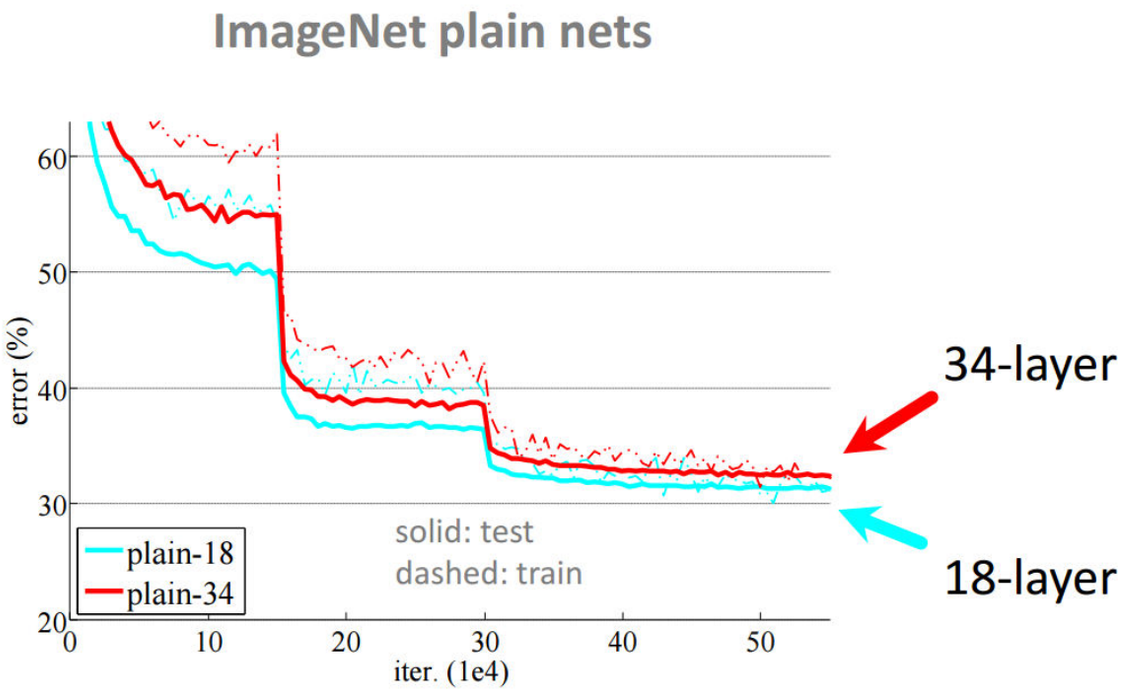
- CIFAR-10
- Deeper ResNets have lower errors

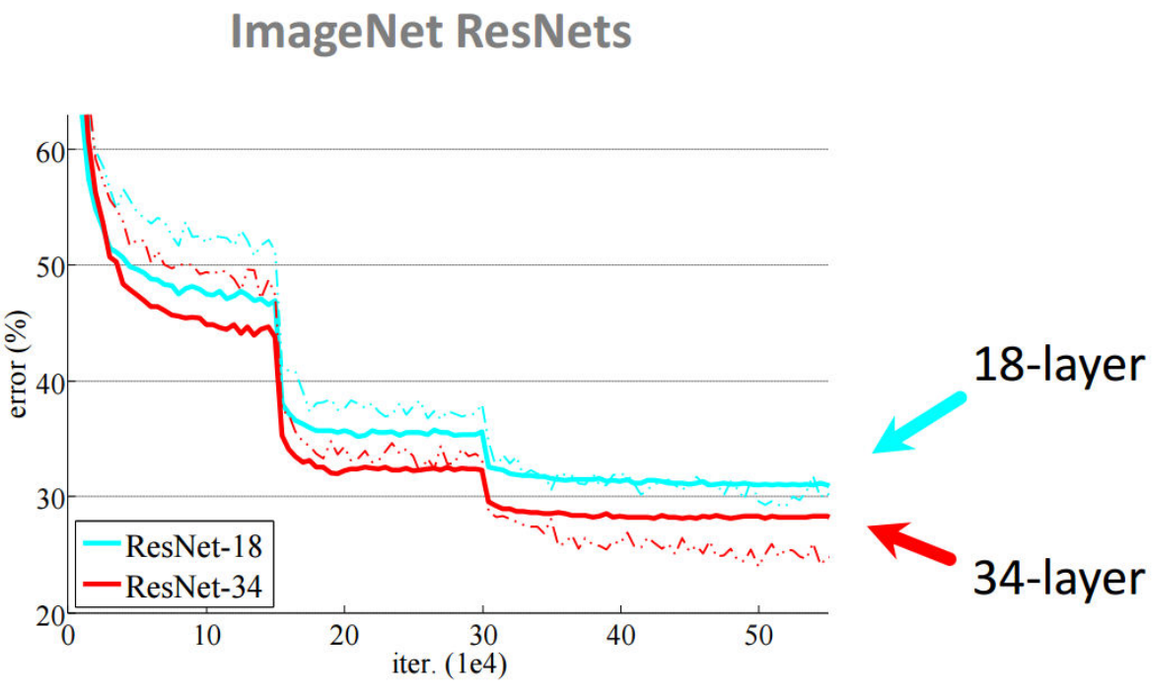




Deeper ResNets

- ImageNet
- Deeper ResNets have lower errors





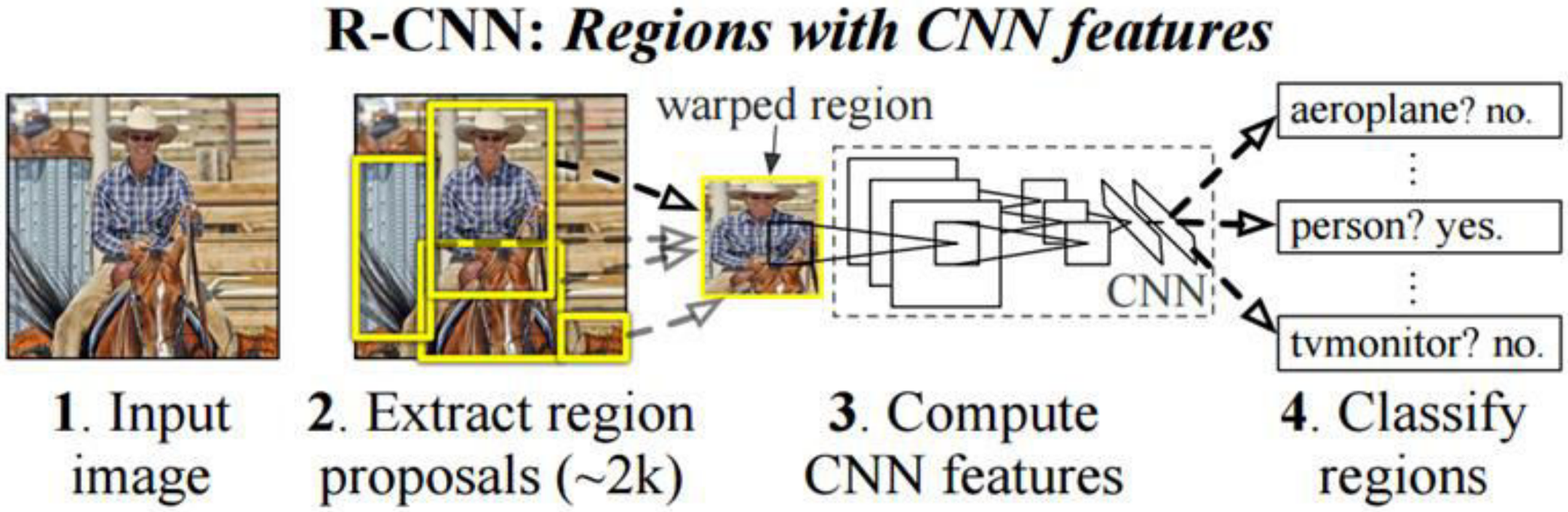
Microsoft

Region Based CNNs

Object Detection

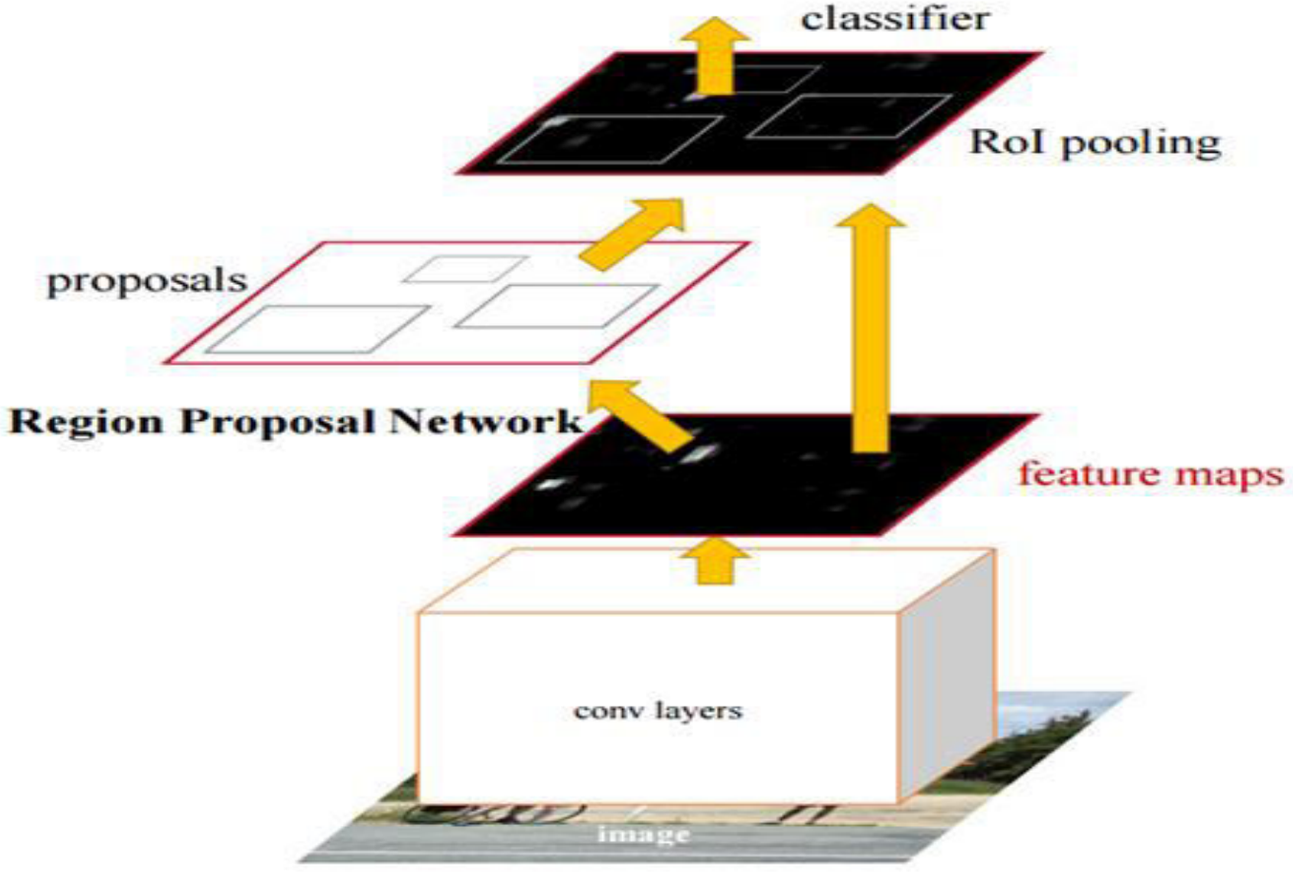
Object detection

1. Region proposal step - Selective search
2. Classification step



Region Based CNNs Contributions

- Localisation and classification of objects
- Faster R-CNN is the benchmark for object detection programs today
- Faster R-CNN:



Microsoft