Recap AlexNet: What’s next?

How to improve AlexNet architecture?

+++Deep?

+++Convolutional?

+++Fully connected?

All?

⇒A lot of empirical studies

- ⇒Tuning various design parameters
- ⇒what really works?

⇒Winners: GoogLeNet, VGG, ResNet

Recap AlexNet: What’s next?

How to improve AlexNet architecture?

+++Deep?

+++Convolutional?

+++Fully connected?

All?

⇒A lot of empirical studies

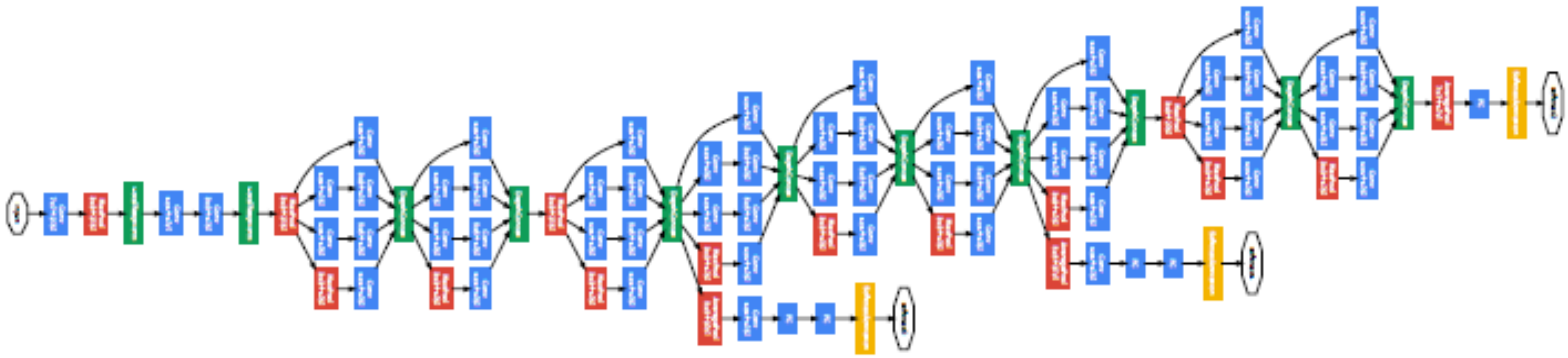
- ⇒Tuning various design parameters
- ⇒what really works?

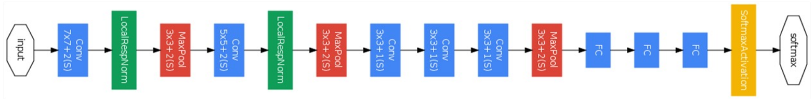
⇒Winners: **GoogLeNet**, VGG, ResNet

GoogLeNet (2014)

Winner of ILSVRC -2014. Very deep network with 22 layers:

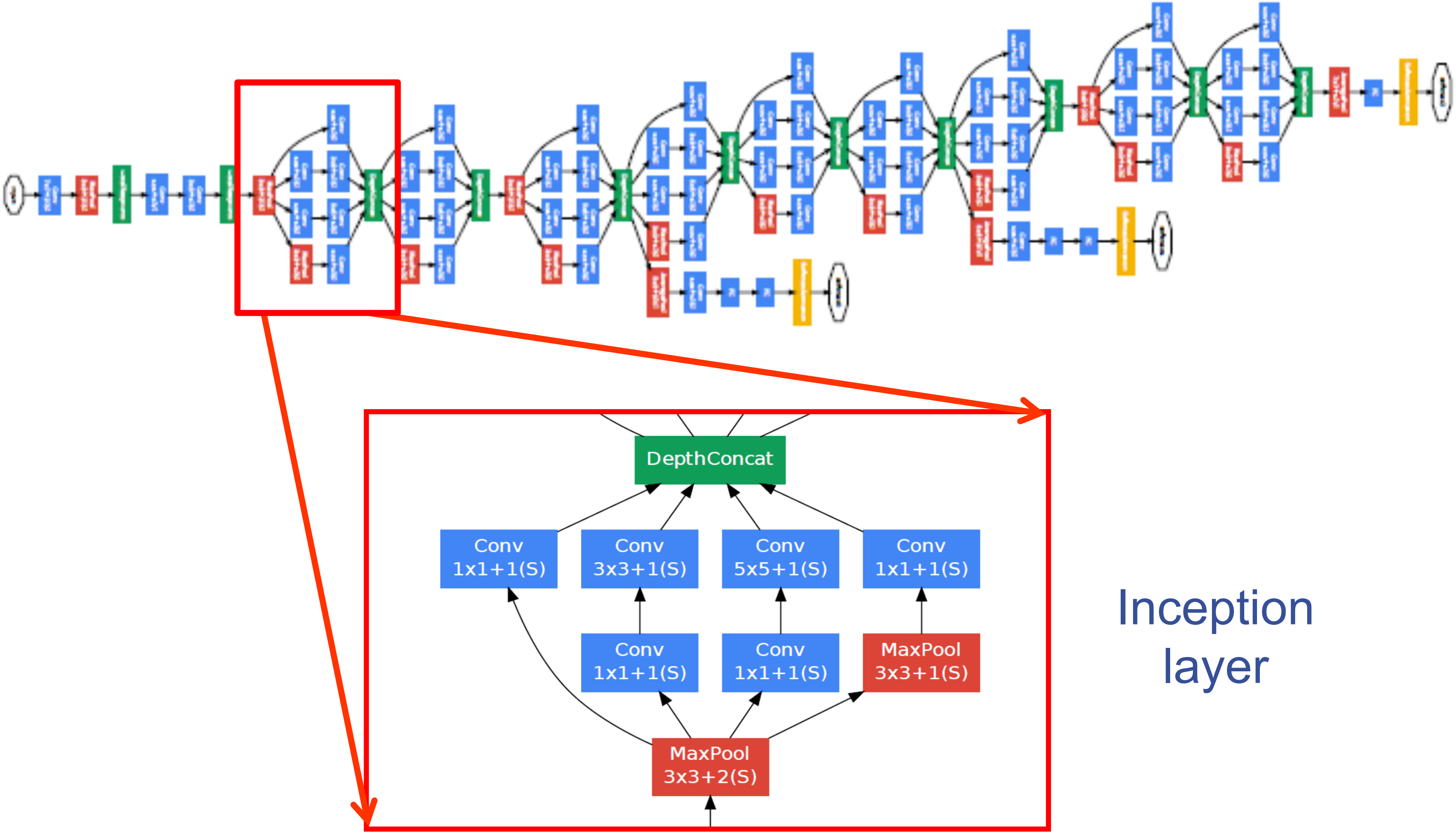
- Network–in-network–in-network
- Removed fully connected layers → small # of parameters (5M weights)



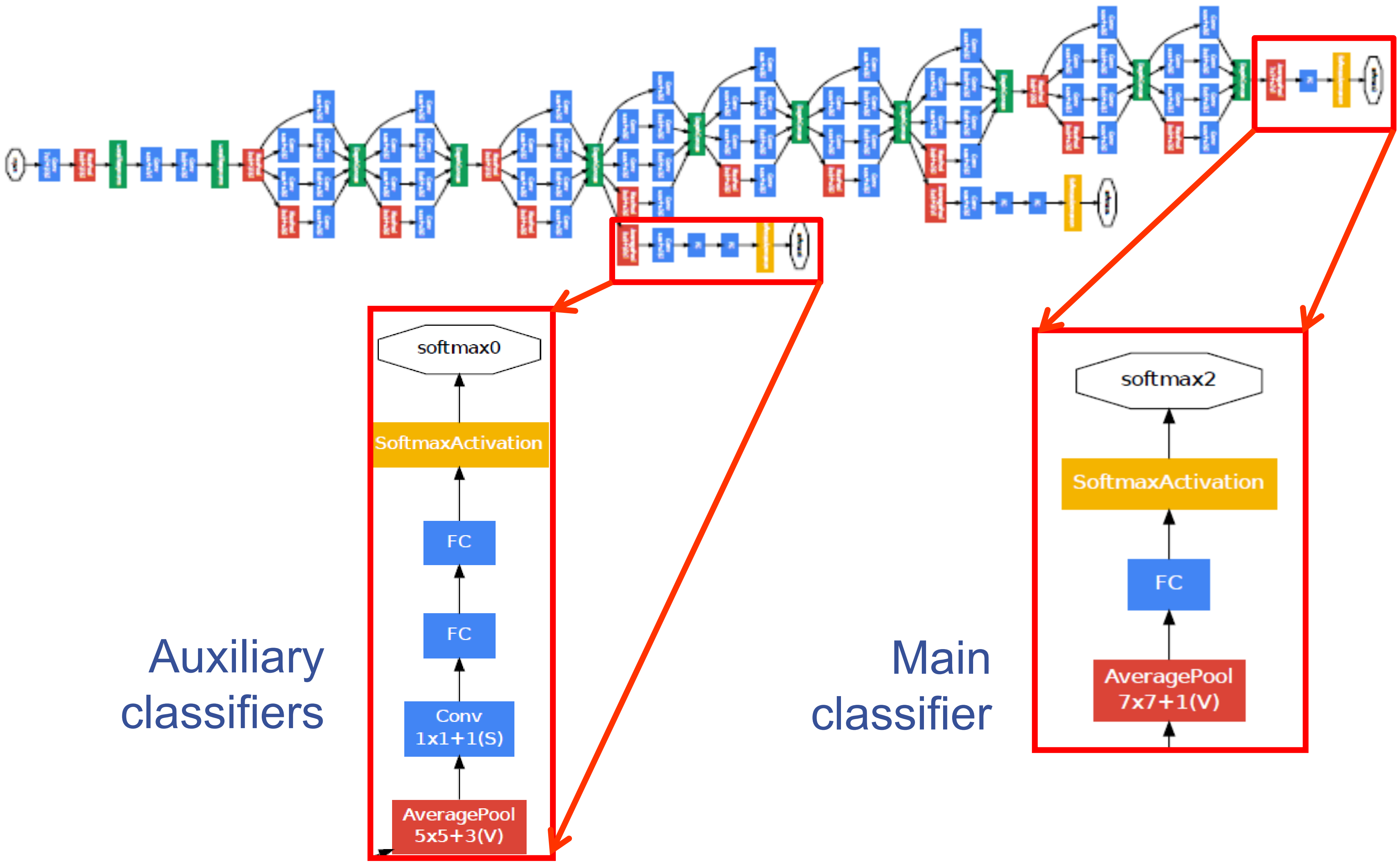


Convolution
Pooling
Softmax
Other

GoogLeNet (2014)



GoogLeNet (2014)



Recap AlexNet: What’s next?

How to improve AlexNet architecture?

+++Deep?

+++Convolutional?

+++Fully connected?

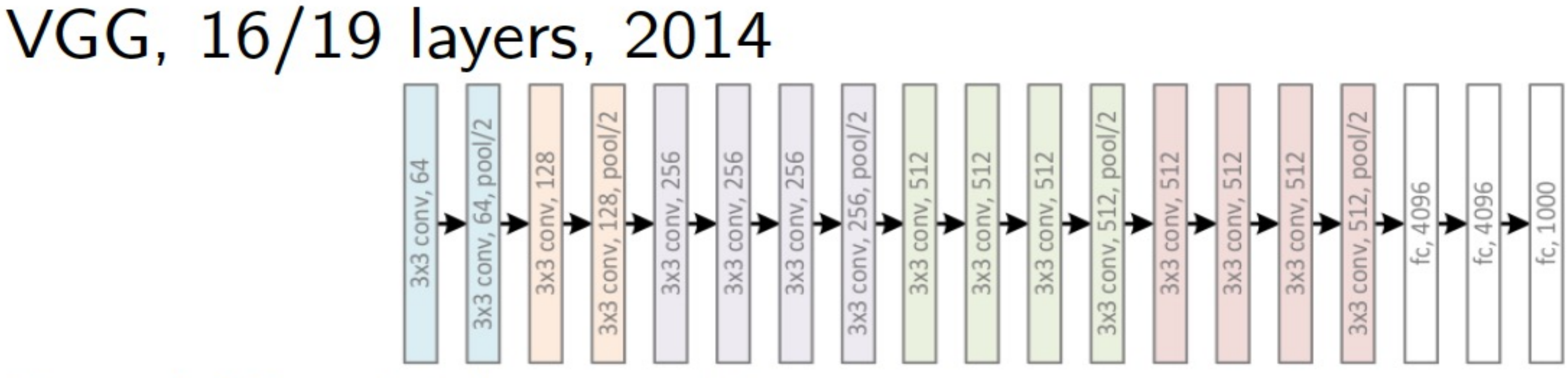
All?

⇒A lot of empirical studies

- ⇒Tuning various design parameters
- ⇒what really works?

⇒Winners: GoogLeNet, **VGG**, ResNet

VGG Net: Archi post-2012 revolution



K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, ICLR 2015

VGG Net

Basic Idea: Investigate the **effect of depth** in large scale image recognition

- **Fix other parameters** of architecture, and steadily increase depth

Fixed configuration:

- Convolutional Layers: from 8 to 16
- Fully Connected Layers: 3
- Stride: 1
- ReLu: Follow all hidden layers
- Max-Pooling: 2x2 window
- Padding: s/t spatial resolution is preserved
- #Convolutional filters: Starting from 64, double after each max-pooling layer until 512
- Filter sizes: 3x3 and 1x1

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

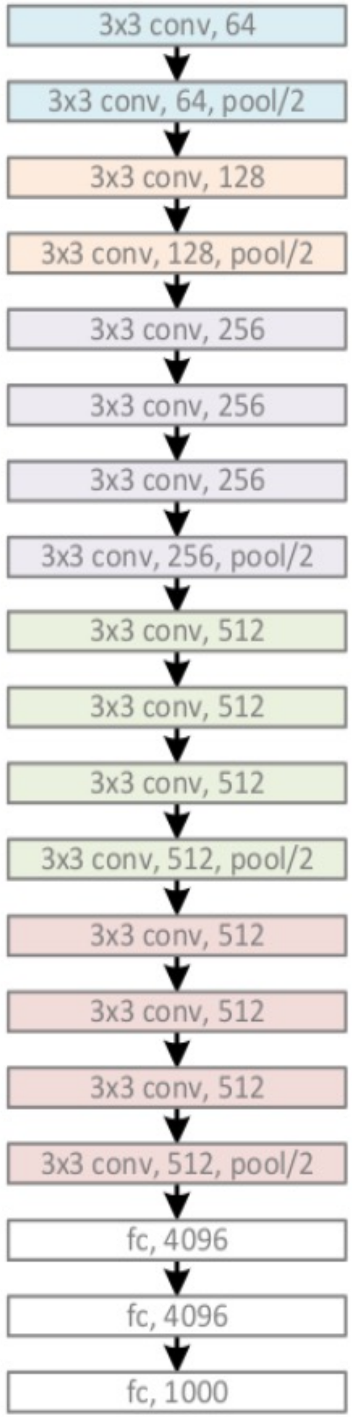


TABLE CREDIT:VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION, ICLR2015

VGG Net

Results:

- First place in localization (25.3% error), second in classification (7.3% error) in ILSVRC 2014 using ensemble of 7 networks
- Outperforms Szegedy et.al (GoogLeNet) in terms of single network classification accuracy (7.1% vs 7.9%)

Observations with VGG testing:

- Deepnets with small filters outperform shallow networks with large filters
  - Shallow version of B: 2 layers of 3x3 replaced with single 5x5 performs worse
- Classification error decreases with increases ConvNet depth
- Important to capture more spatial context (config D vs C)
- Error rate saturated at 19 layers
- Scale jittering at training helps capturing multiscale statistics and leads to better performance

Recap AlexNet: What’s next?

How to improve AlexNet architecture?

+++Deep?

+++Convolutional?

+++Fully connected?

All?

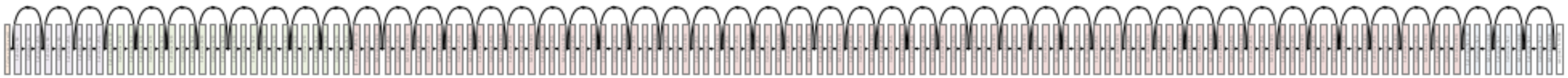
⇒A lot of empirical studies

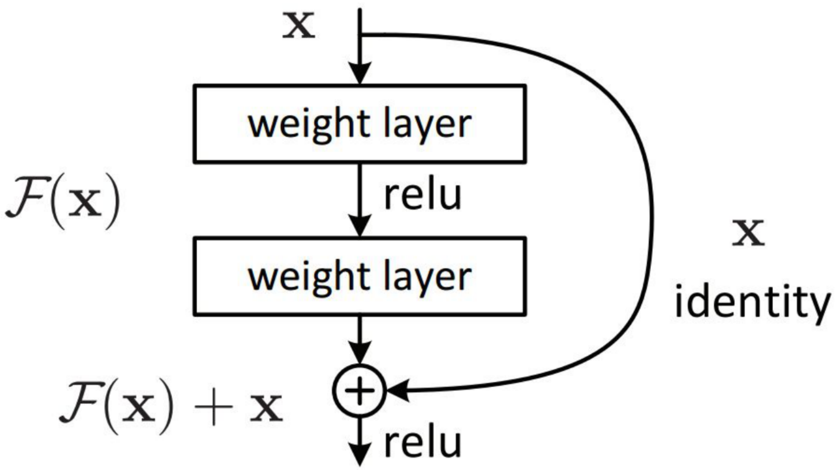
- ⇒Tuning various design parameters
- ⇒what really works?

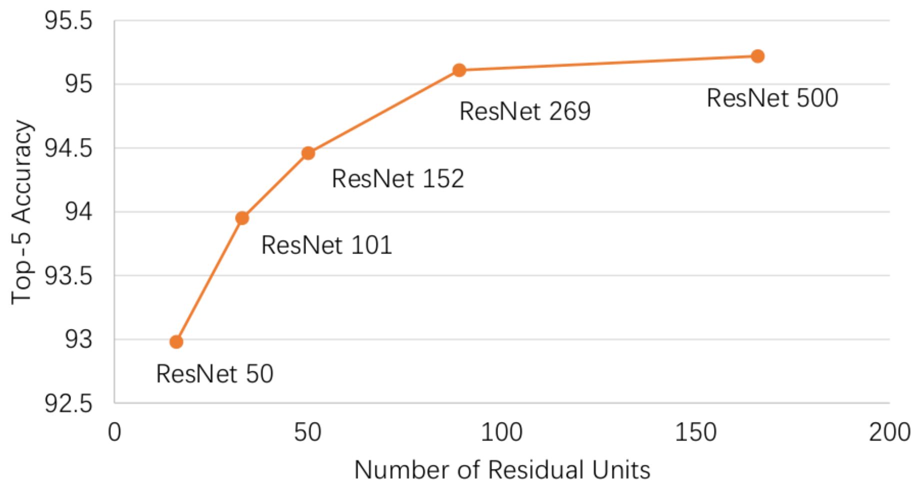
⇒Winners: GoogLeNet, VGG, **ResNet**

Deep ConvNets for image classification

- **ResNet** 152 layers, 60M parameters





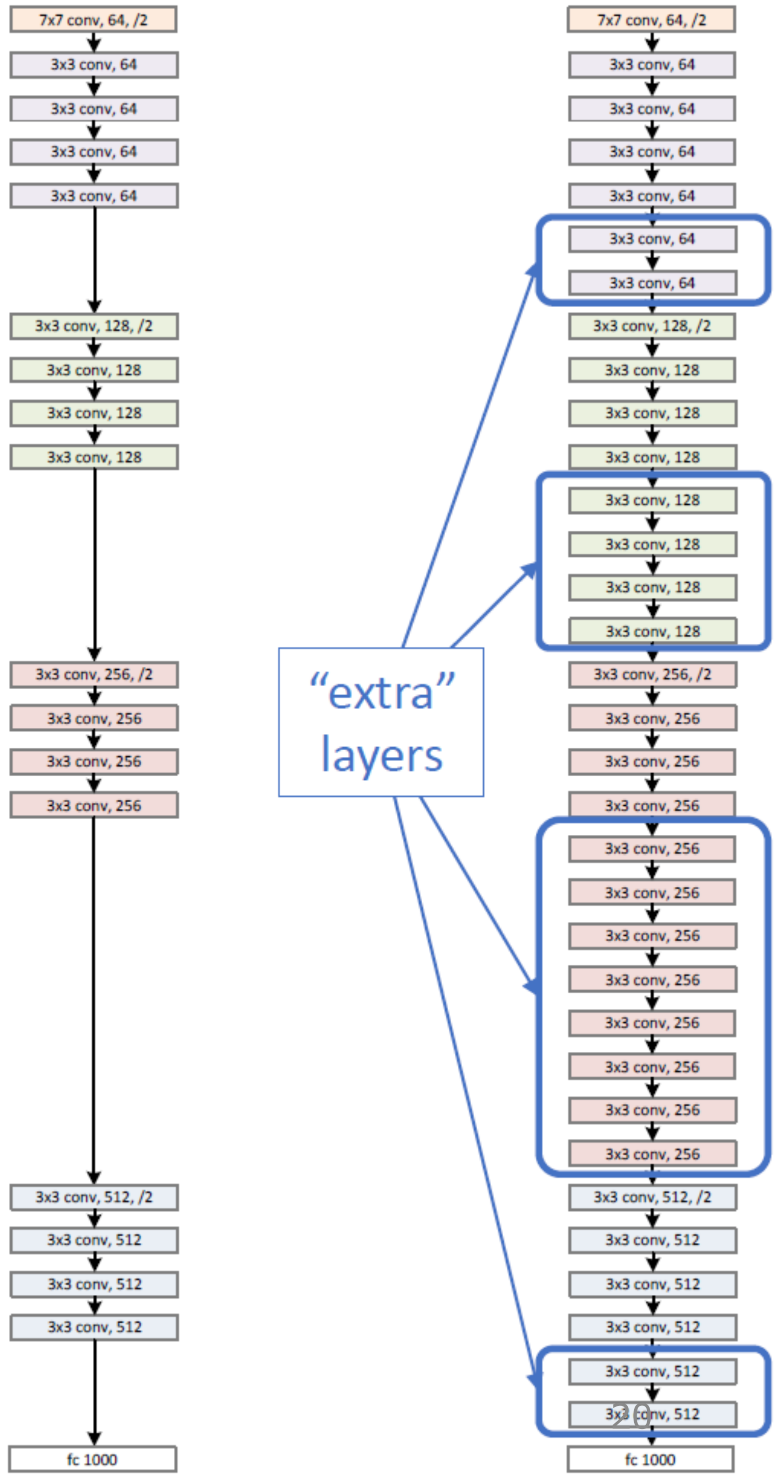


Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun
Deep Residual Learning for Image Recognition.
In *CVPR*, 2016.

Deeper VGG:

Deeper networks maintain the tendency of results

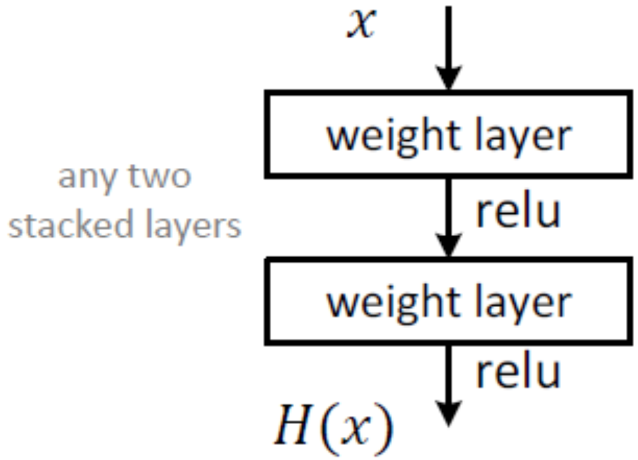
- Features in same level will be almost same
- An amount of changes is fixed
- Adding layers make smaller differences
- Optimal mappings closer to an identity



Residual Network

Plain block

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



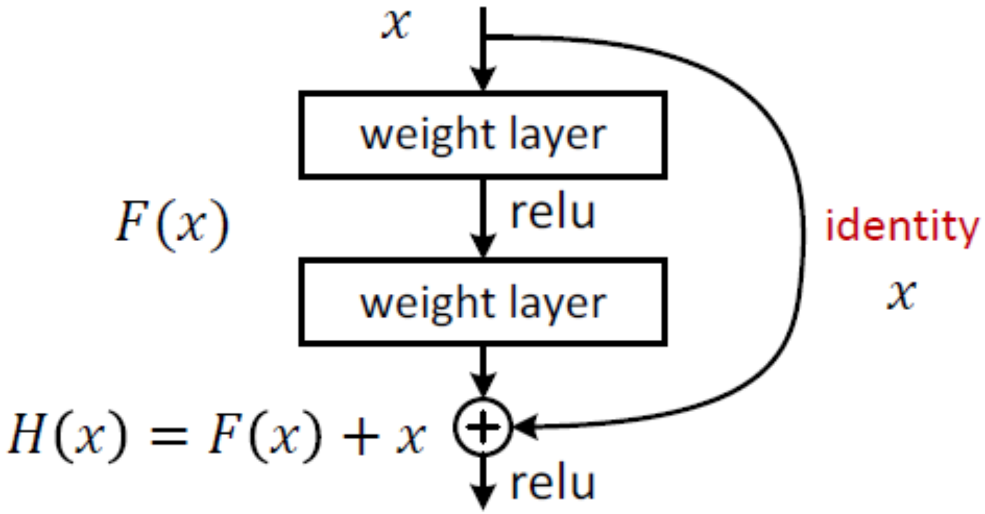
Residual Network

Residual block

If identity were optimal,
easy to set weights as 0

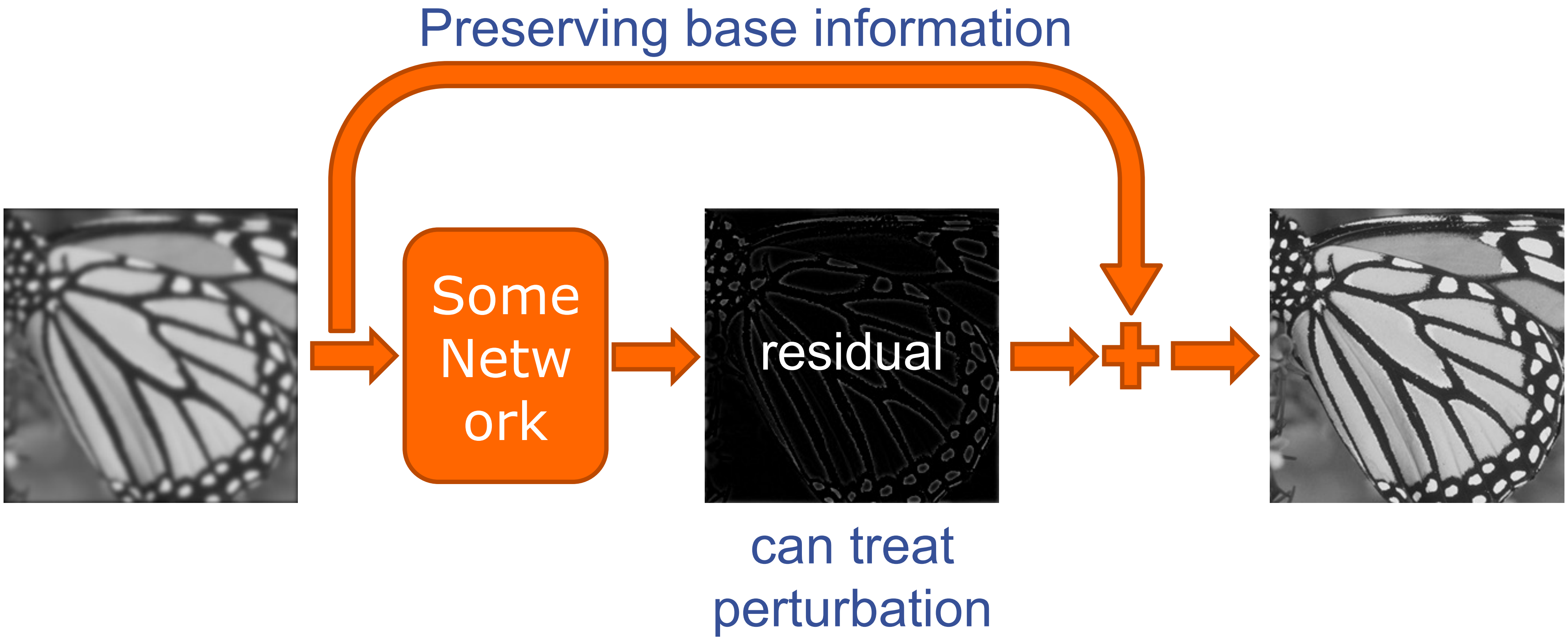
If optimal mapping is
closer to identity, easier to
find small fluctuations

-> Appropriate for treating perturbation as keeping a base information



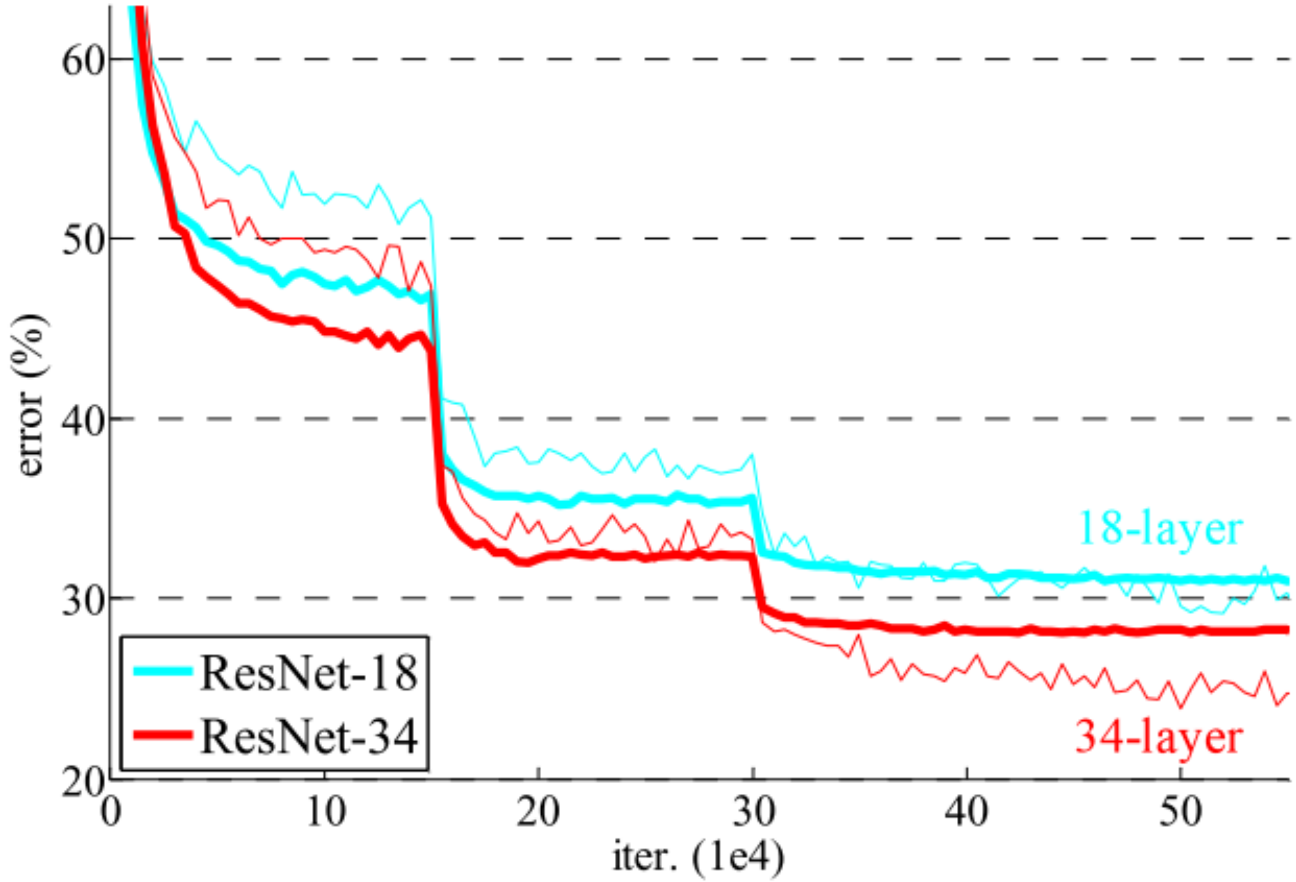
Residual Network

- Difference between an original image and a changed image



Residual Network

Deeper ResNets have lower training error



Conclusion

- ResNet: currently the best ConvNet archi for large scale image classification
- Not yet consensus about the design of the Net, Neural Architecture Search
- Fully Convolutional Net (FCN) very interesting option