DENSELY CONNECTED CONVOLUTIONAL NETWORKS

Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger

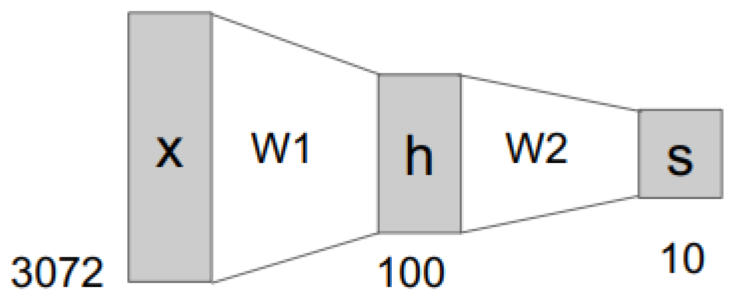
Tsinghua University

Raja Haseeb

Contents

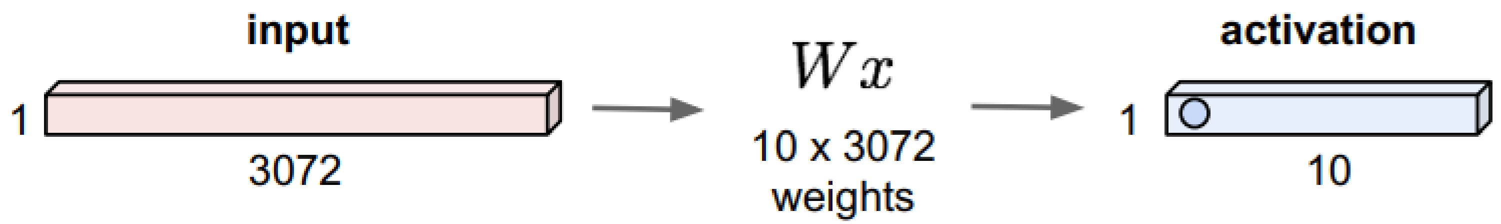
Neural Networks

- Linear score function: $f = Wx$
- 2-layer neural network: $f = W_2 \max(0, W_1 x)$



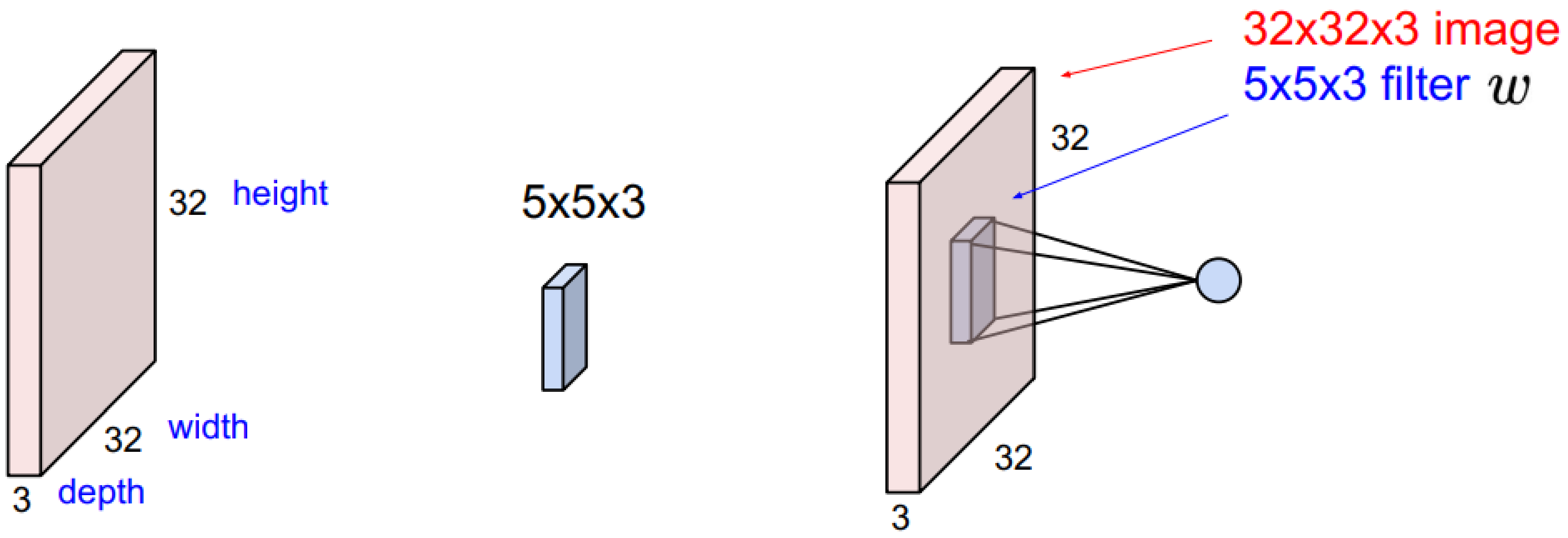
Fully Connected Layer

- 32 * 32 * 3 image => stretch to 3072 * 1



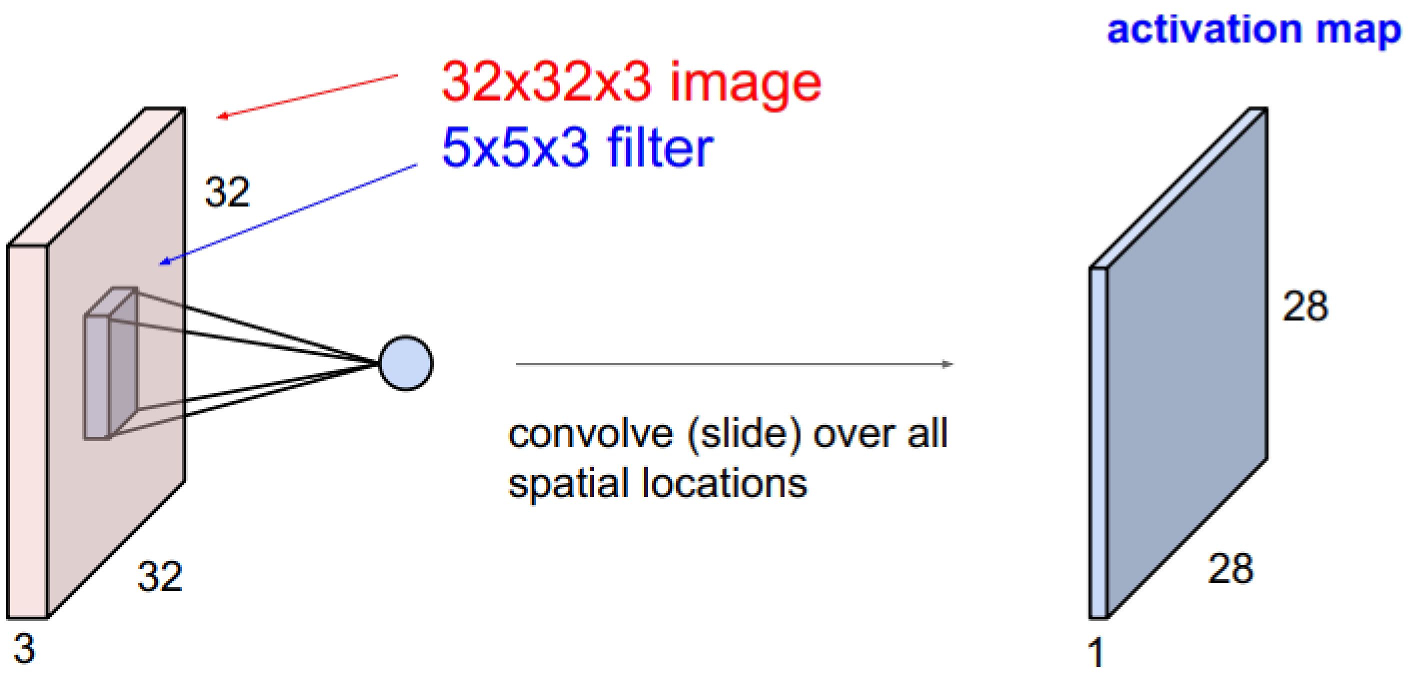
Convolutional Layer

- 32 * 32 * 3 image => preserve spatial structure

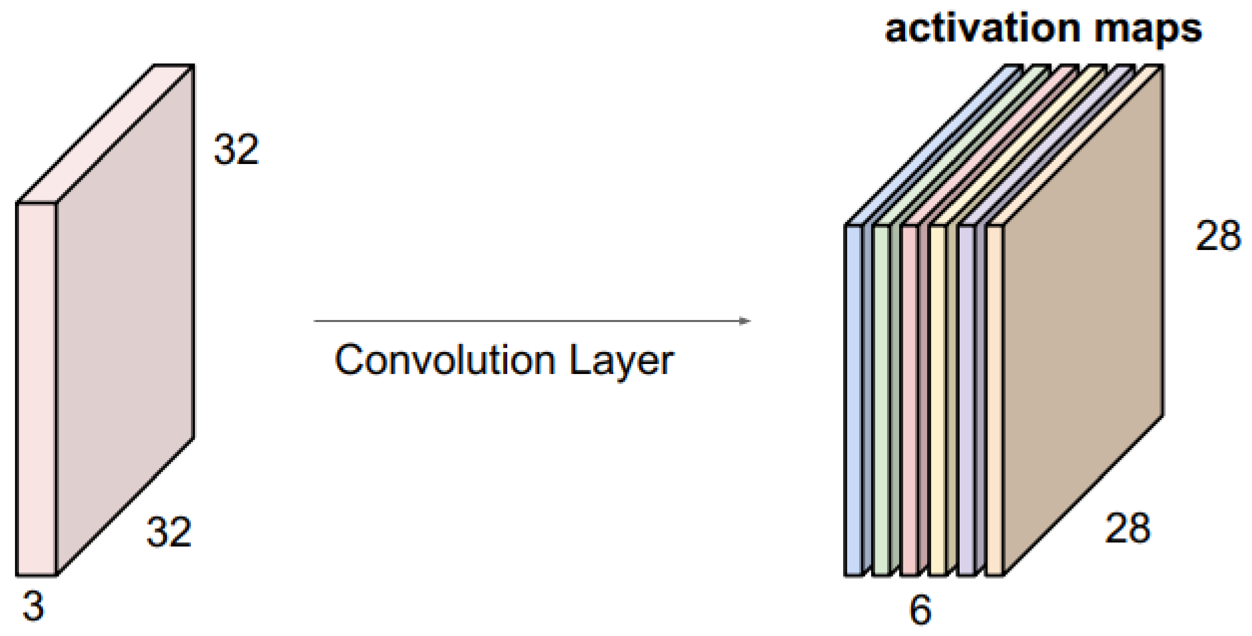


- Filter is convolved with the image i.e. *slide over image spatially, computing dot products*

Convolutional Layer



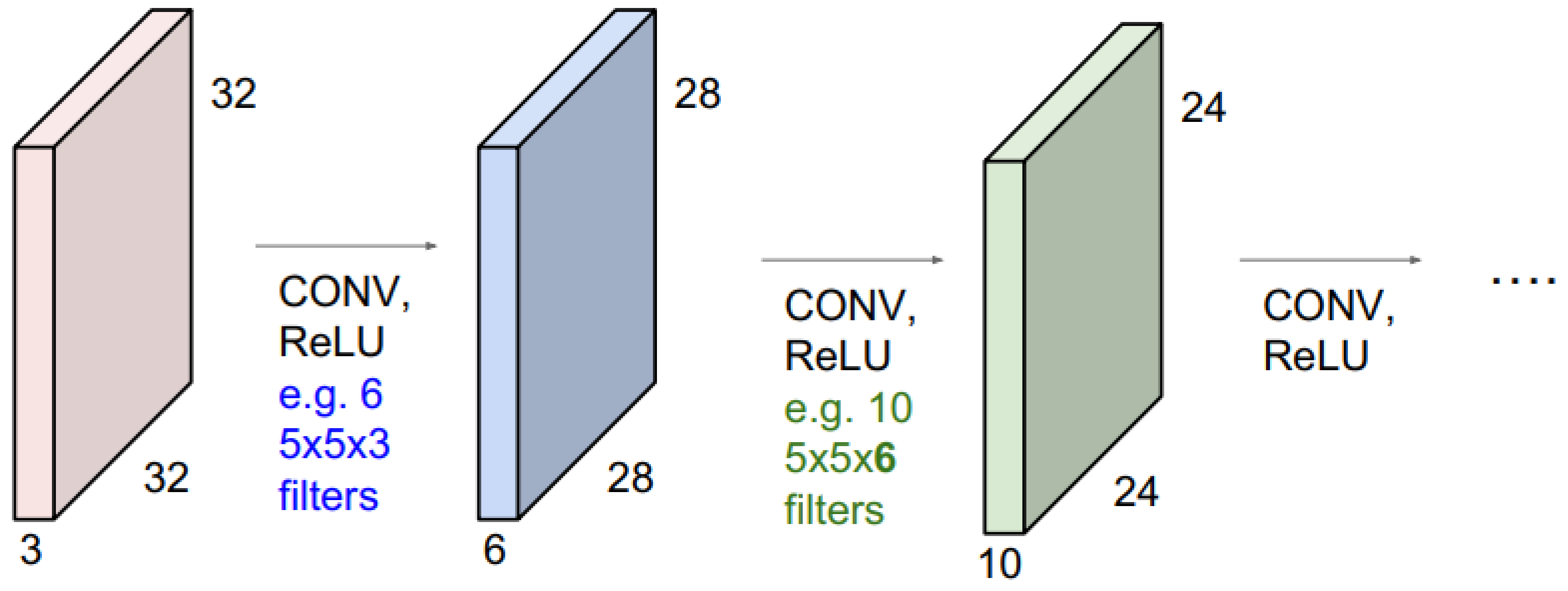
Convolutional Layer



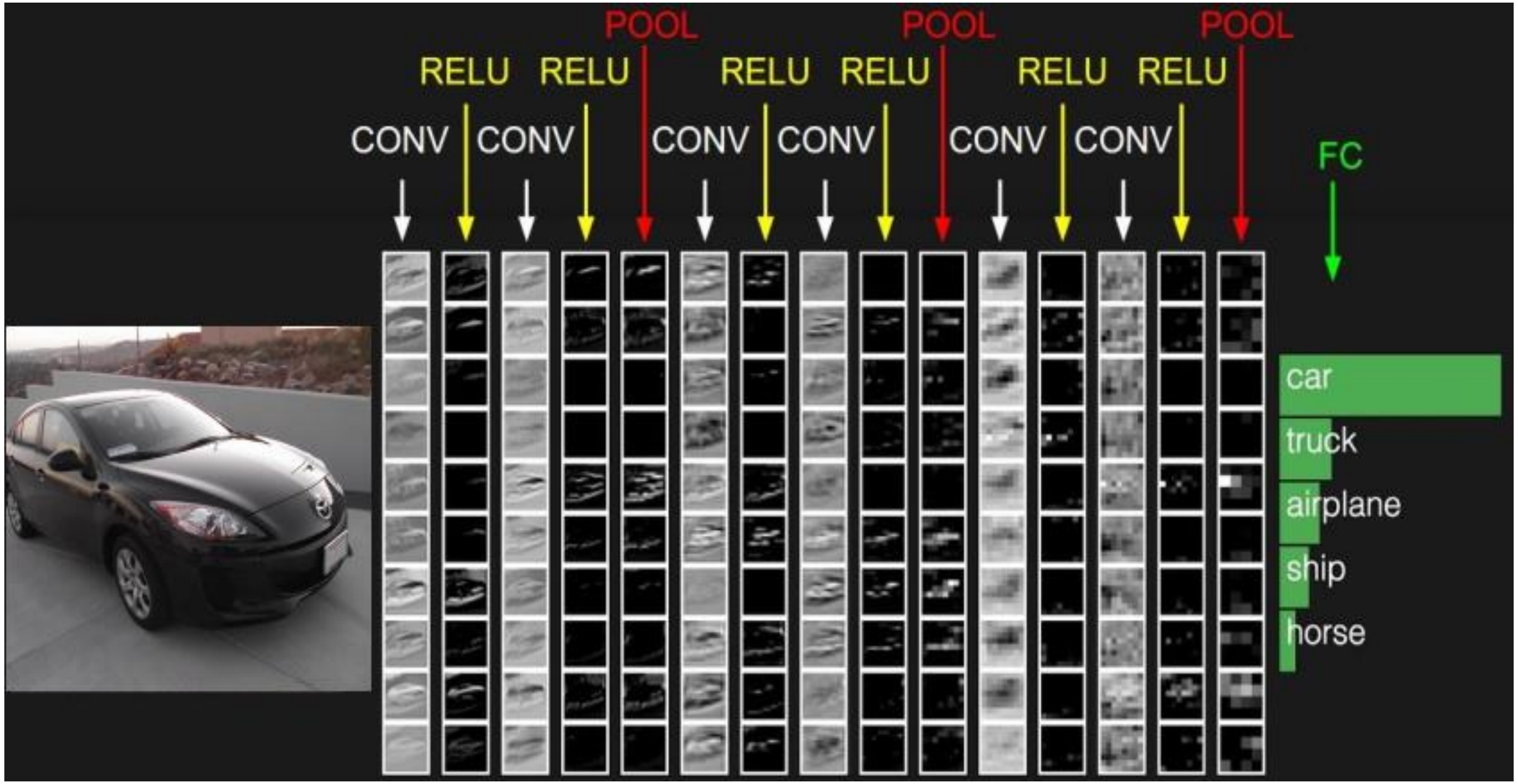
CONVNET

ConvNet

- ConvNet is a sequence of Convolution Layers, interspersed with activation functions

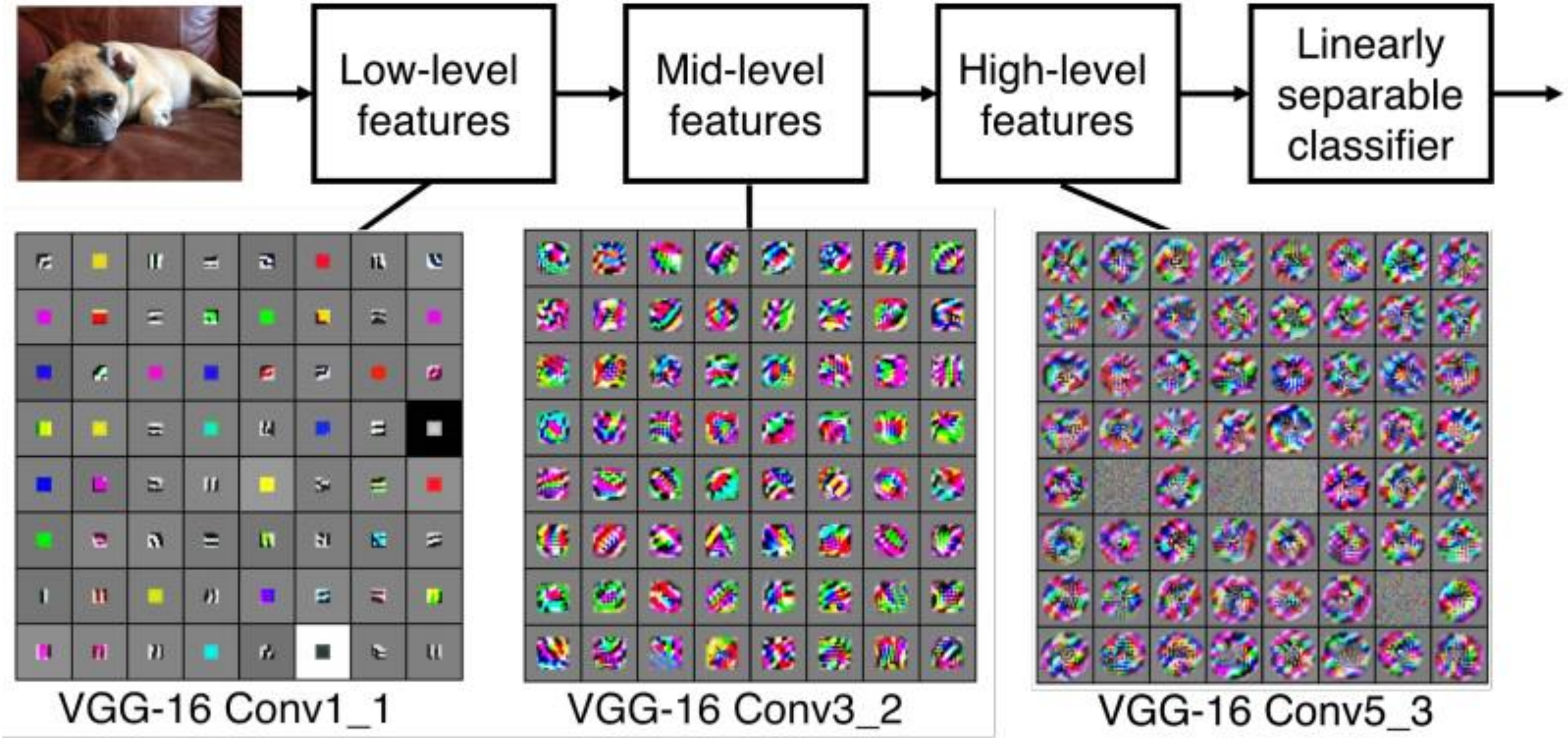


ConvNet



POOL
POOL
POOL
RELU
RELU
RELU
RELU
RELU
RELU
CONV
CONV
CONV
CONV
CONV
CONV
FC
car
truck
airplane
ship
horse

ConvNet

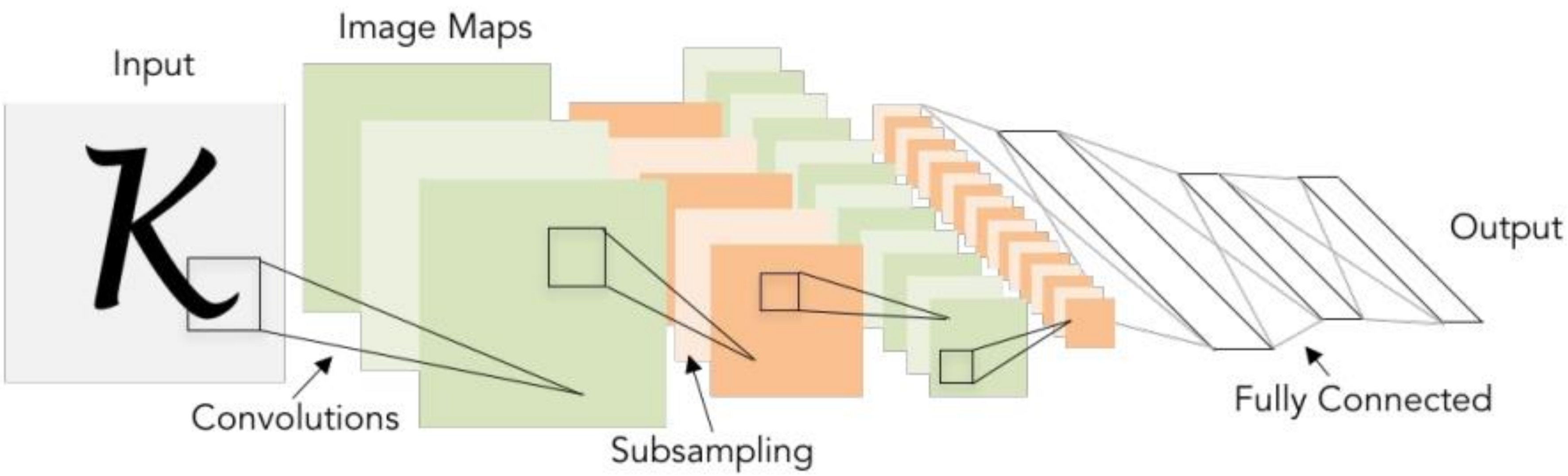


Low-level features
Mid-level features
High-level features
Linearly separable classifier
VGG-16 Conv1_1
VGG-16 Conv3_2
VGG-16 Conv5_3

HISTORY

Review: LeNeT-5

- In 1989, Alex Weibel introduced Time Delay Neural Networks (TDNN) which is considered first CNN
- LeCun et al., 1998
- He actually was working on them since 1980s



- Conv filters were 5x5, applied at stride 1
  Subsampling (Pooling) layers were 2x2 applied at stride 2
  i.e. architecture is [CONV-POOL-CONV-POOL-FC-FC]

Review: The Gap

- In the years from 1998 to 2010 neural network were in incubation
- People considered them complex and hard to optimize
- Some researchers slowly progressed
- More data and resources was available with the rise of technology

Review: The Gap

"Ask anyone in machine learning what kept neural network research alive and they will probably mention one or all of these three names: Geoffrey Hinton, fellow Canadian Yoshua Bengio and Yann LeCun, of Facebook and New York University."

"It was the worst possible time," says Bengio, a professor at the Université de Montréal and co-director of the CIFAR program since it was renewed last year. "Everyone else was doing something different. Somehow, Geoff convinced them."

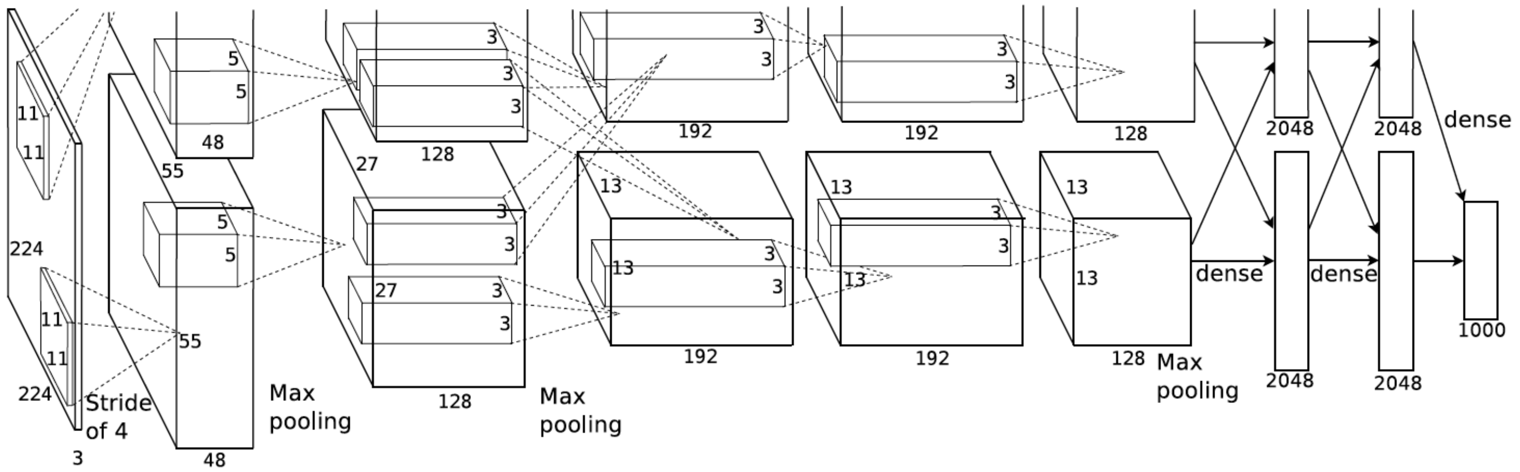
"We should give (CIFAR) a lot of credit for making that gamble."

CIFAR "had a huge impact in forming a community around deep learning," adds LeCun, the CIFAR program's other co-director. "We were outcast a little bit in the broader machine learning community: we couldn't get our papers published. This gave us a place where we could exchange ideas."

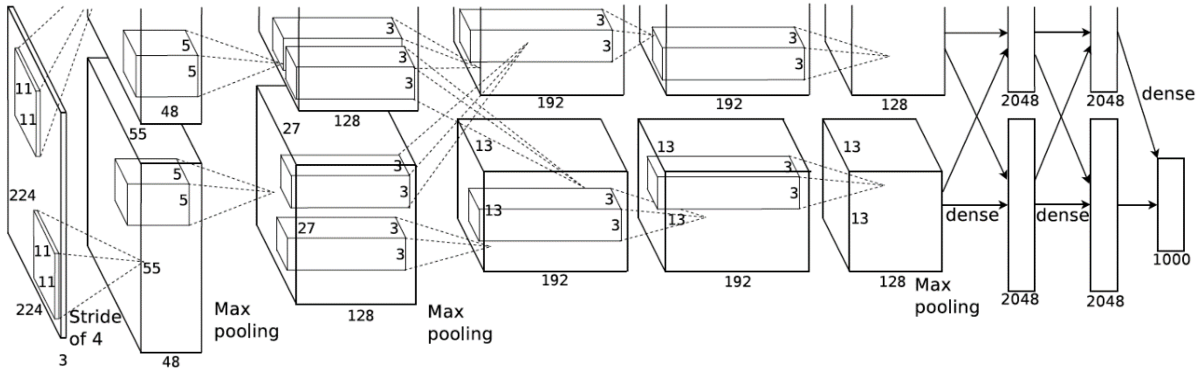
Review: AlexNet

- *Krizhevsky et al., 2012*
- Architecture (8 layers):

CONV1
MAX POOL1
NORM1
CONV2
MAX POOL2
NORM2
CONV3
CONV4
CONV5
Max POOL3
FC6
FC7
FC8



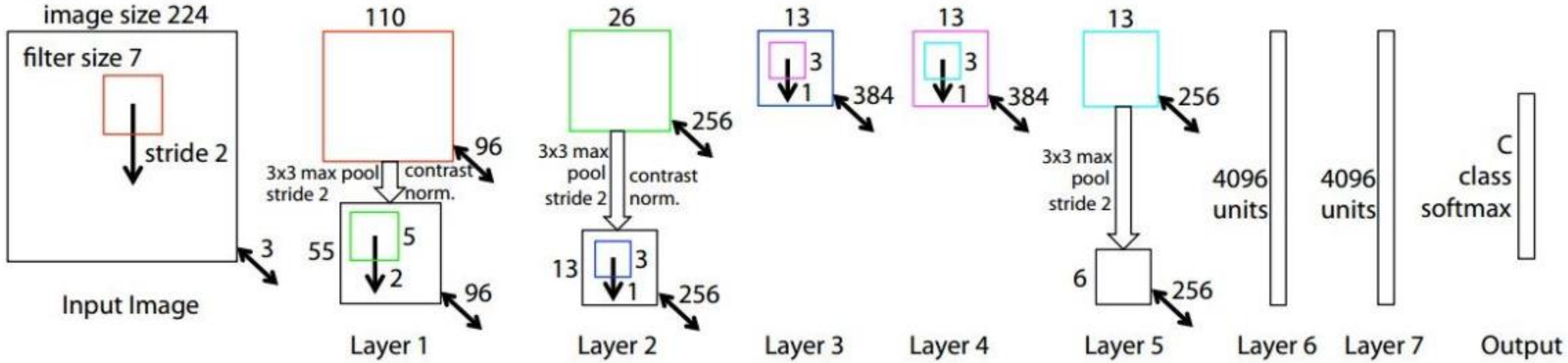
AlexNet



Full (simplified) AlexNet architecture:
[227x227x3] INPUT
[55x55x96] CONV1: 96 11x11 filters at stride 4, pad 0
[27x27x96] MAX POOL1: 3x3 filters at stride 2
[27x27x96] NORM1: Normalization layer
[27x27x256] CONV2: 256 5x5 filters at stride 1, pad 2
[13x13x256] MAX POOL2: 3x3 filters at stride 2
[13x13x256] NORM2: Normalization layer
[13x13x384] CONV3: 384 3x3 filters at stride 1, pad 1
[13x13x384] CONV4: 384 3x3 filters at stride 1, pad 1
[13x13x256] CONV5: 256 3x3 filters at stride 1, pad 1
[6x6x256] MAX POOL3: 3x3 filters at stride 2
[4096] FC6: 4096 neurons
[4096] FC7: 4096 neurons
[1000] FC8: 1000 neurons (class scores)

Review: ZFNet

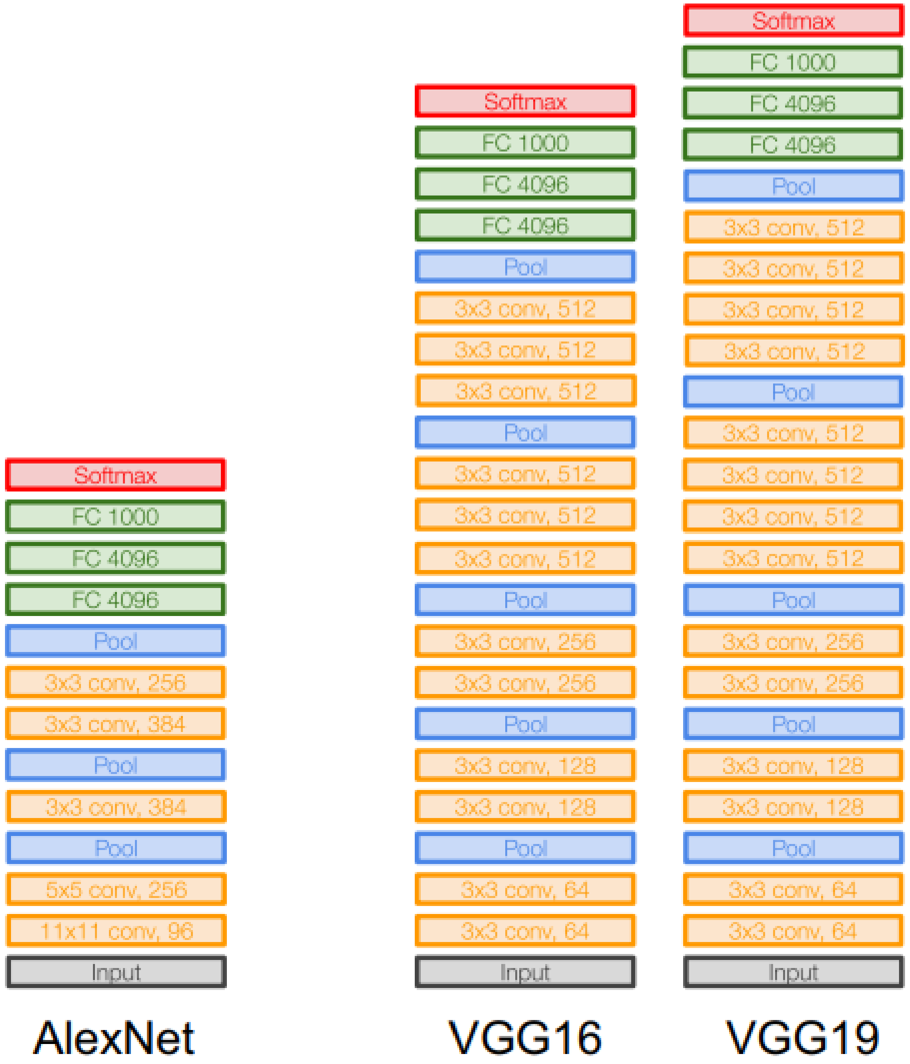
- Improved hyperparameters of AlexNet
- Architecture:



- ImageNet top 5 error reduced from 16.4% to 11.7%

Review: VGGNet

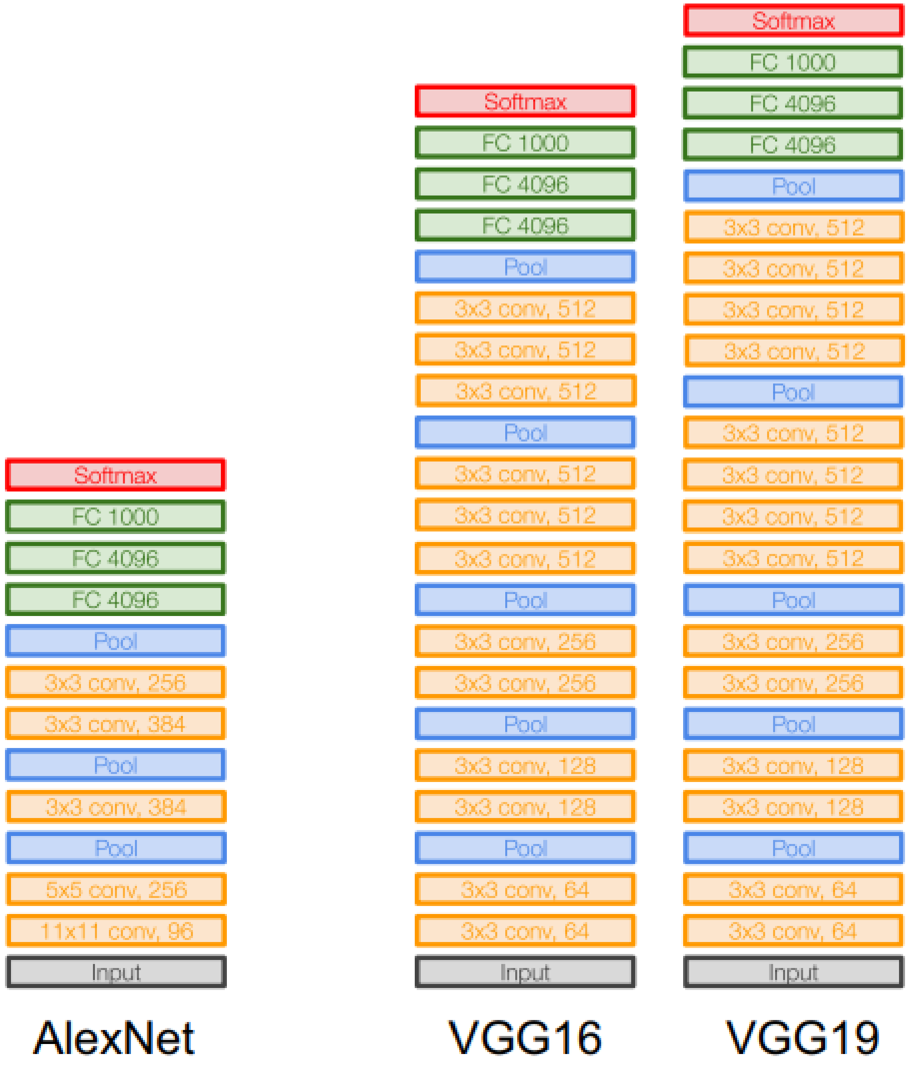
- *Karen Simonyan and Andrew Zisserman, 2014*
- Small filters, deeper networks
- 16 – 19 layers instead of 8 layers in AlexNet
- Only 3*3 convolutions
- Around 138M parameters
- 7.3% top 5 error in ILSVRC'14



Review: VGGNet

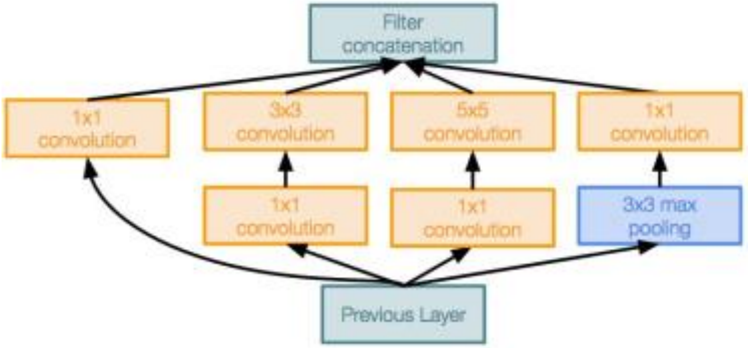
- *Why use small filters?*

Stack of three 3x3 conv (stride 1) layers has same effective receptive field as one 7x7 conv layer

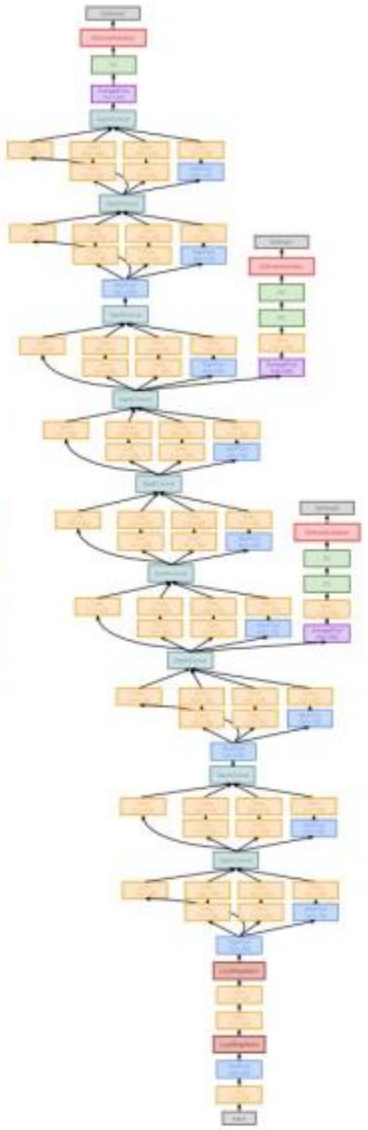


Review: GoogLeNet

- *Szegedy et al., 2014*
- Deeper networks with computation
- 22 layers
- Efficient inception modules
- No FC layers
- Only 5 million parameters
  (12x less than AlexNet)
- ILSVRC’14 classification winner
  (6.7% top 5 error)



Inception module

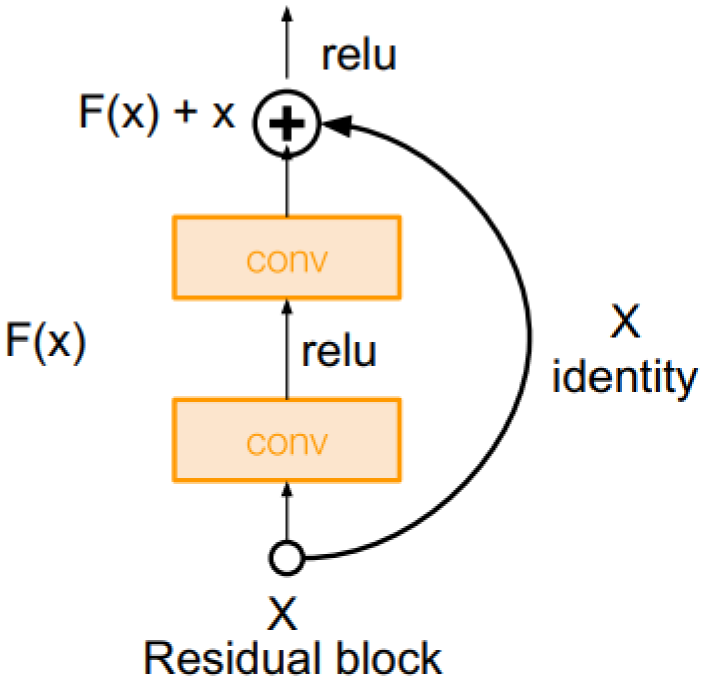


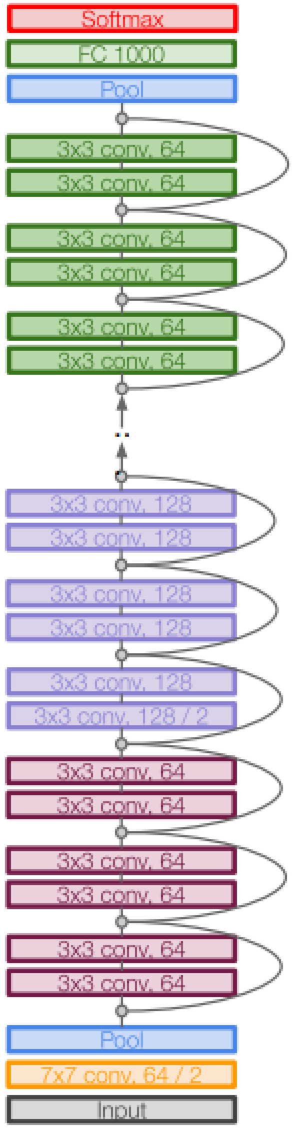
Issues in going deeper

- Computational power (Alexnet had to use use two GPUs)
- Vanishing gradient problem
- A small gradient means that the weights and biases of the initial layers will not be updated effectively with each training session.

Review: ResNet

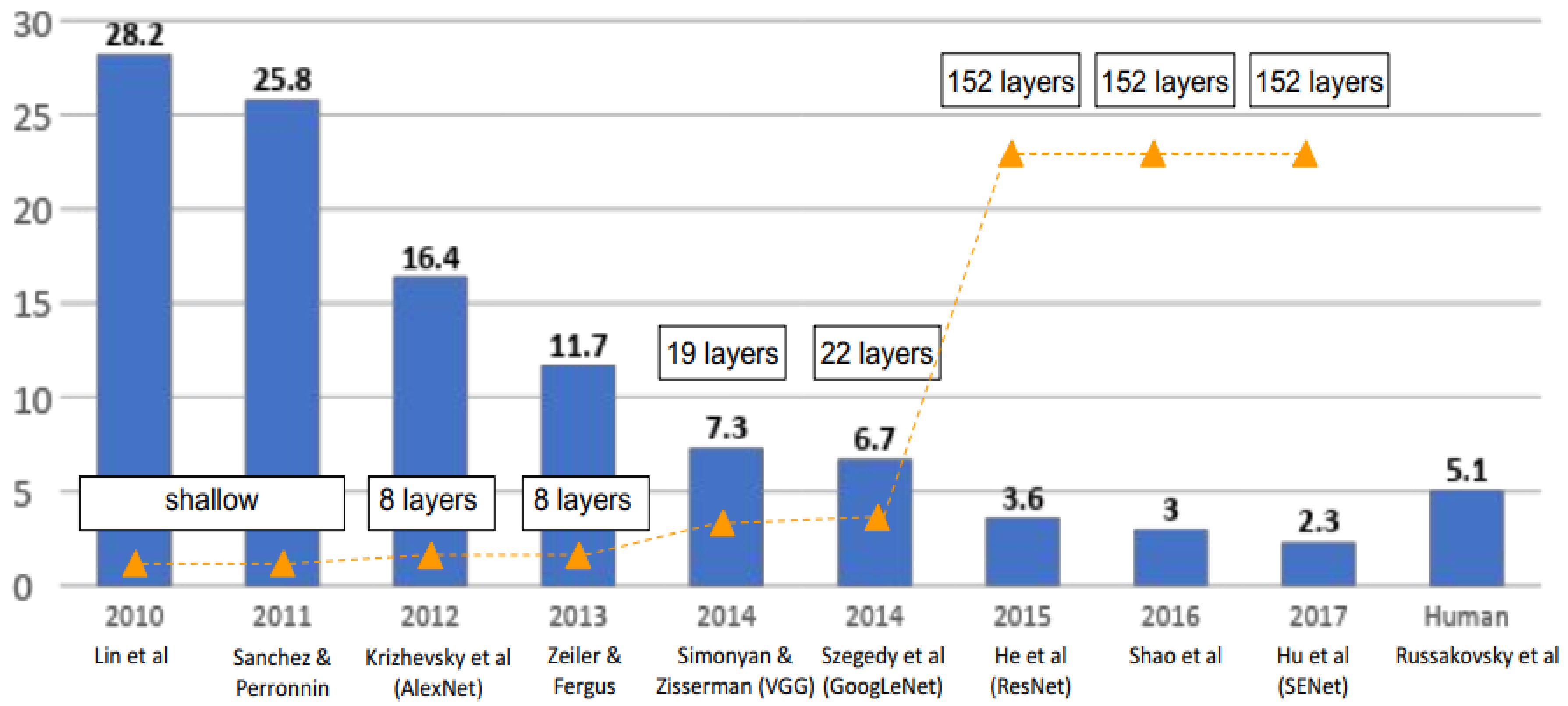
- *He et al., 2014*
- Very deep networks using residual connections
- 152 layer model for ImageNet
- Efficient inception modules
- No FC layers
- ResNet-50 has 23M parameters
- ILSVRC'15 classification winner (3.6% top 5 error)



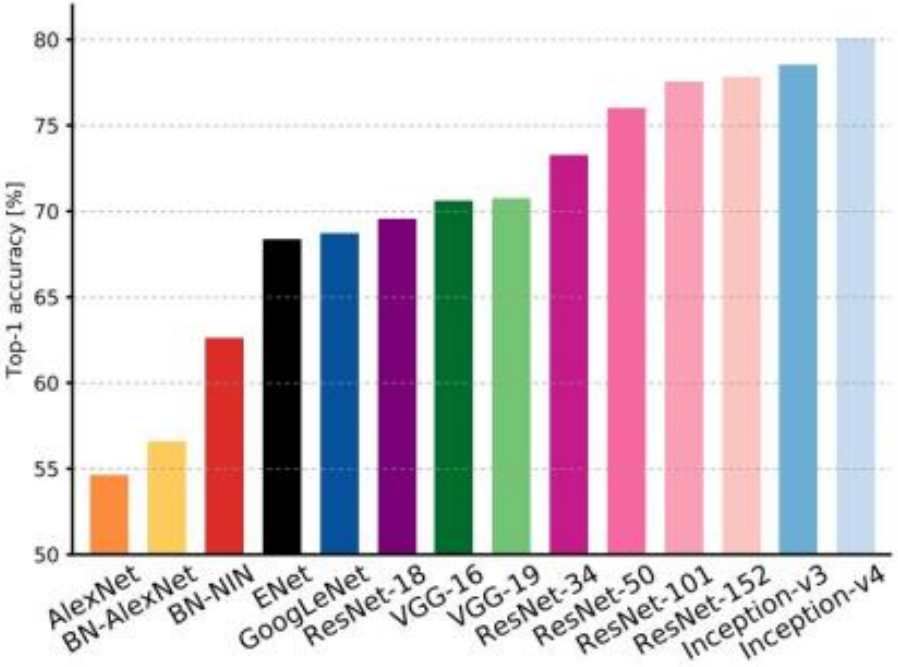


COMPARISON

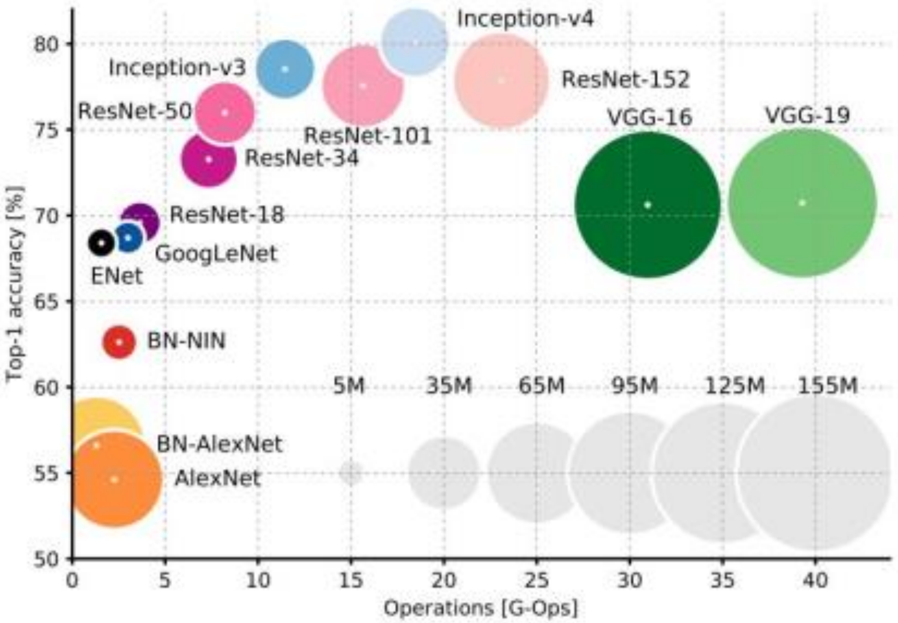
ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



Comparison



Top-1 accuracy [%]
80
75
70
65
60
55
50
AlexNet
BN-AlexNet
BN-NIN
ENet
GoogLeNet
ResNet-18
VGG-16
VGG-19
ResNet-34
ResNet-50
ResNet-101
ResNet-152
Inception-v3
Inception-v4

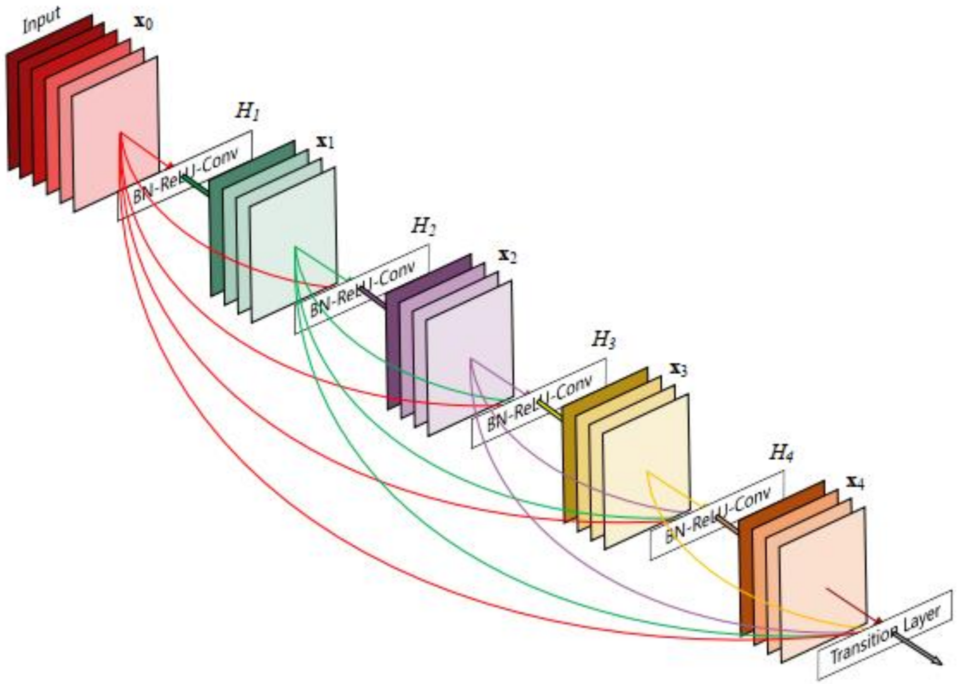


Top-1 accuracy [%]
Inception-v4
Inception-v3
ResNet-152
ResNet-50
VGG-16
VGG-19
ResNet-101
ResNet-34
ResNet-18
GoogLeNet
ENet
BN-NIN
5M
35M
65M
95M
125M
155M
BN-AlexNet
AlexNet
Operations [G-Ops]
0
5
10
15
20
25
30
35
40

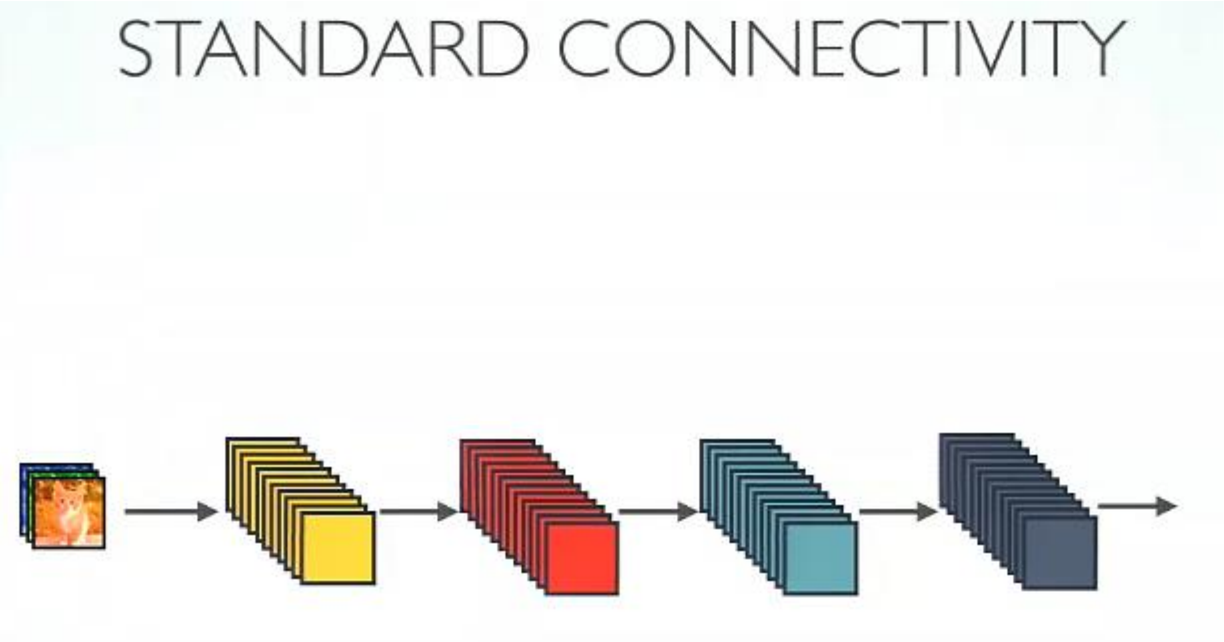
DenseNet

DenseNet

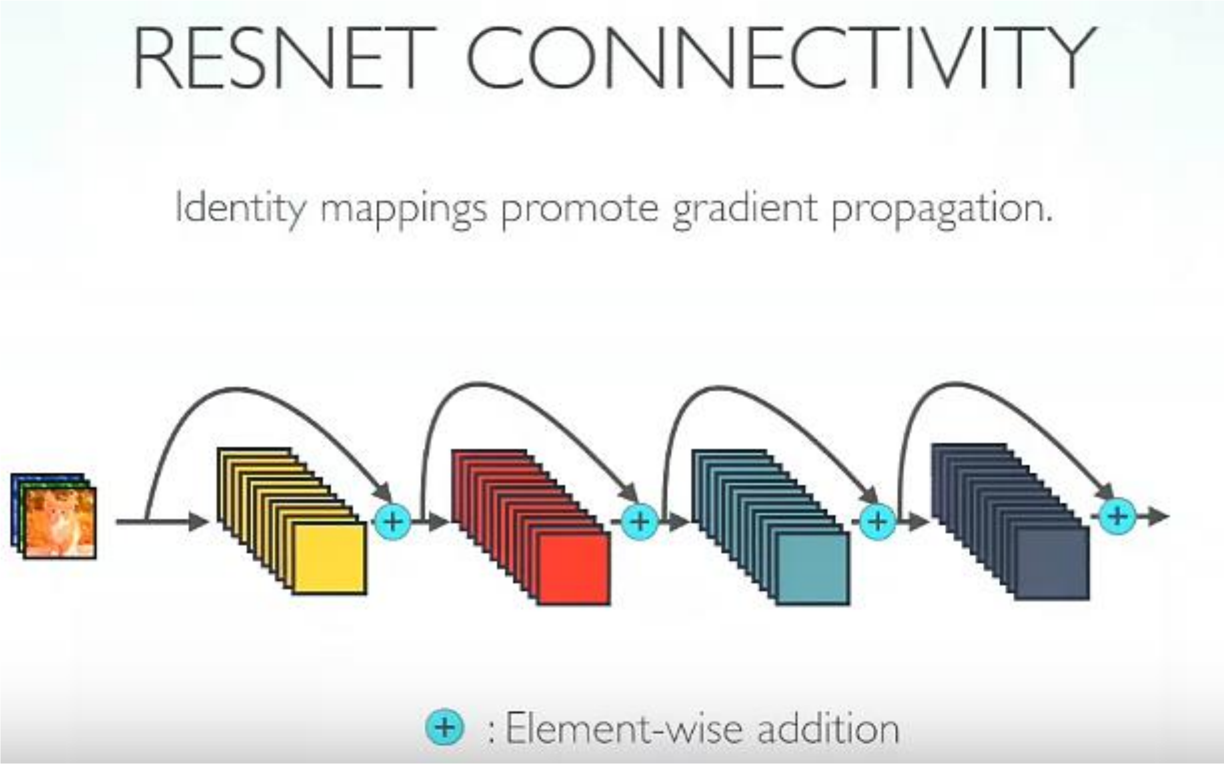
- *Gao Huang et al., 2018*



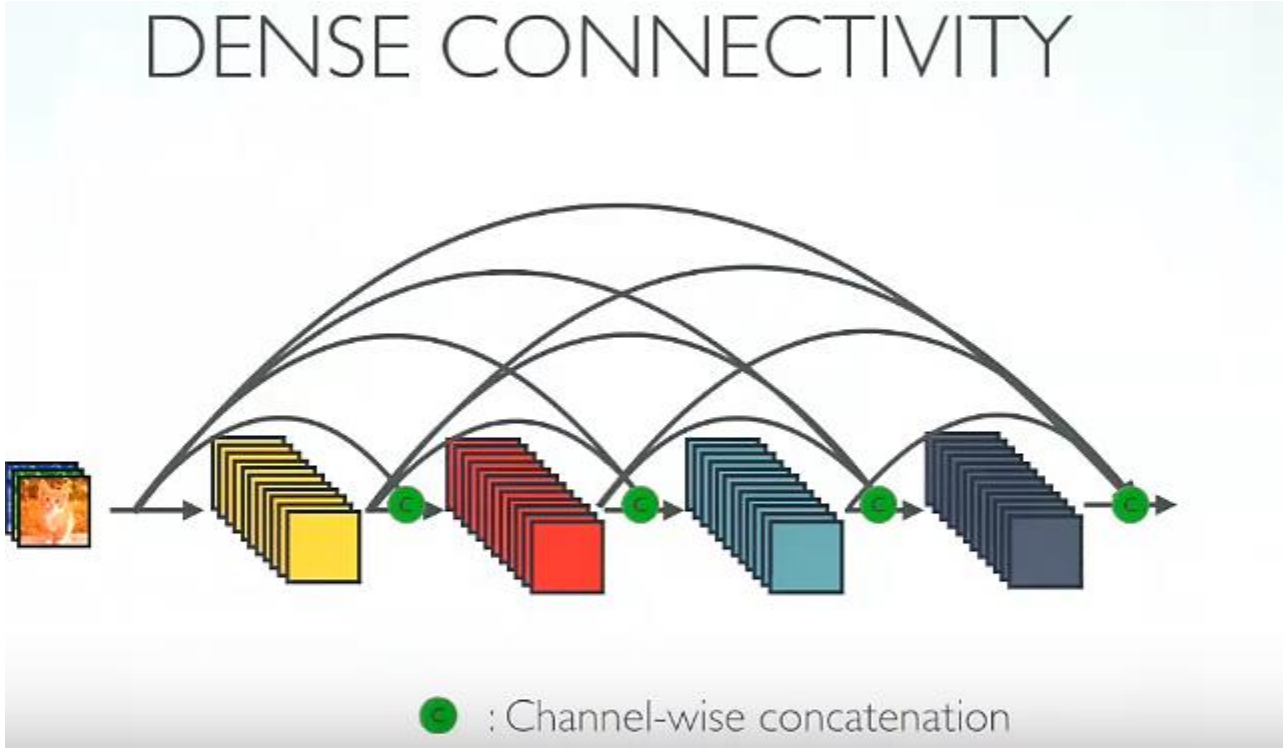
DenseNet



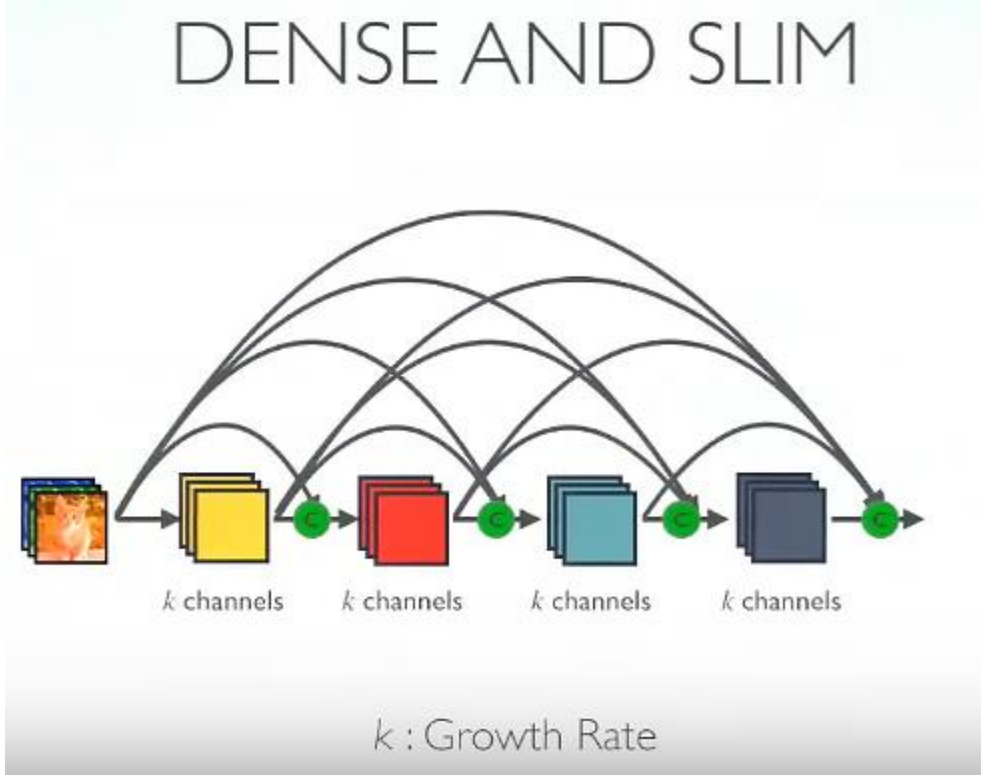
DenseNet


RESNET CONNECTIVITY
Identity mappings promote gradient propagation.
: Element-wise addition

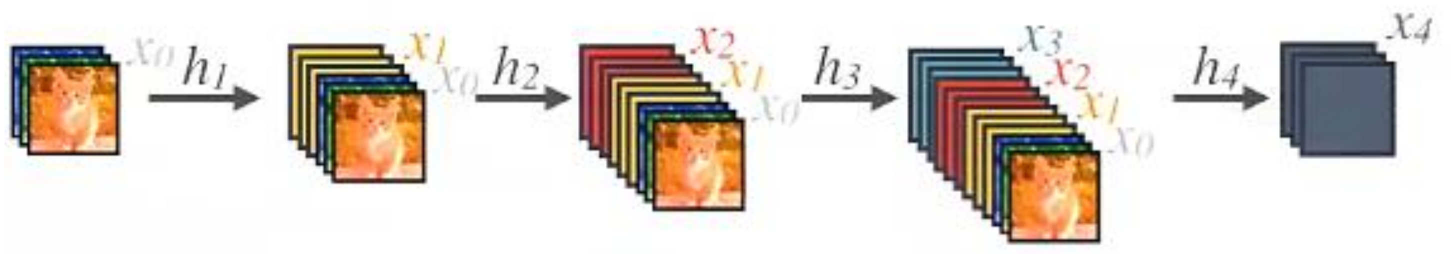
DenseNet



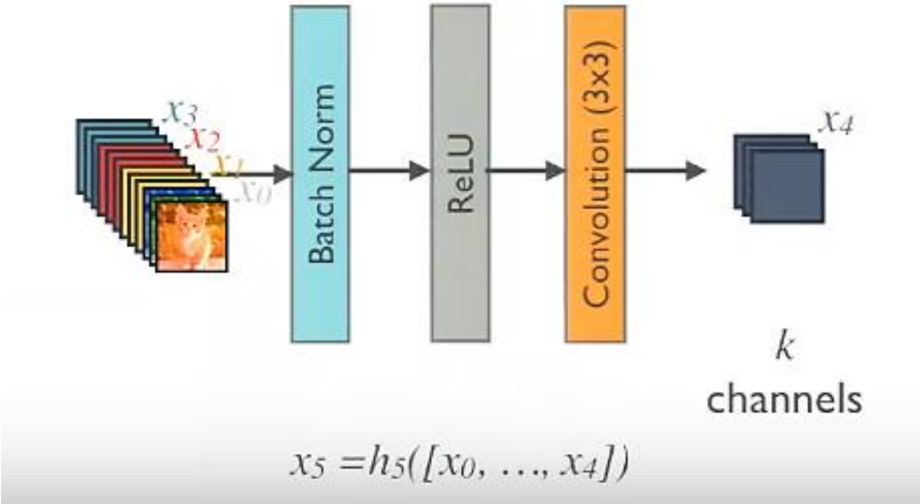
DenseNet


DENSE AND SLIM
k channels
k channels
k channels
k channels
k : Growth Rate

Forward Propagation

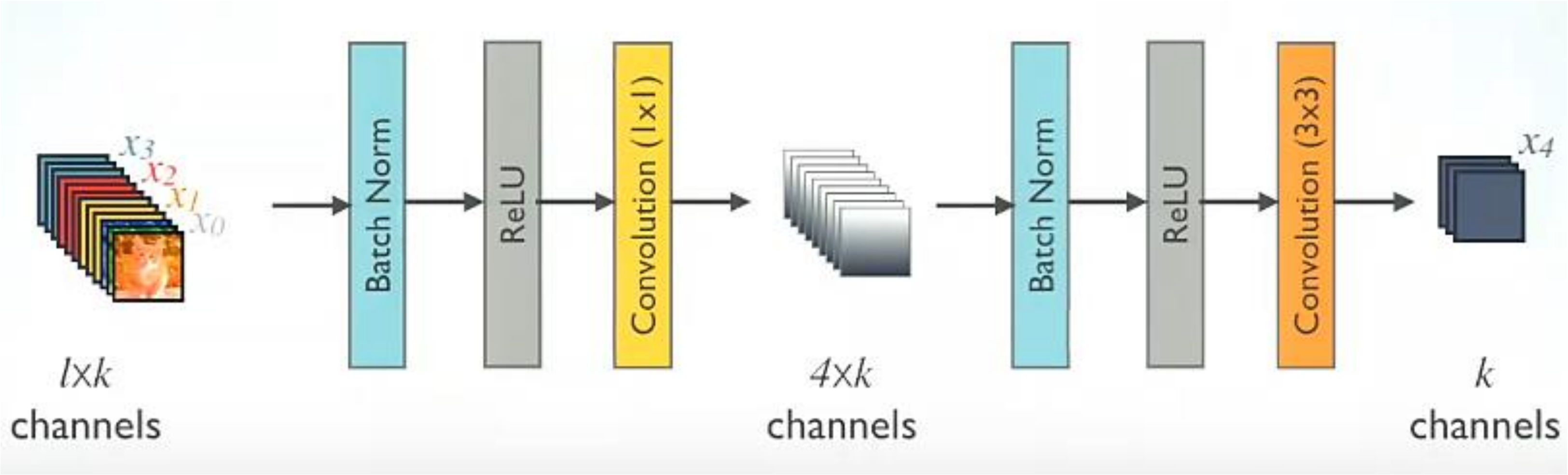


Composite layer in DenseNet



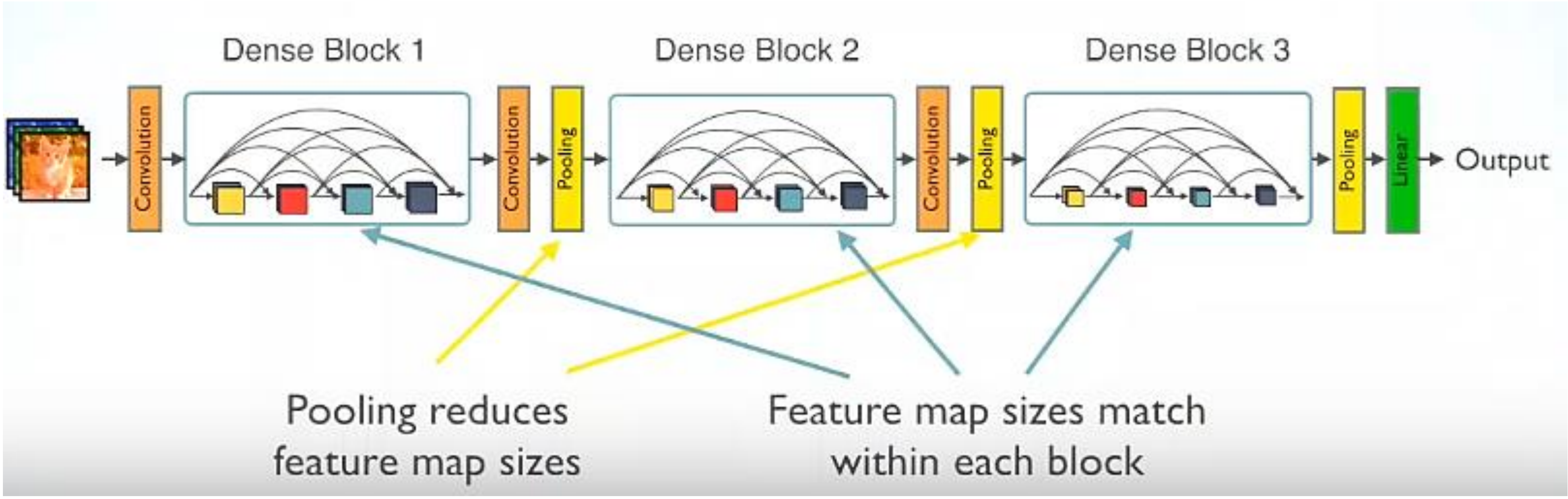
Composite layer in DenseNet with bottleneck

- High parameter and computational efficiency



DenseNet

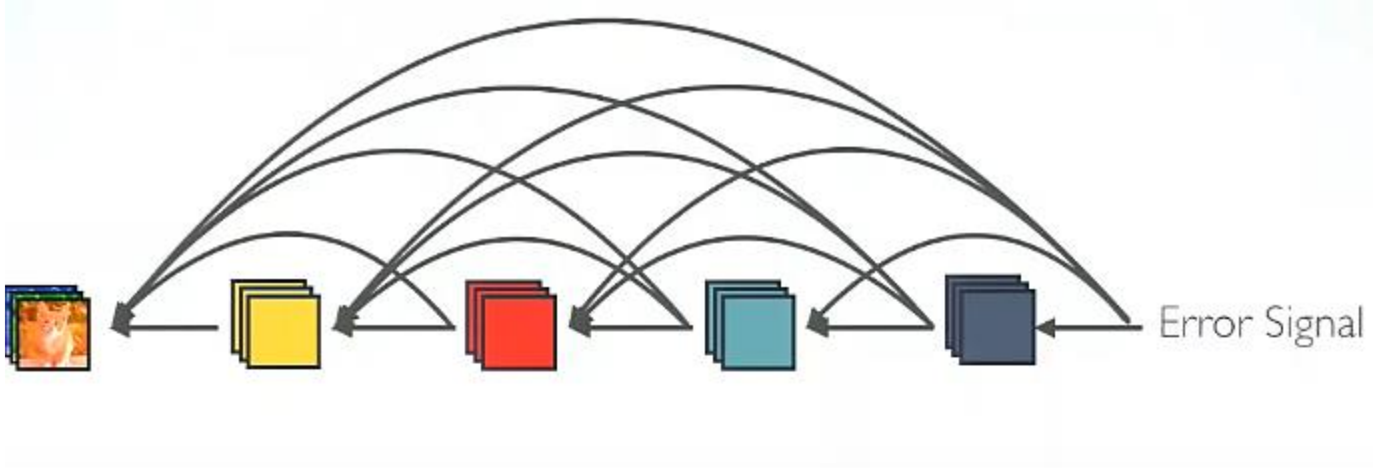
- A deep DenseNet with three dense blocks



ADVANTAGES OF DENSE CONNECTIVITY

ADVANTAGE 1: Strong Gradient Flow

- Implicit deep supervision

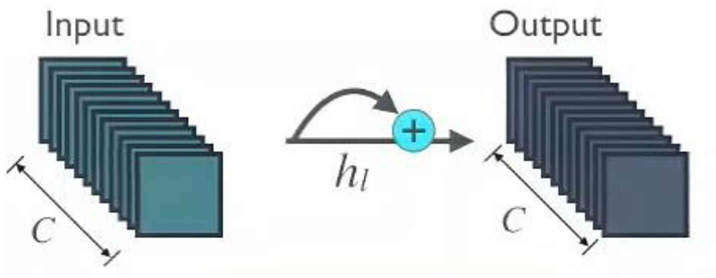


Deeply supervised net [Lee, Xie, Gallagher, Zhang, Tu] (2015)

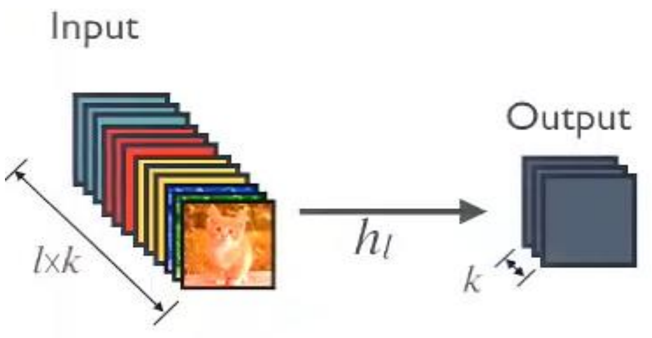
ADVANTAGE 2: Parameter and Computational Efficiency

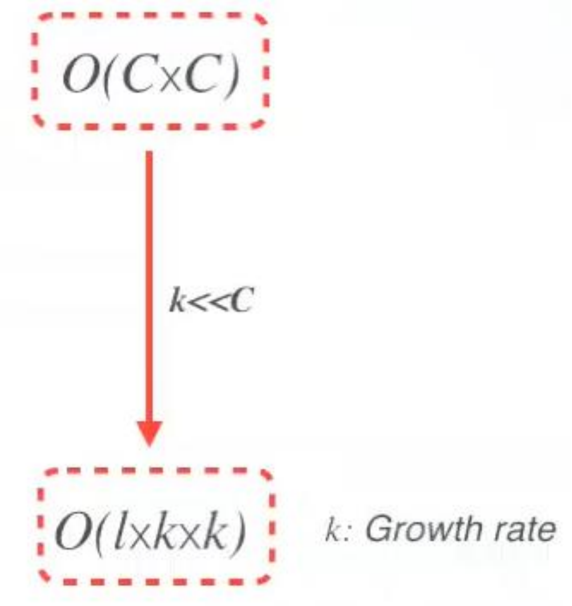
ResNet connectivity:

#parameter



DenseNet connectivity:



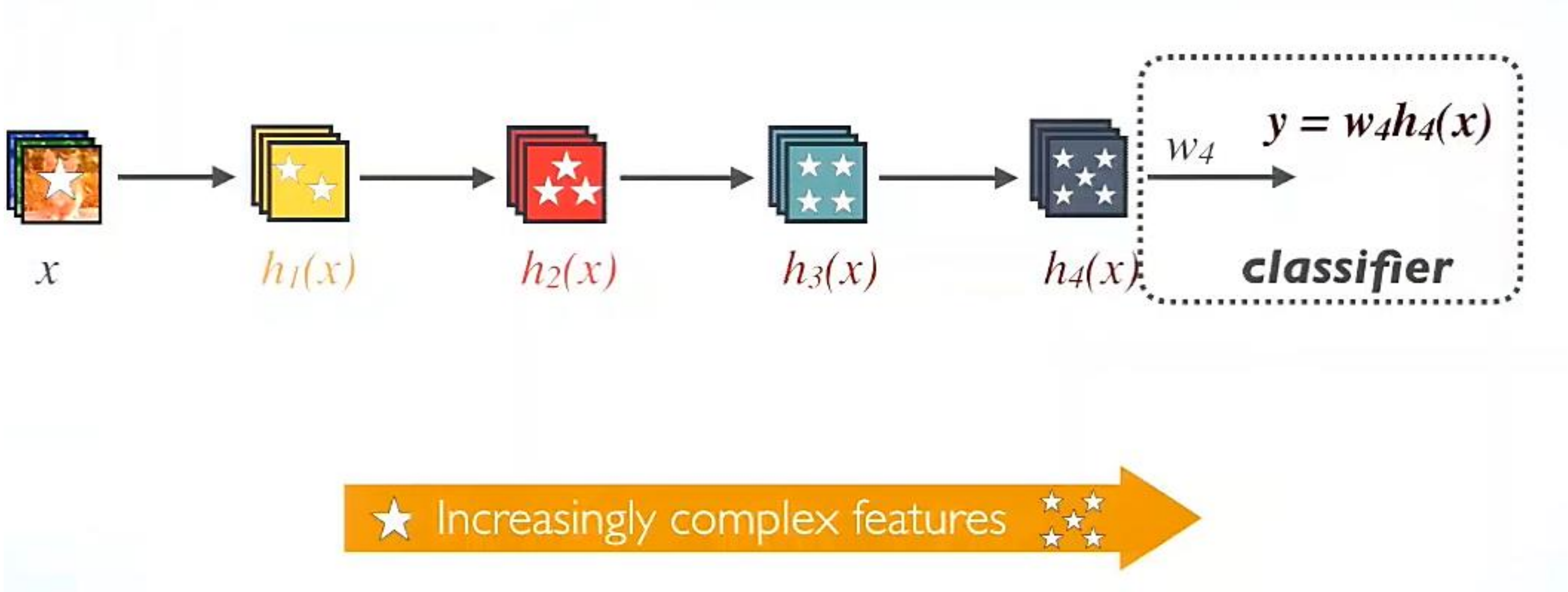


ADVANTAGE 3: Diverse and richer patterns

- Features from all the preceding layers are used

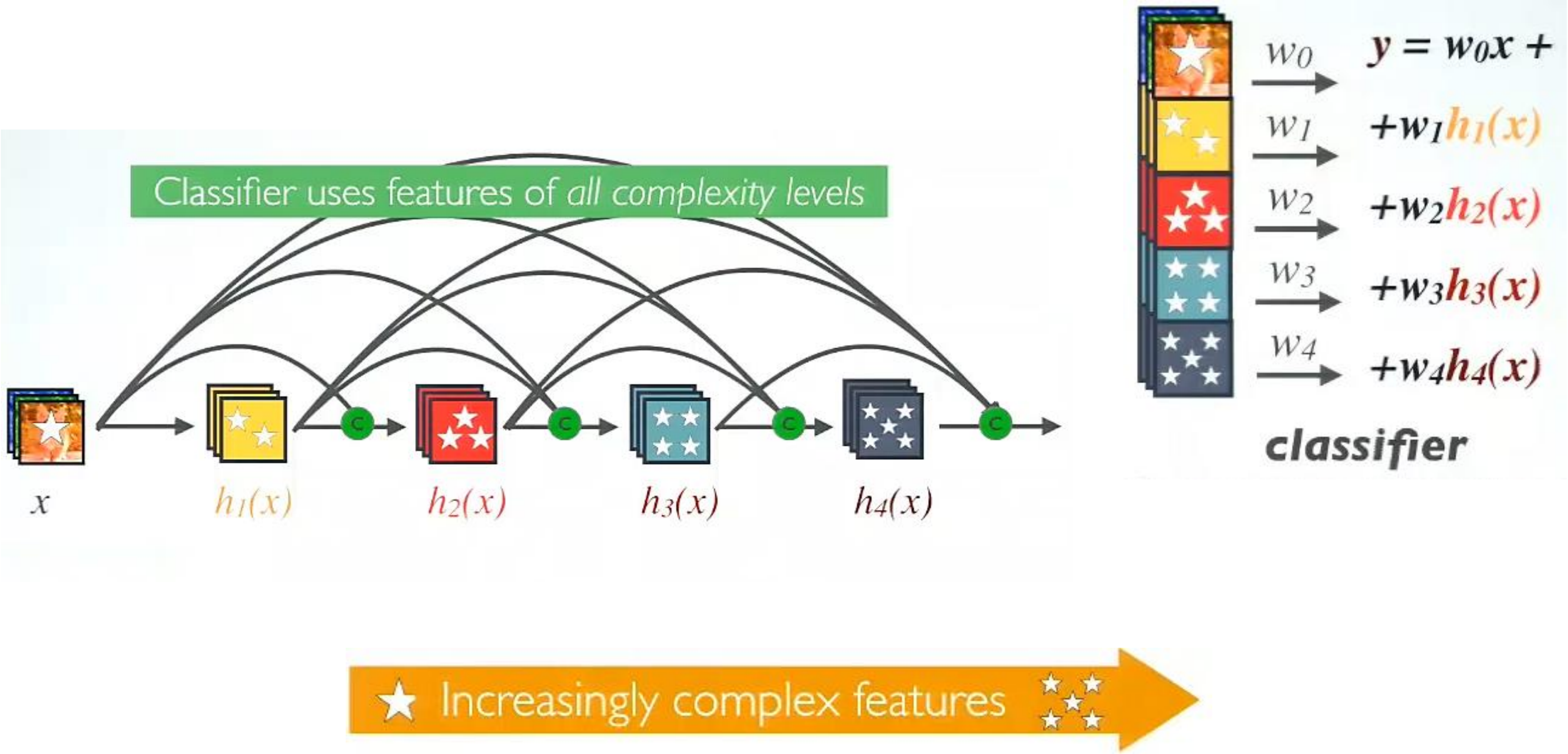
ADVATANGE 4: Maintains low complexity features

Standard connectivity:



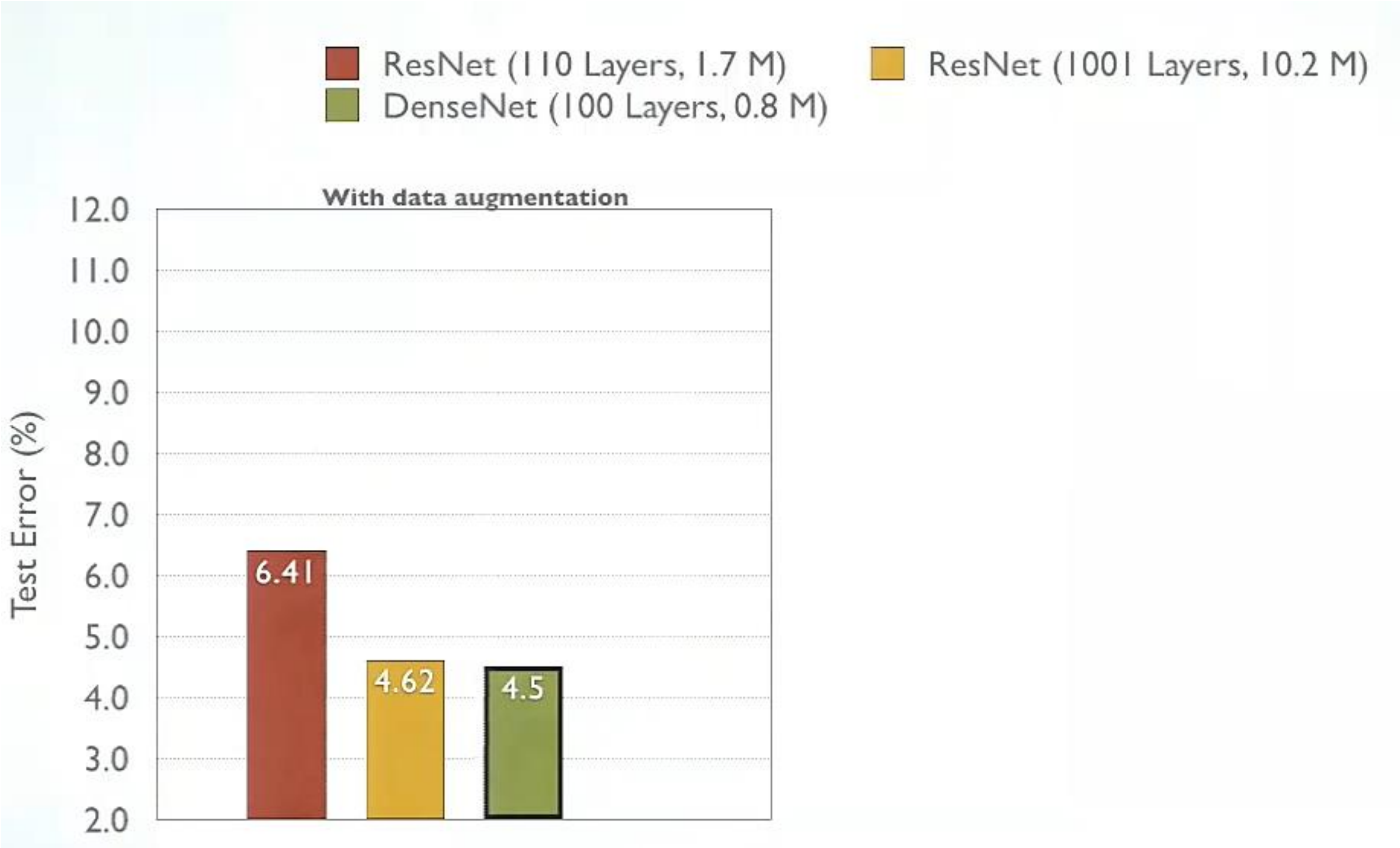
ADVATANGE 4: Maintains low complexity features

DenseNet connectivity:

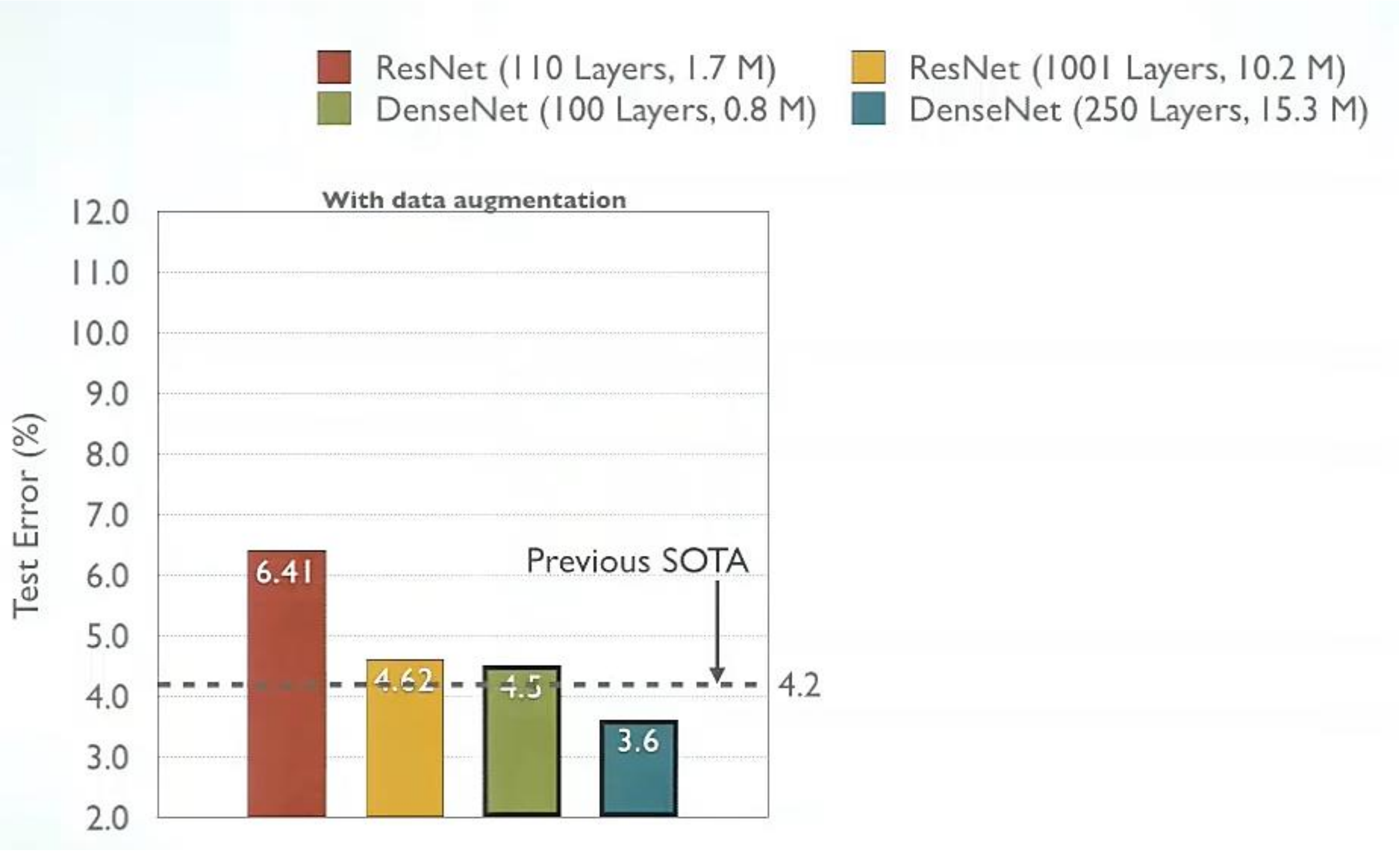


RESULTS

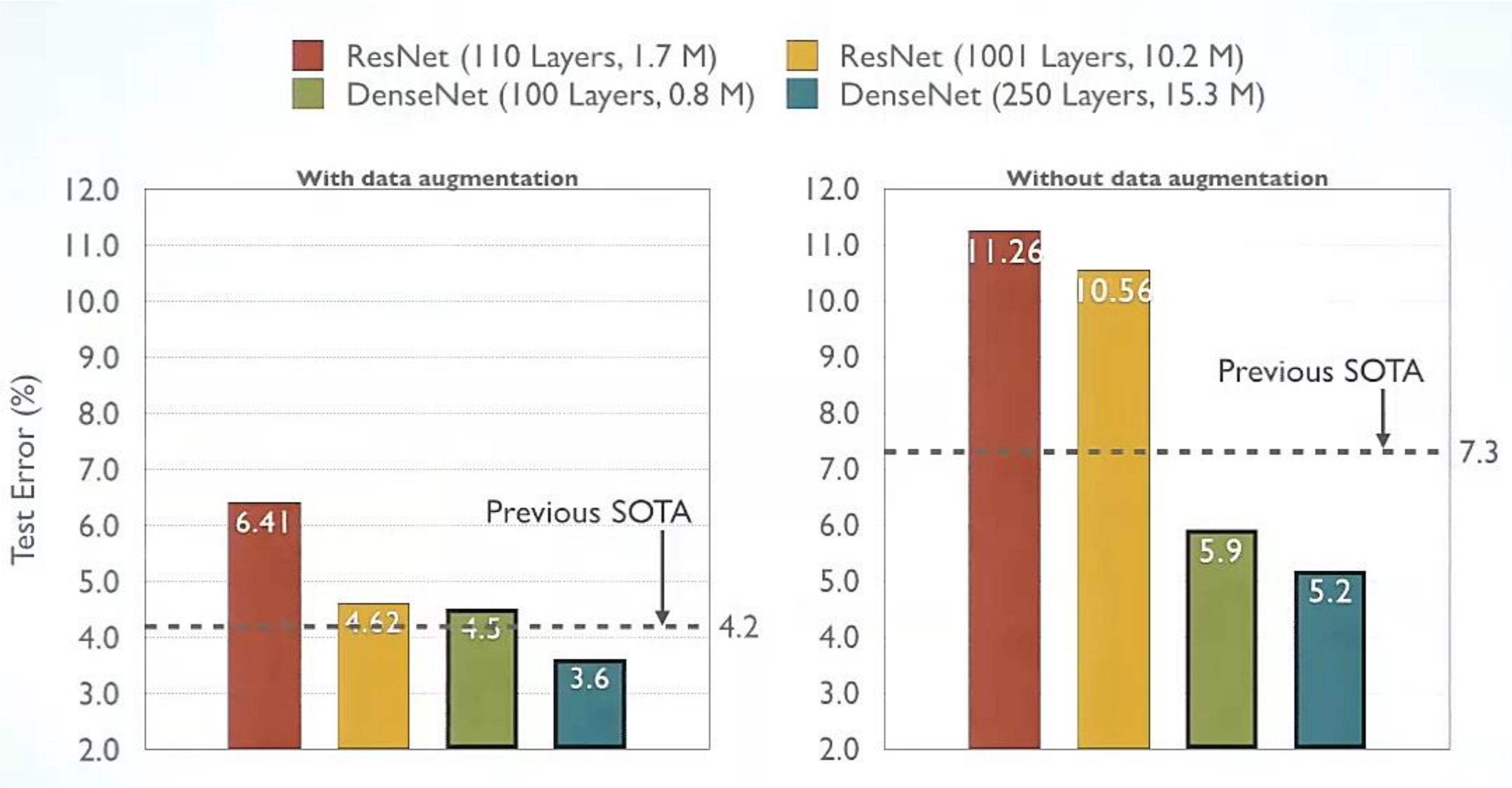
Results on **CIFAR-10**



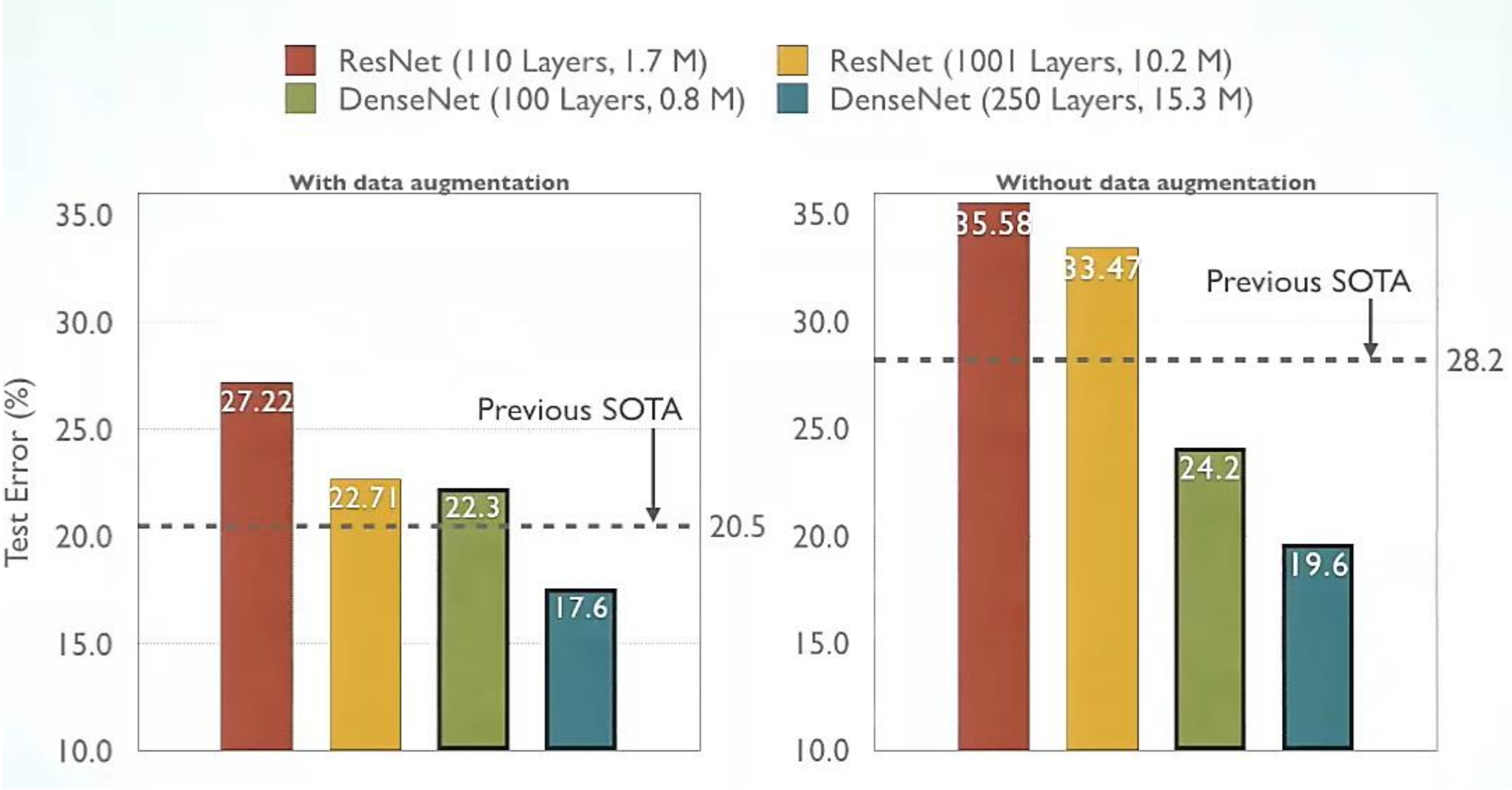
Results on **CIFAR-10**



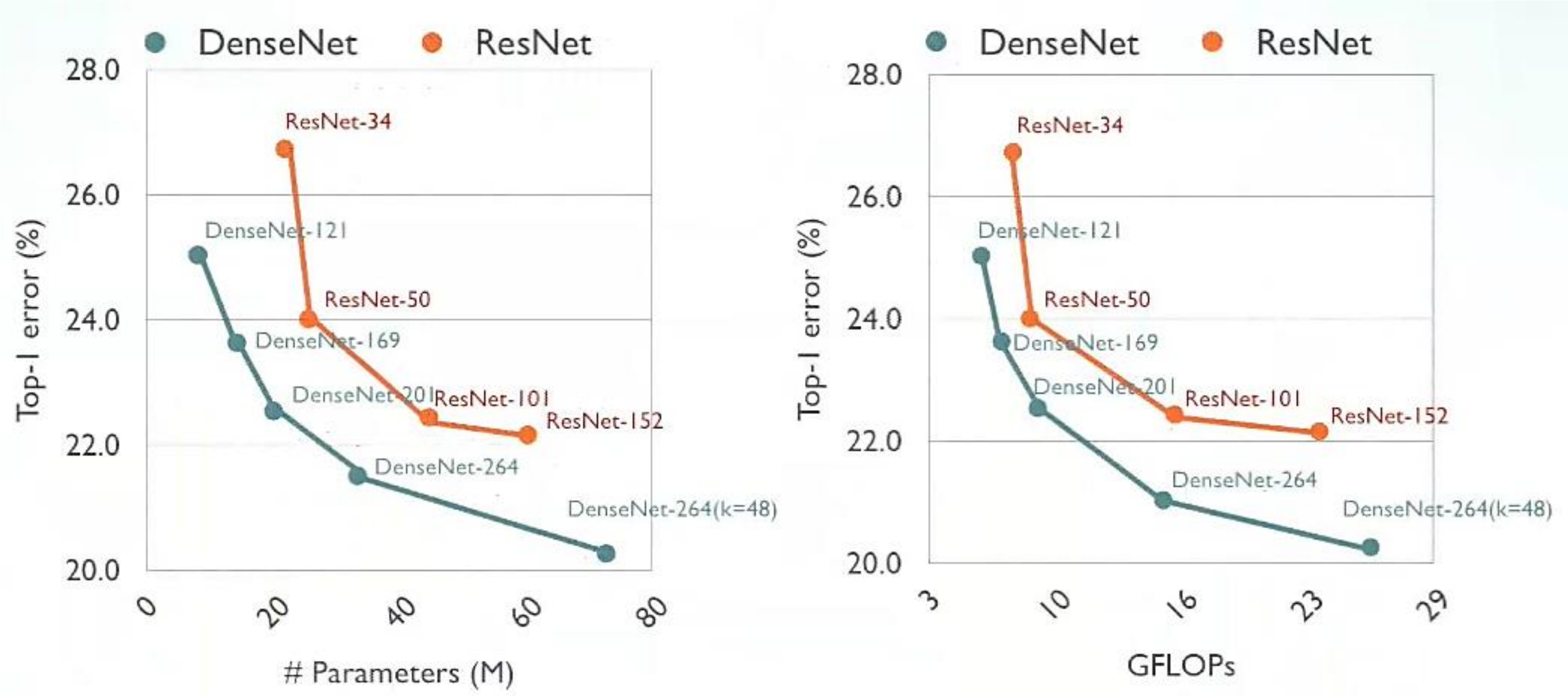
Results on **CIFAR-10**



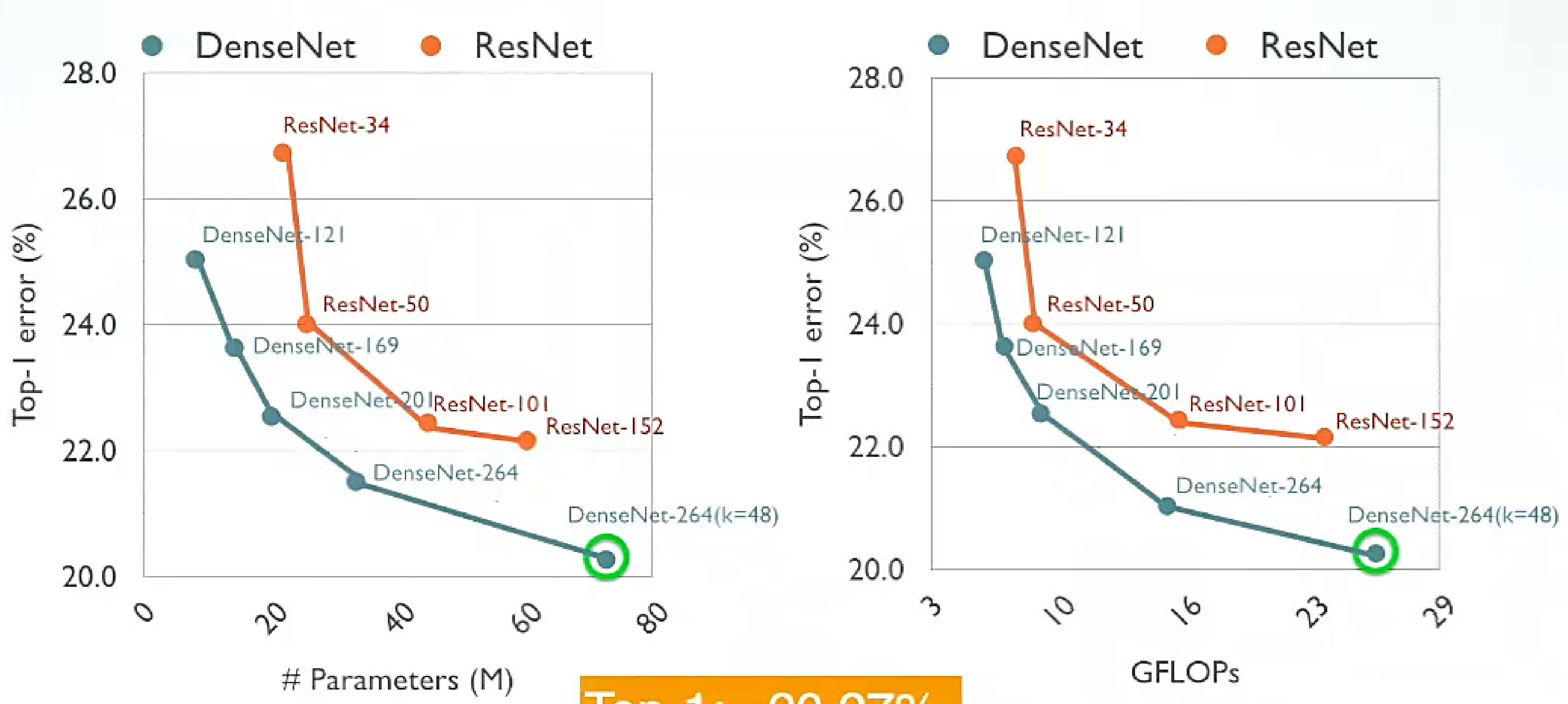
Results on **CIFAR-100**



Results on **IMAGENET**



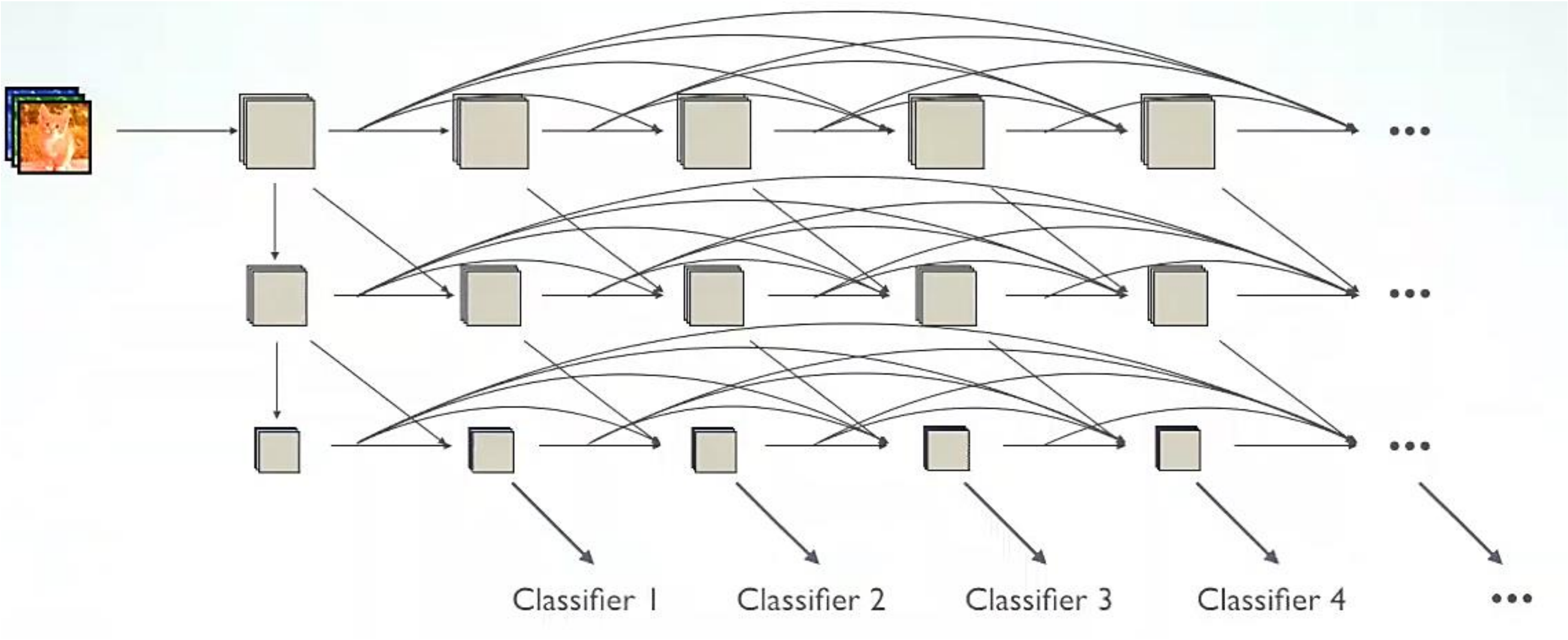
Results on **IMAGENET**



Top-1: 20.27%
Top-5: 5.17%

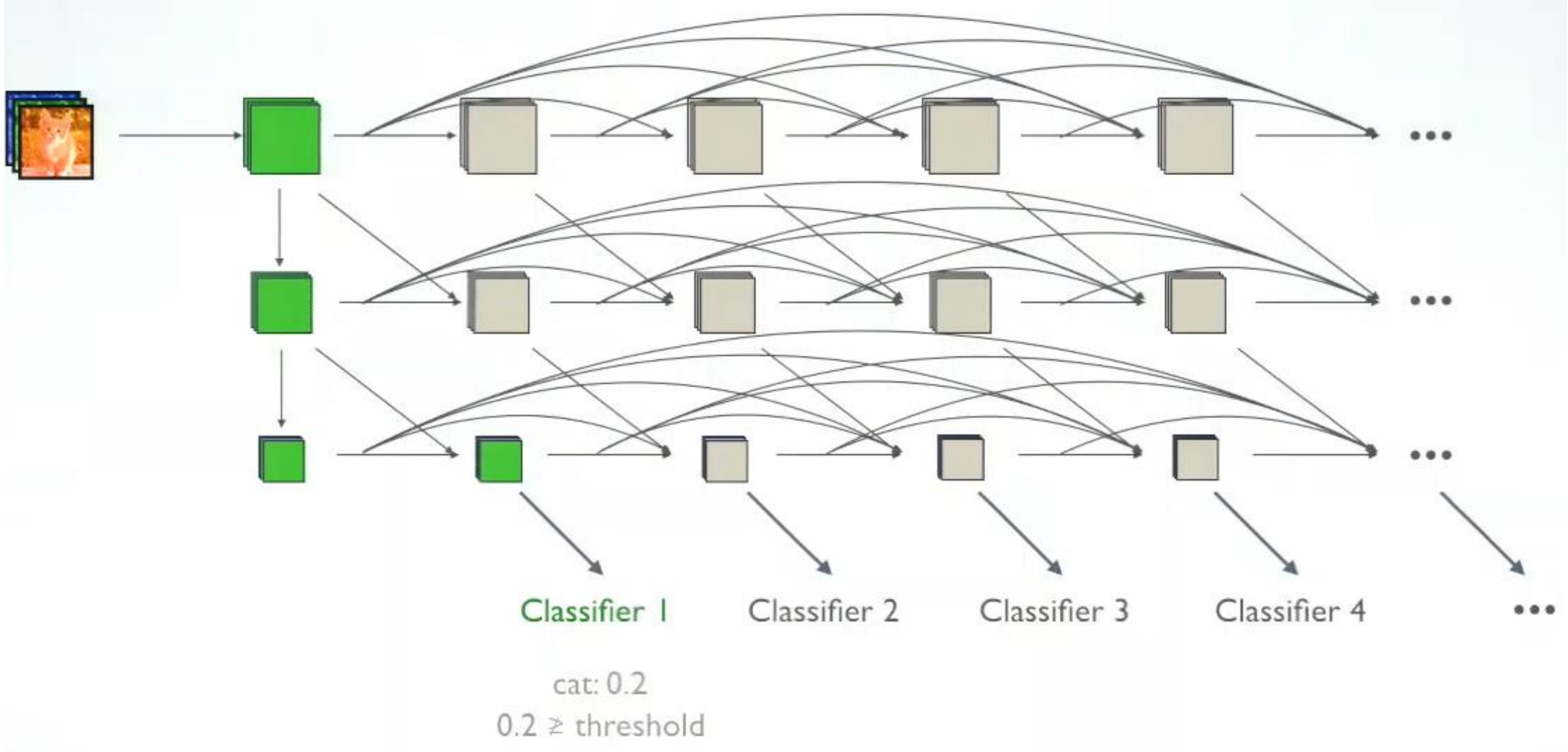
MULTI-SCALE DENSENET

Preview



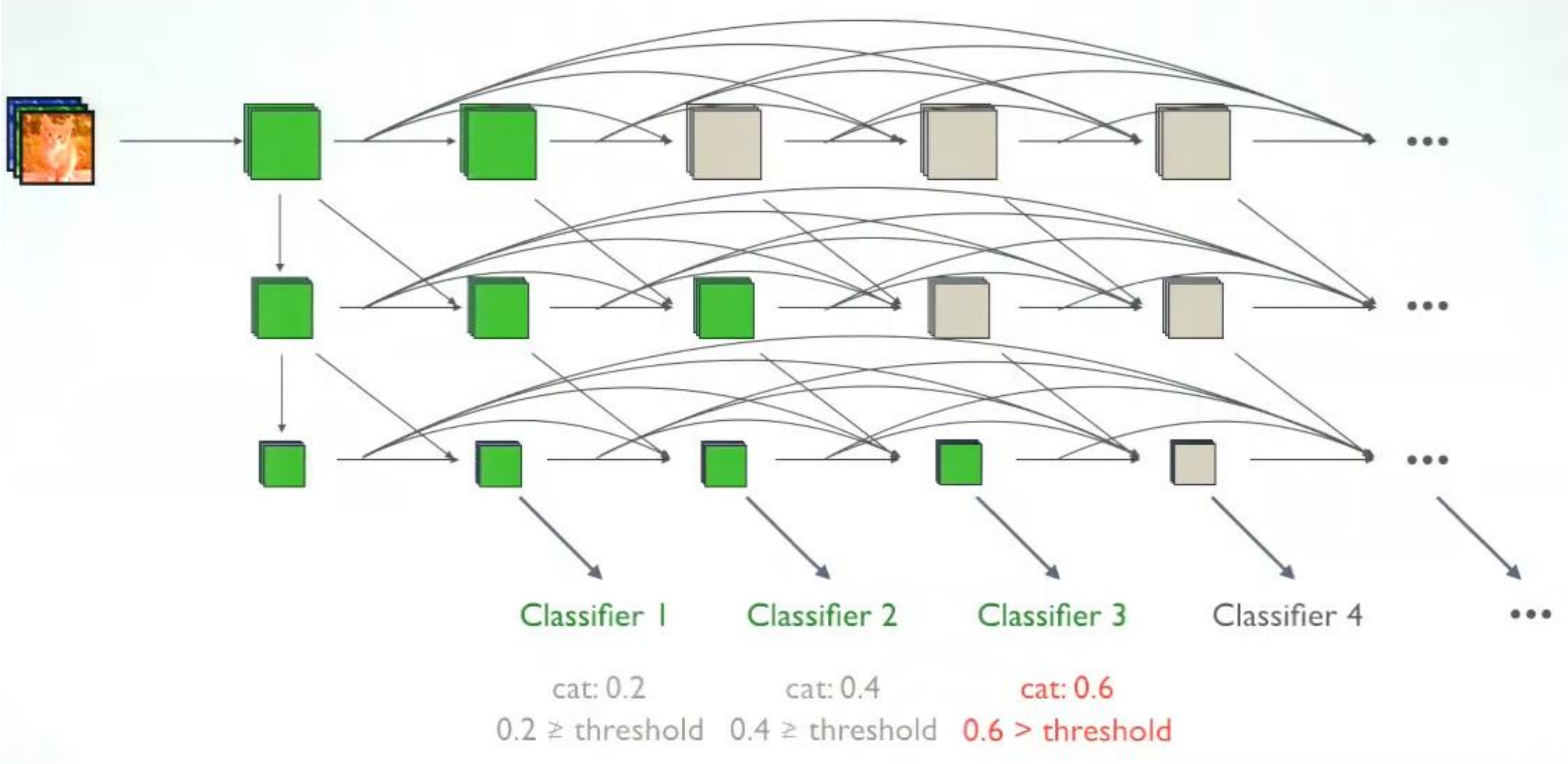
Multi-scale DenseNet: [Huang, Chen, Li, Wu, van der Maaten, Weinberger] (arXiv Preprint: 1703.09844)

Inference



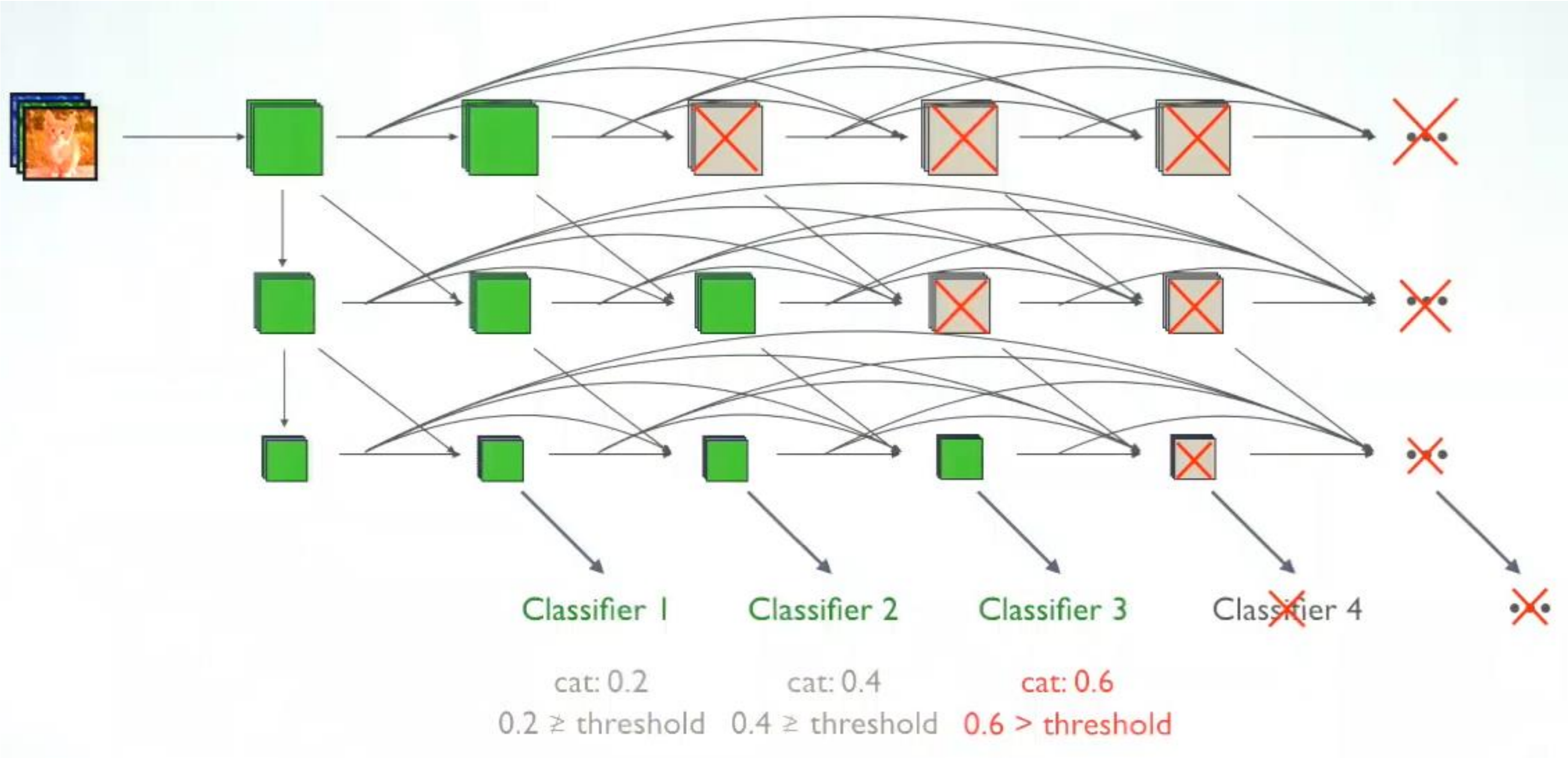
Multi-scale DenseNet: [Huang, Chen, Li, Wu, van der Maaten, Weinberger] (arXiv Preprint: 1703.09844)

Inference



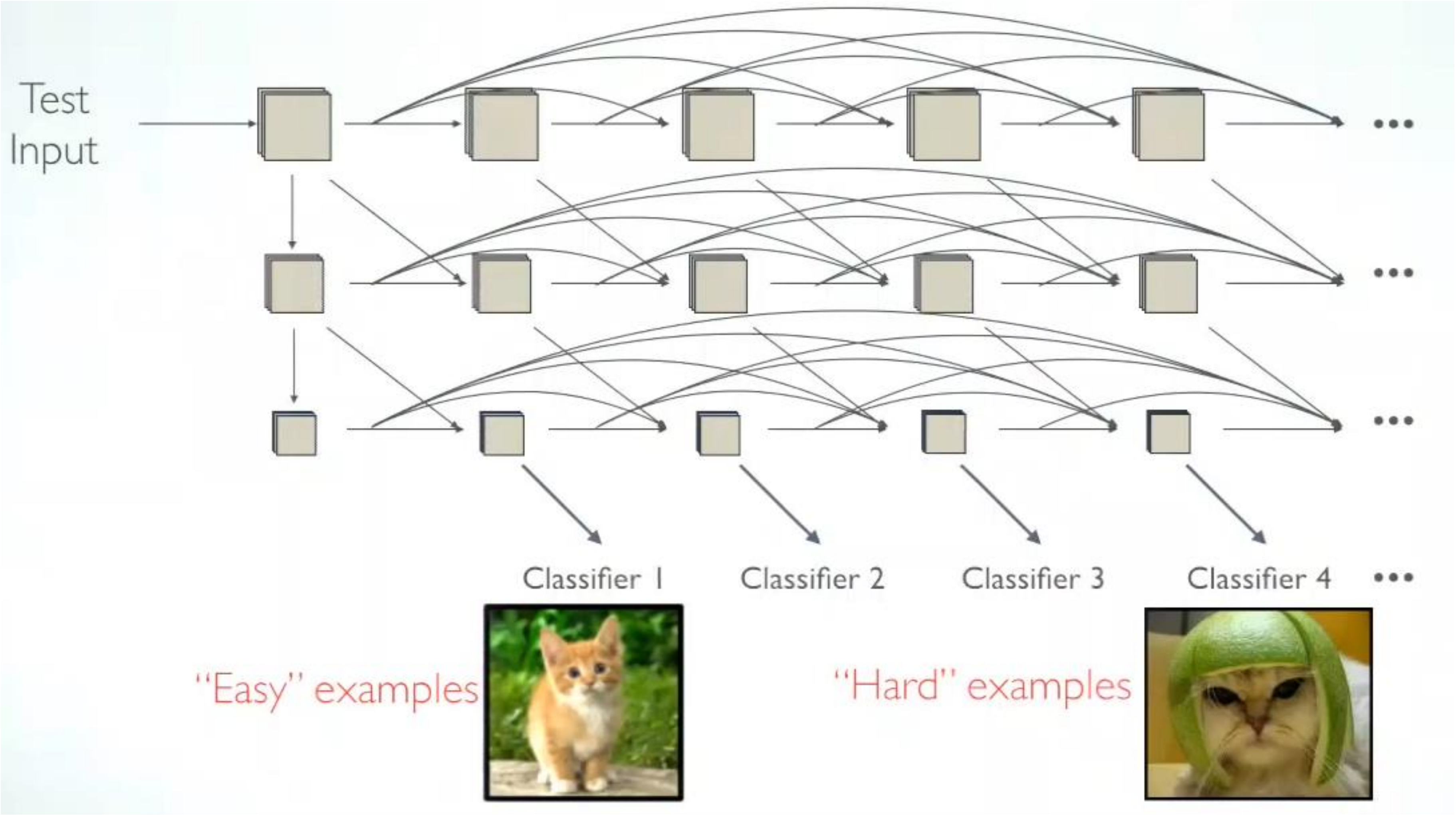
Multi-scale DenseNet: [Huang, Chen, Li, Wu, van der Maaten, Weinberger] (arXiv Preprint: 1703.09844)

Inference



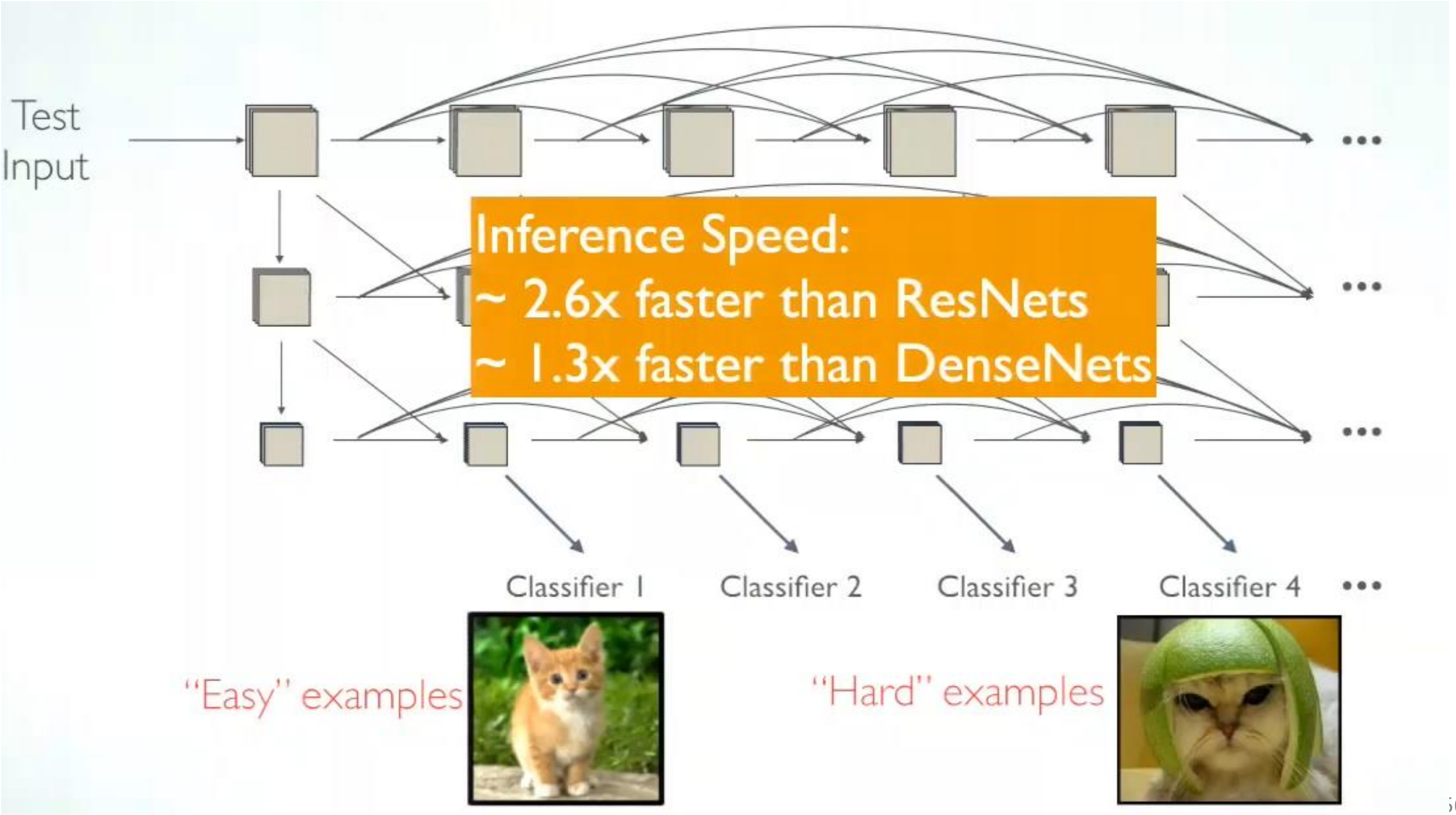
Multi-scale DenseNet: [Huang, Chen, Li, Wu, van der Maaten, Weinberger] (arXiv Preprint: 1703.09844)

Inference



Test
Input
Classifier 1
Classifier 2
Classifier 3
Classifier 4
"Easy" examples
"Hard" examples

Advantage


Test Input
Inference Speed:
~ 2.6x faster than ResNets
~ 1.3x faster than DenseNets
Classifier 1
Classifier 2
Classifier 3
Classifier 4
"Easy" examples
"Hard" examples

CONCLUSION

Final Remarks

- Proposed a new CNN architecture
- Scaling with no optimization difficulties
- Results indicate consistent improvement with increasing number of parameters
- No performance degradation or overfitting
- State-of-the-art results
- Less parameters and computations
- Can learn more compact and accurate models

Follow-up work

- Some other networks like Efficient net, NoisyStudent network and their variants have shown better benchmarks on Imagenet dataset.
- NASNet and Genetic Algorithms

References

- Stanford CS231n: Convolutional Neural Networks for Visual Recognition
- Huang, Gao, et al. "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.*
- Huang, Gao, et al. "Multi-scale dense convolutional networks for efficient prediction." *arXiv preprint arXiv:1703.09844* 2 (2017).


THANK
YOU
GRACIAS
ARIGATO
SHUKURIA
DANKSCHEEN
TASHAKKUR ATU
YAQHANYELAY
SUKSAMA
EKHMET
BÏYAN
SHUKRIA
TINGKI
GRAZIE
MEHRBANI
PALDIES
BOLZÏN
MERCI
GOZAIMASHITA
EFCHARISTO
JUSPAXAR
KOMAPSUMNIDA
MAAKE


ANY
QUESTIONS?