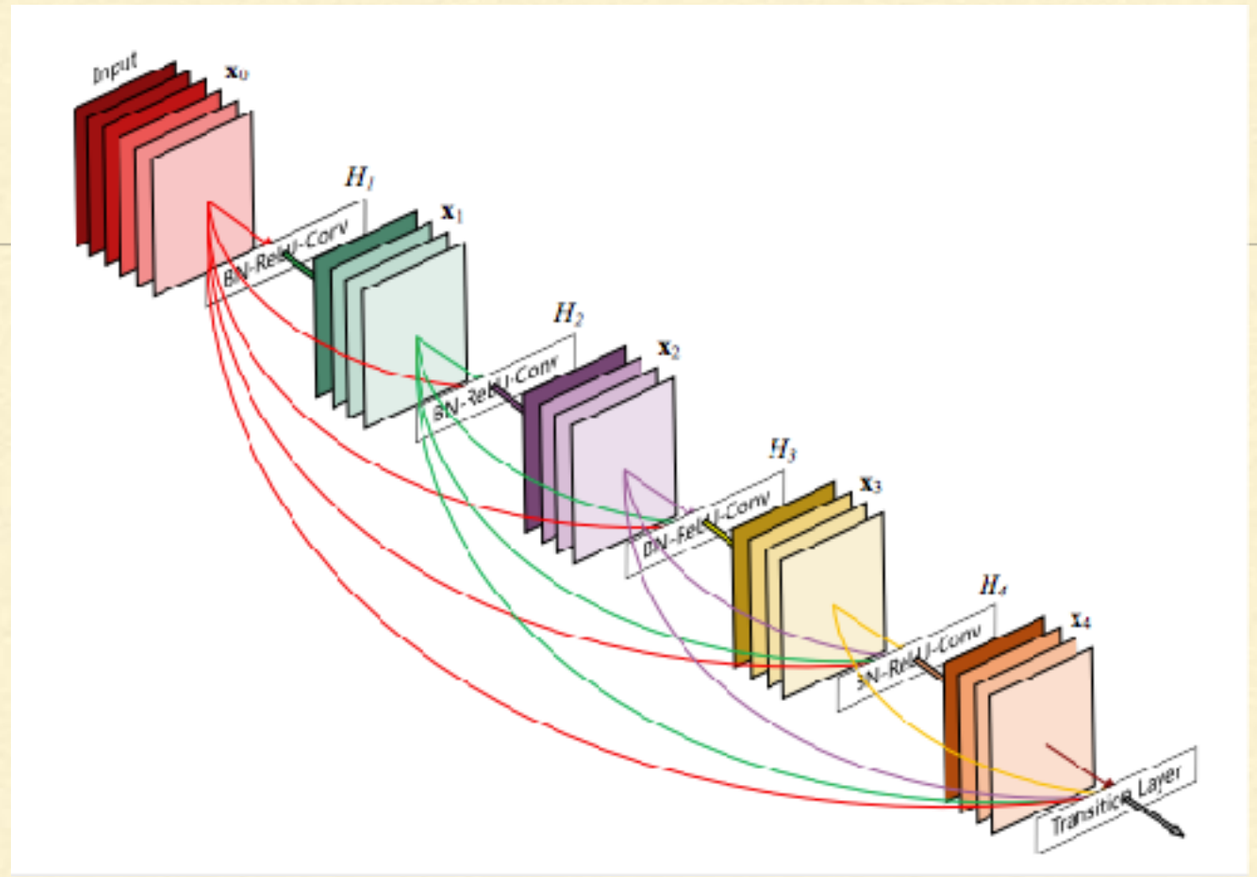
DENSELY CONNECTED CONVOLUTIONAL NETWORKS

Gao Huang* Cornell University

Zhuang liu* Tsinghua University

Laurens Van der Maaten Facebook AI Research

CVPR 2017 Best Paper

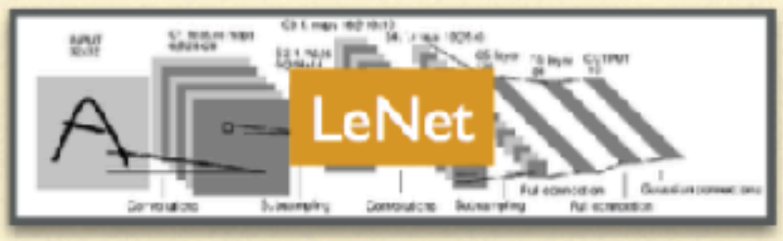


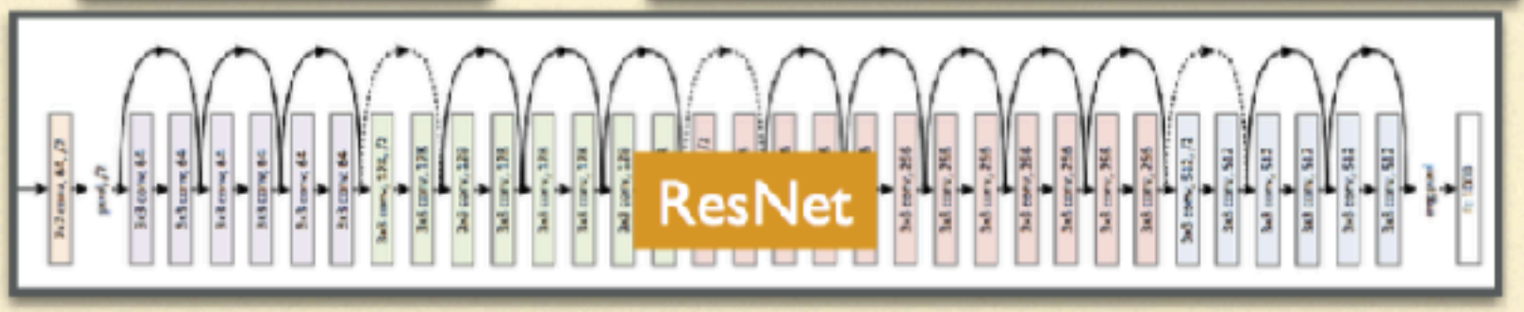
Reporter： Liao Xingyu

OUTLINE

- Development and Challenges of CNNs
- Stochastic Depth Network
- DenseNet
- Future Work

ARCHITECTURES FOR DEEP CONVOLUTIONAL NETWORKS





ZFNet
Network in Network
Inception-ResNet
Xception
FractalNet
DenseNet
MobileNet
ShuffleNet
SENet
Dual Path Network
…

DEVELOPMENT OF CNNS

- Shallower —> **Deeper**
- Large filter kernels —> **Small filter kernels**
- Multiple fully connected layers —> **All convolutional layers**
- Layer-wise connection —> **Cross-layer connection**

CHALLENGES IN DEEP LEARNING

- Optimization difficulty
  - Gradient vanishing / exploding
- Generalization degradation
  - Overfitting
- Resource demanding
  - Long training and inference time

STOCHASTIC DEPTH NETWORK

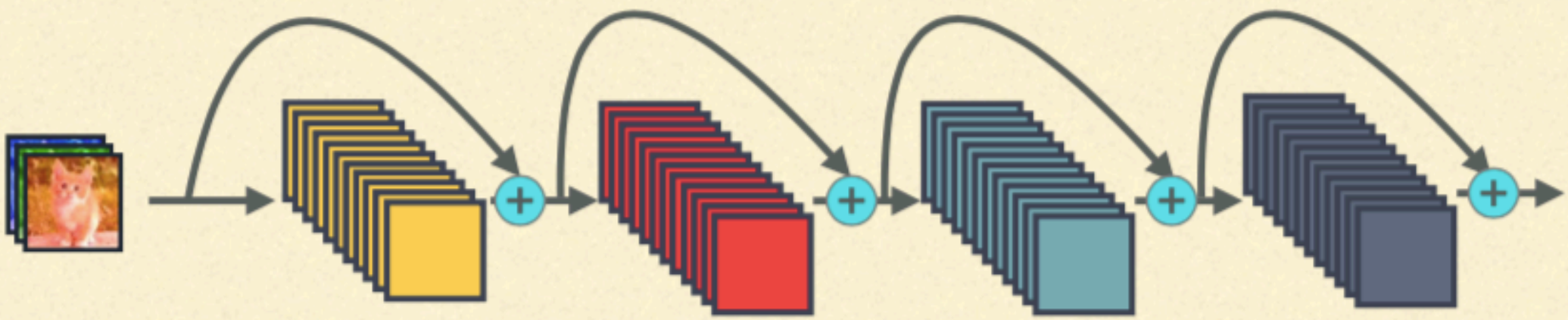
Speeds Up Training

Eases Optimization

Fights against Overfitting

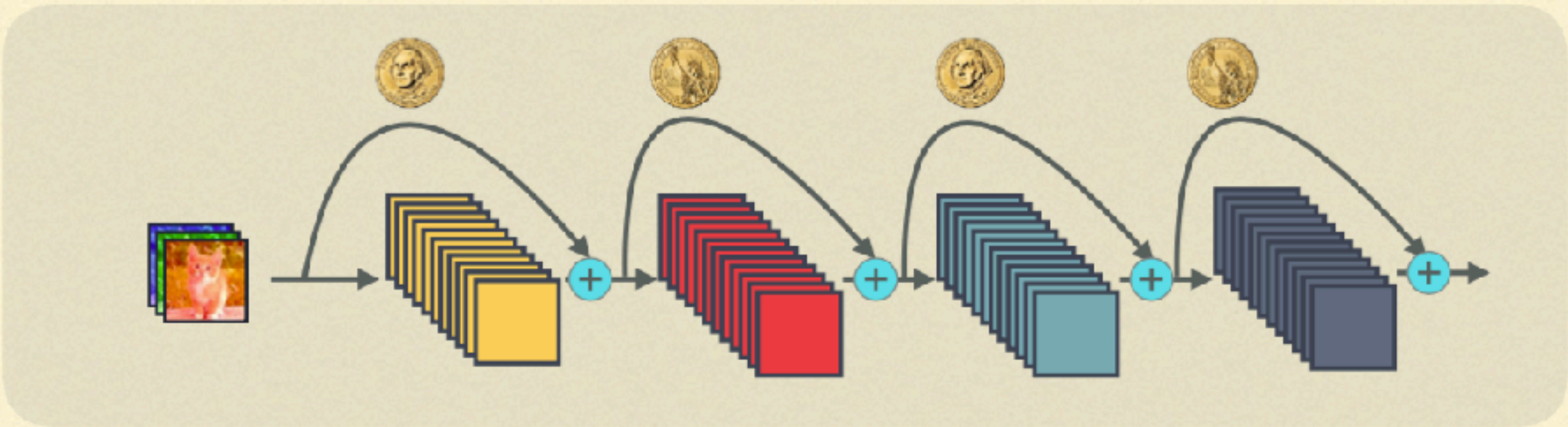
RESNET CONNECTIVITY

Identity mappings promote gradient propagation.

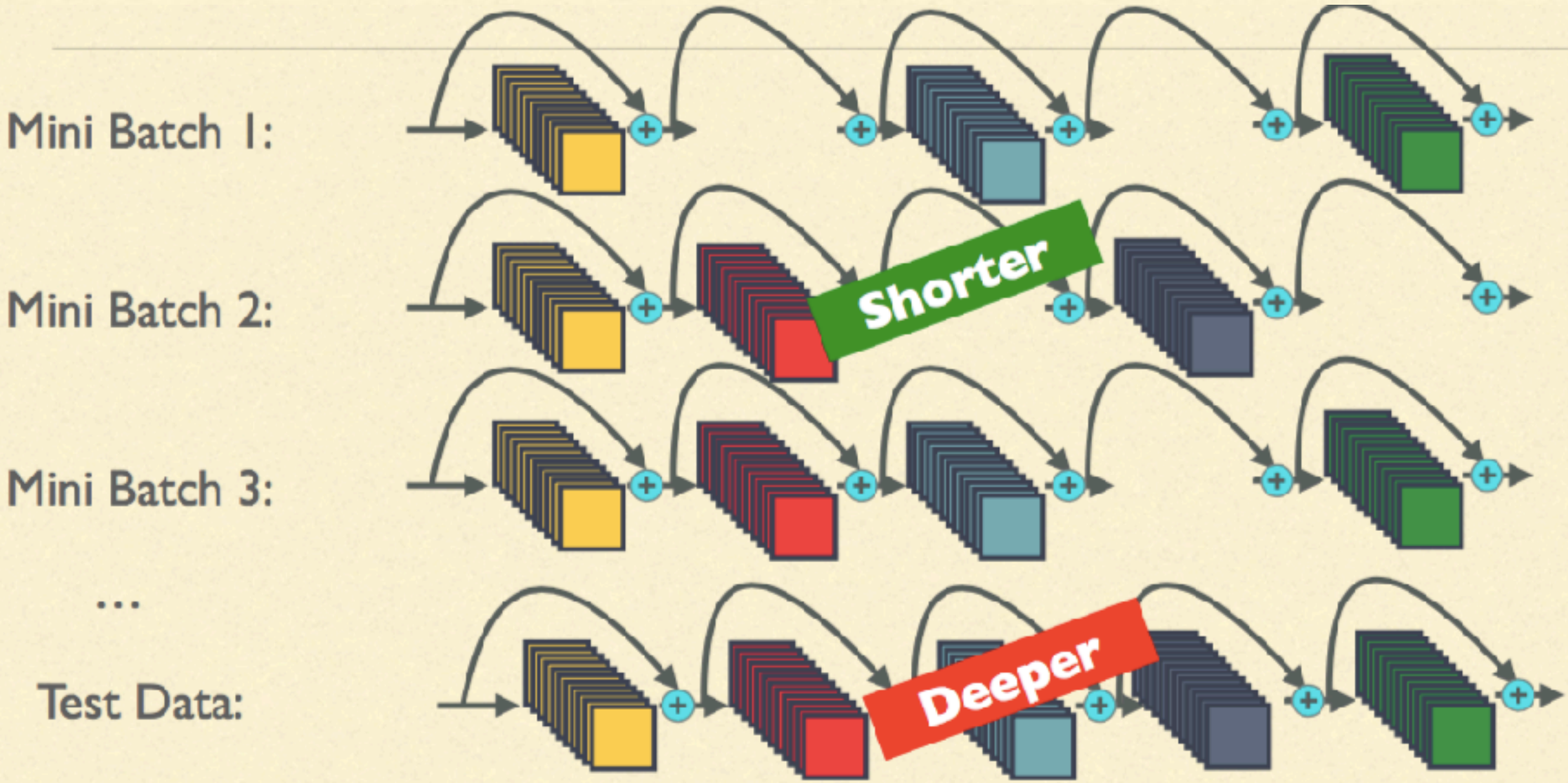


⊕ : Element-wise addition

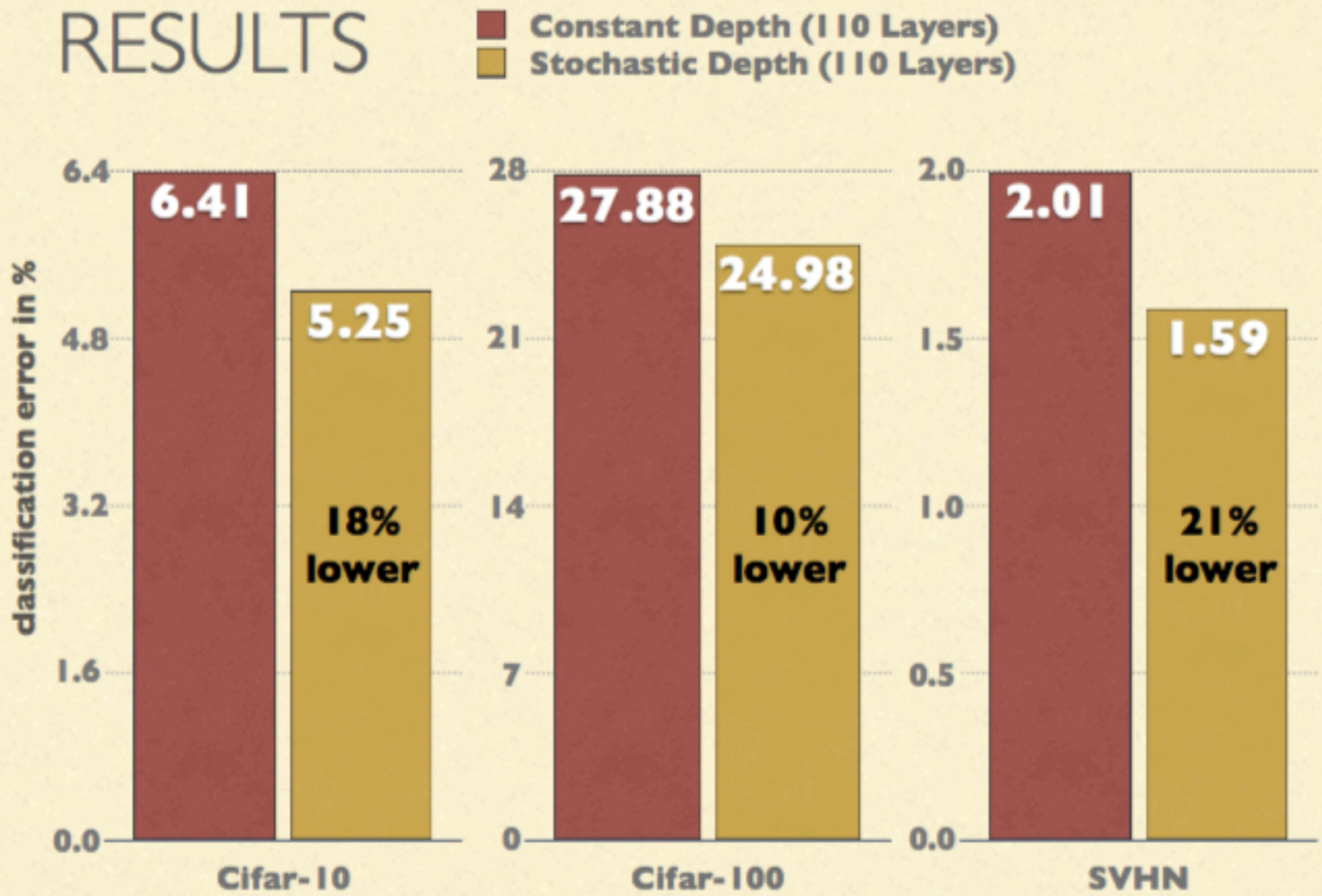
STOCHASTIC DEPTH



STOCHASTIC DEPTH



RESNETS WITH STOCHASTIC DEPTH

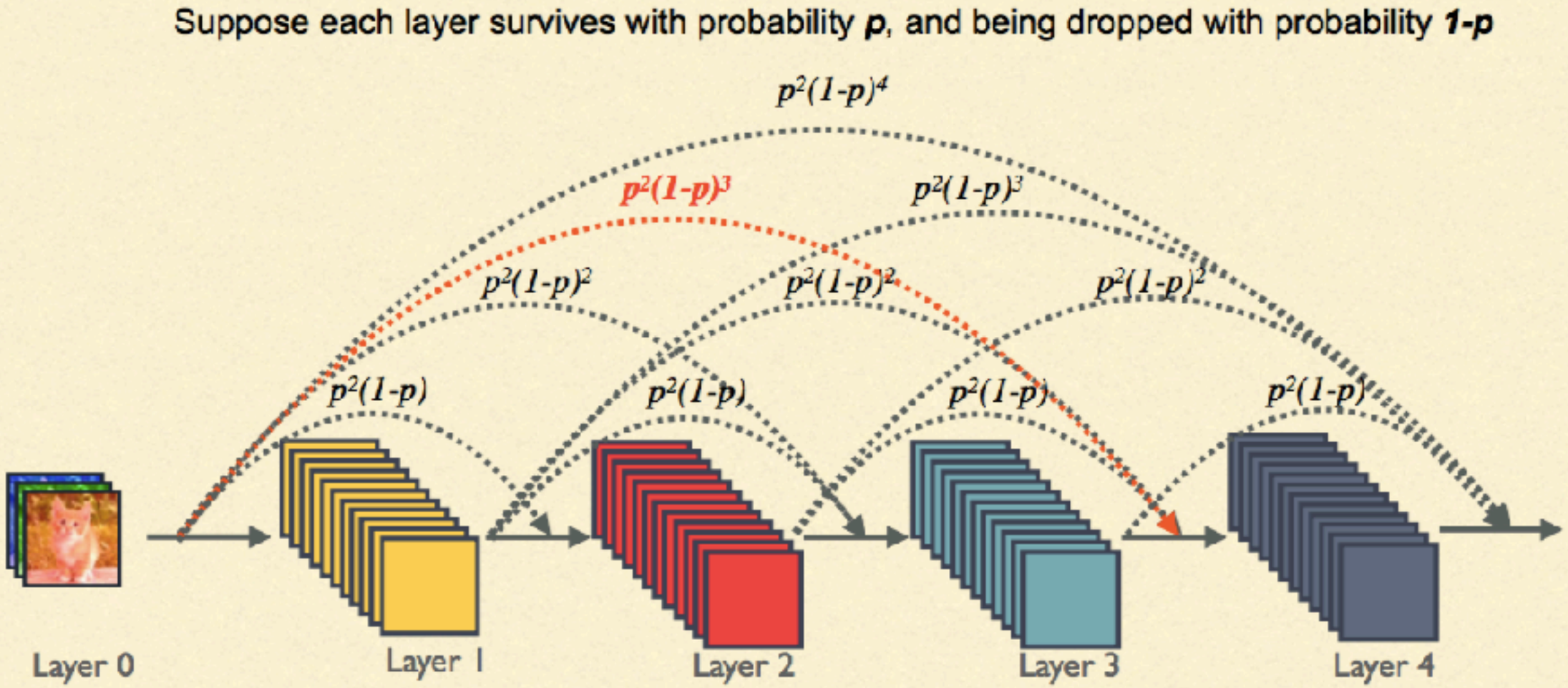


STOCHASTIC DEPTH

With stochastic depth, we can train ResNets:

- faster
- with deeper architectures
- with better generalization performance

UNDERSTAND STOCHASTIC DEPTH

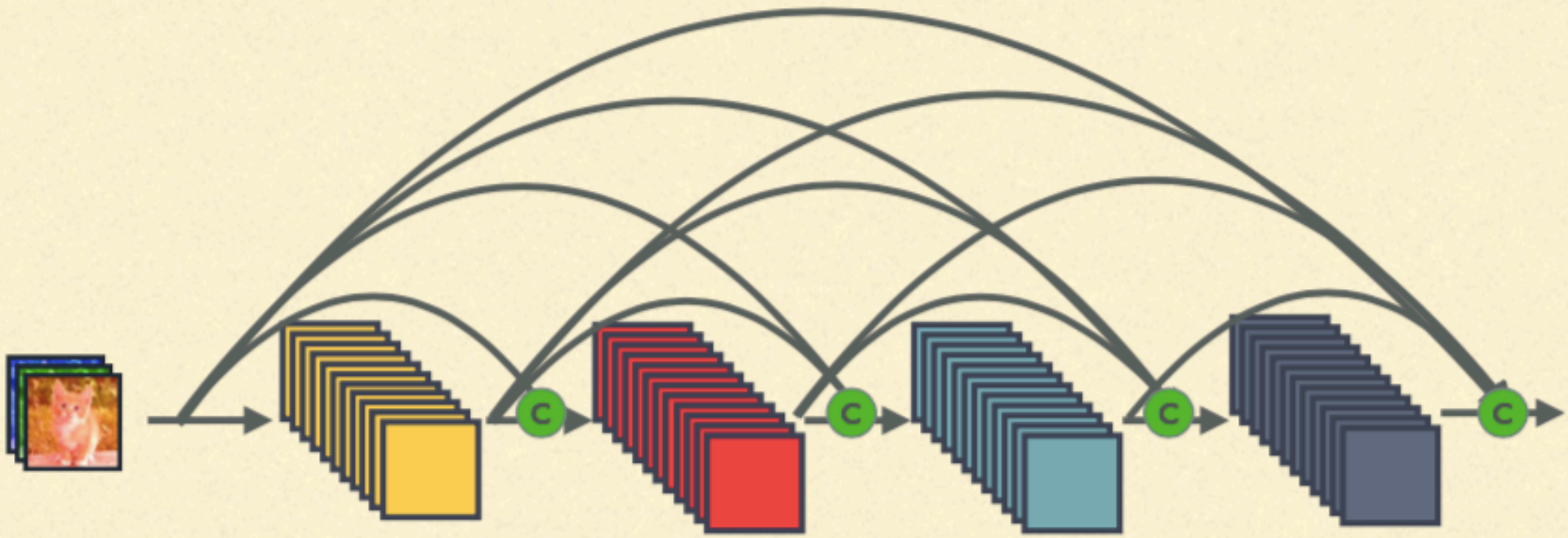


DENSENET

Compact Network Architecture

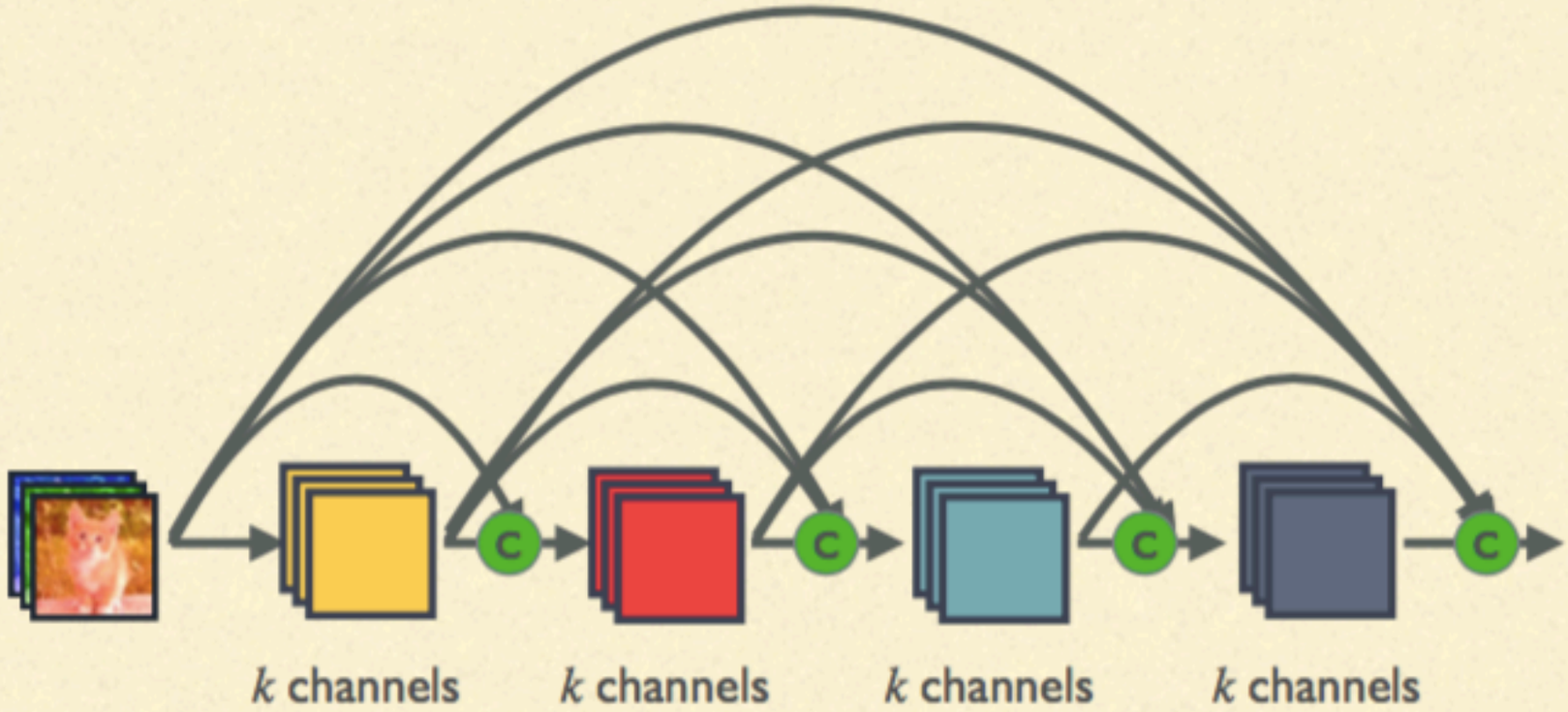
Better Generalization

DENSE CONNECTIVITY



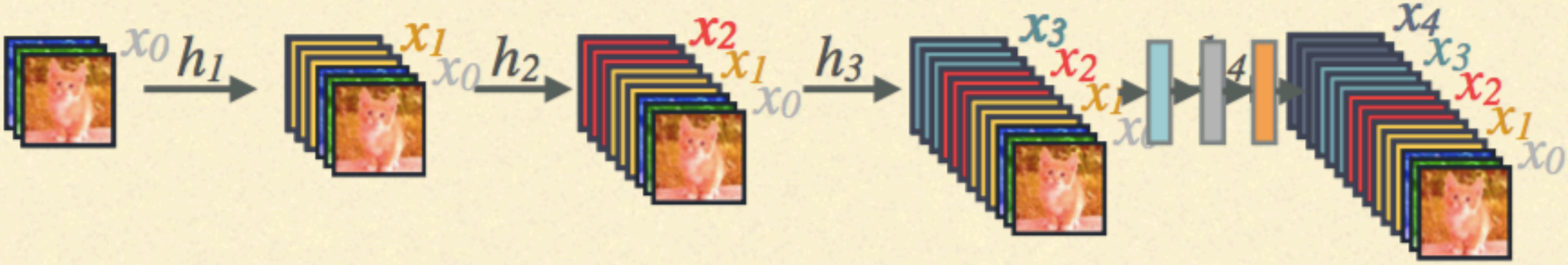
C : Channel-wise concatenation

DENSE AND SLIM

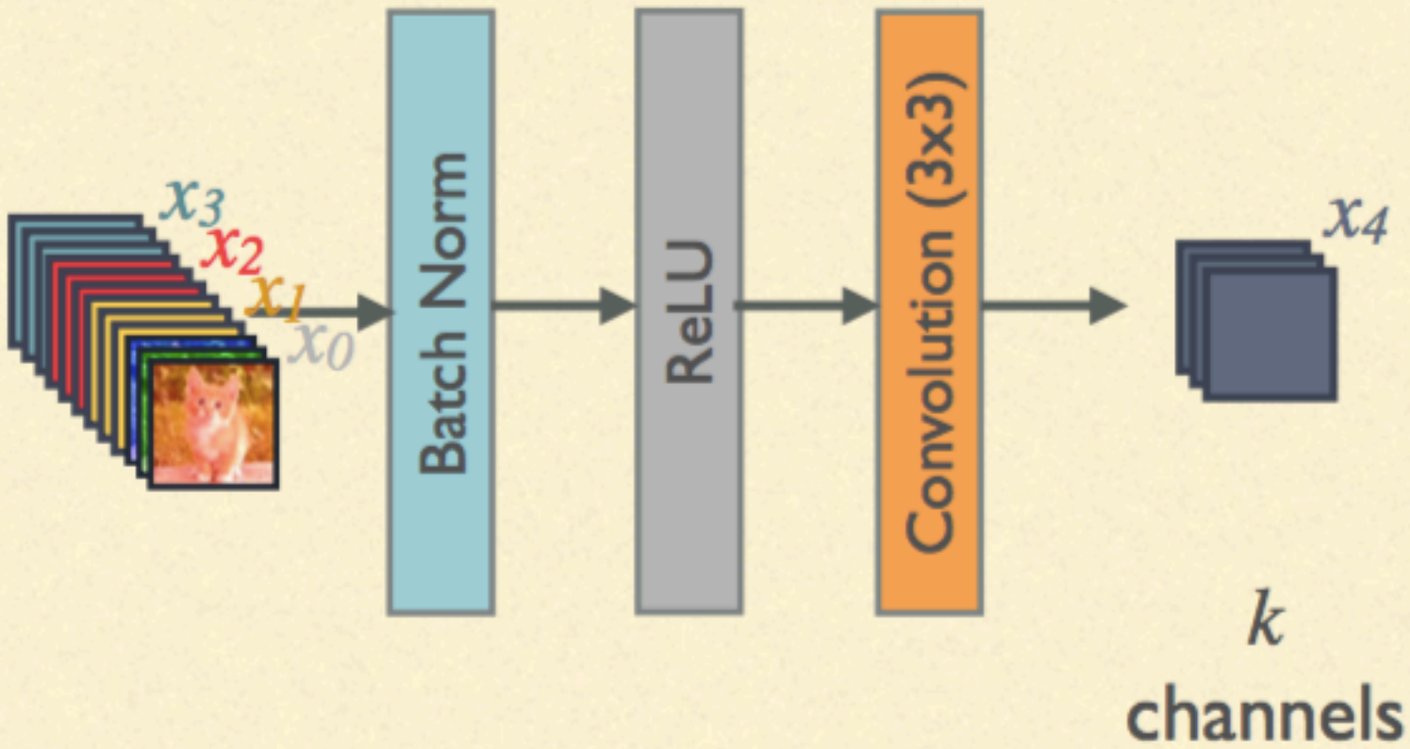


k : Growth Rate

FORWARD PROPAGATION

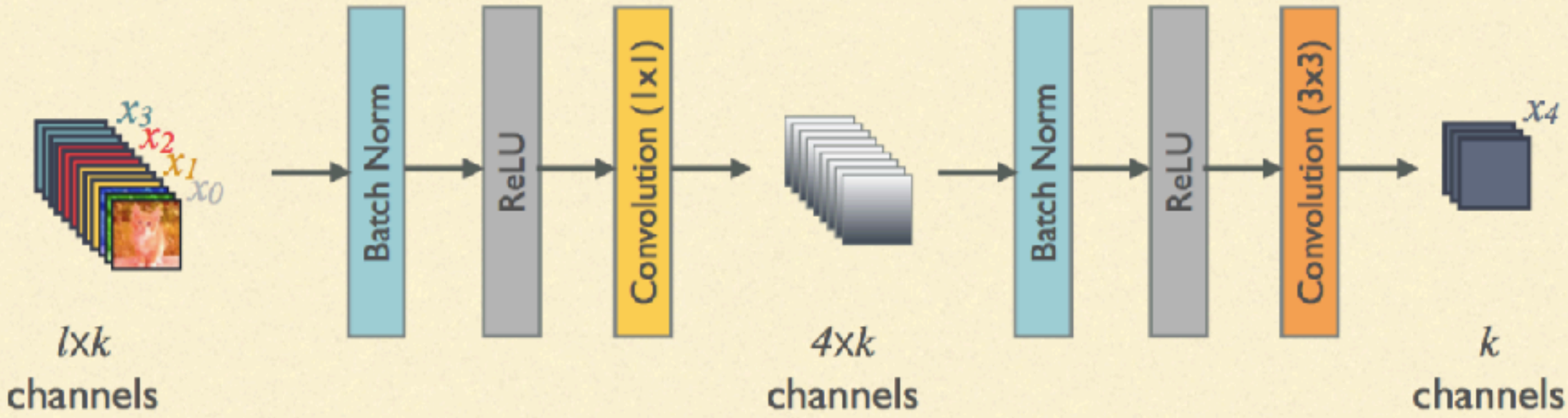


SINGLE LAYER IN DENSENET



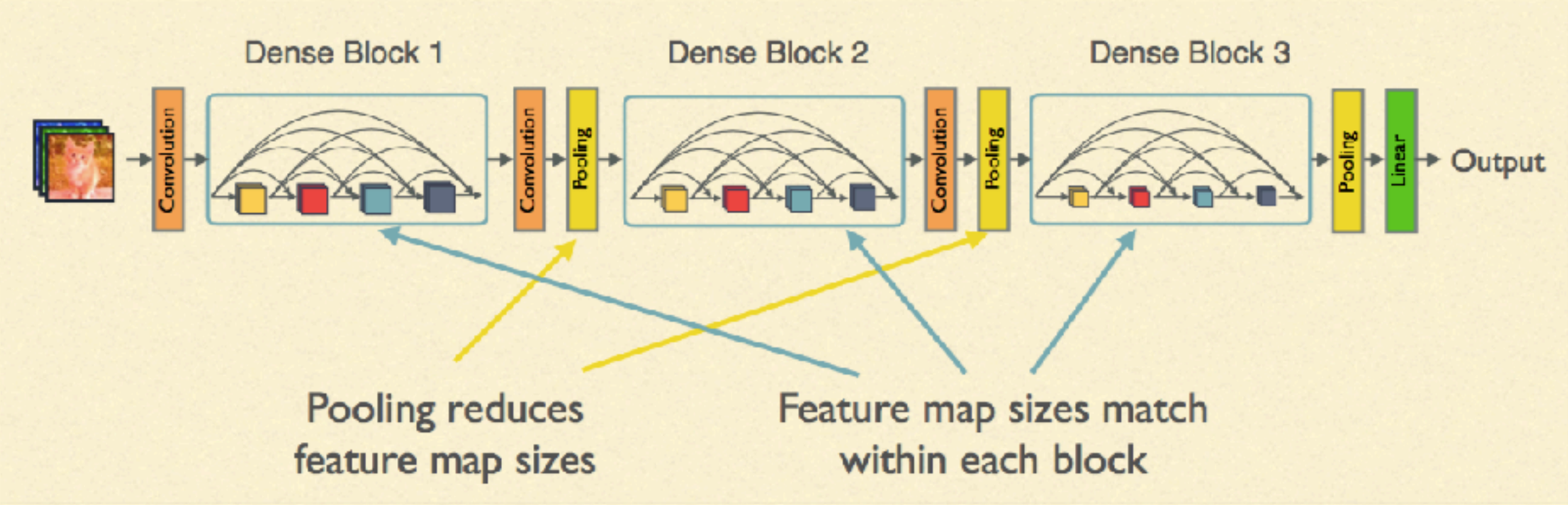
$$x_5 = h_5([x_0, \ldots, x_4])$$

BOTTLENECK LAYER

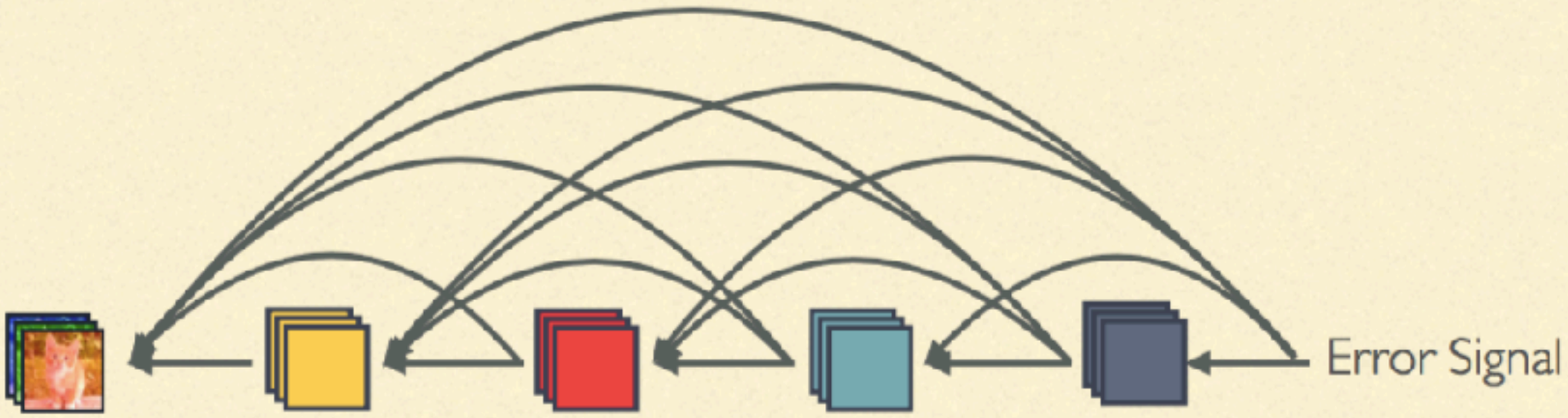


Higher parameter and computational efficiency

DENSENET


Dense Block 1
Dense Block 2
Dense Block 3
Convolution
Convolution
Pooling
Convolution
Pooling
Pooling
Linear
Output
Pooling reduces
feature map sizes
Feature map sizes match
within each block

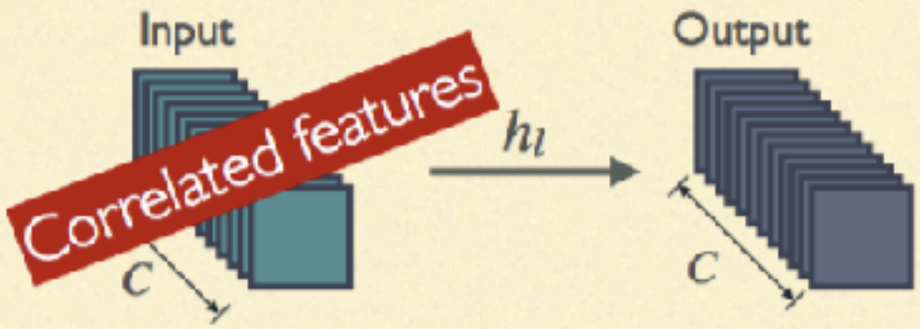
ADVANTAGE 1: STRONG GRADIENT FLOW



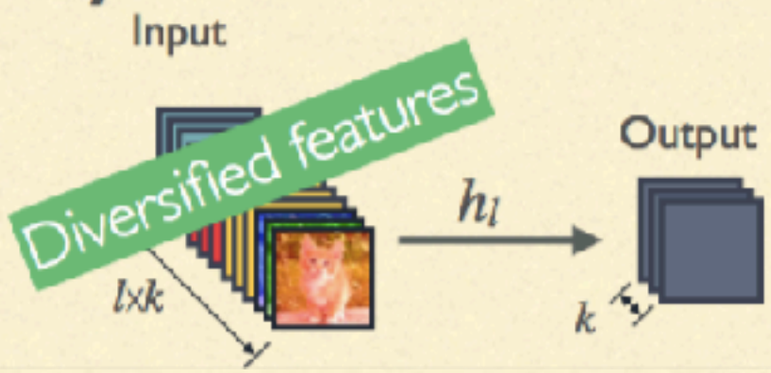
Implicit "deep supervision"

ADVANTAGE 2: PARAMETER & COMPUTATIONAL EFFICIENCY

Standard connectivity:



Dense connectivity:



#parameters:

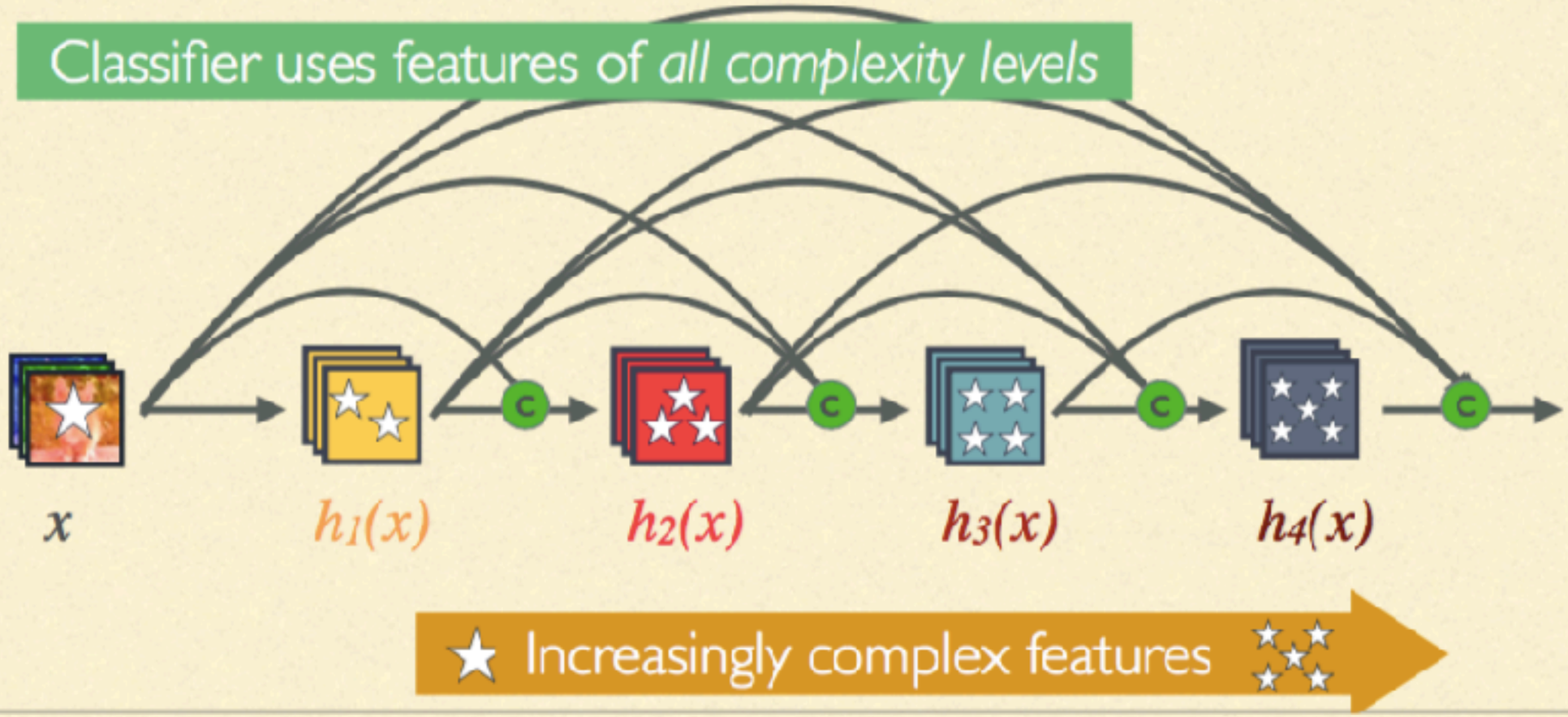
$O(C \times C)$

$k \ll C$

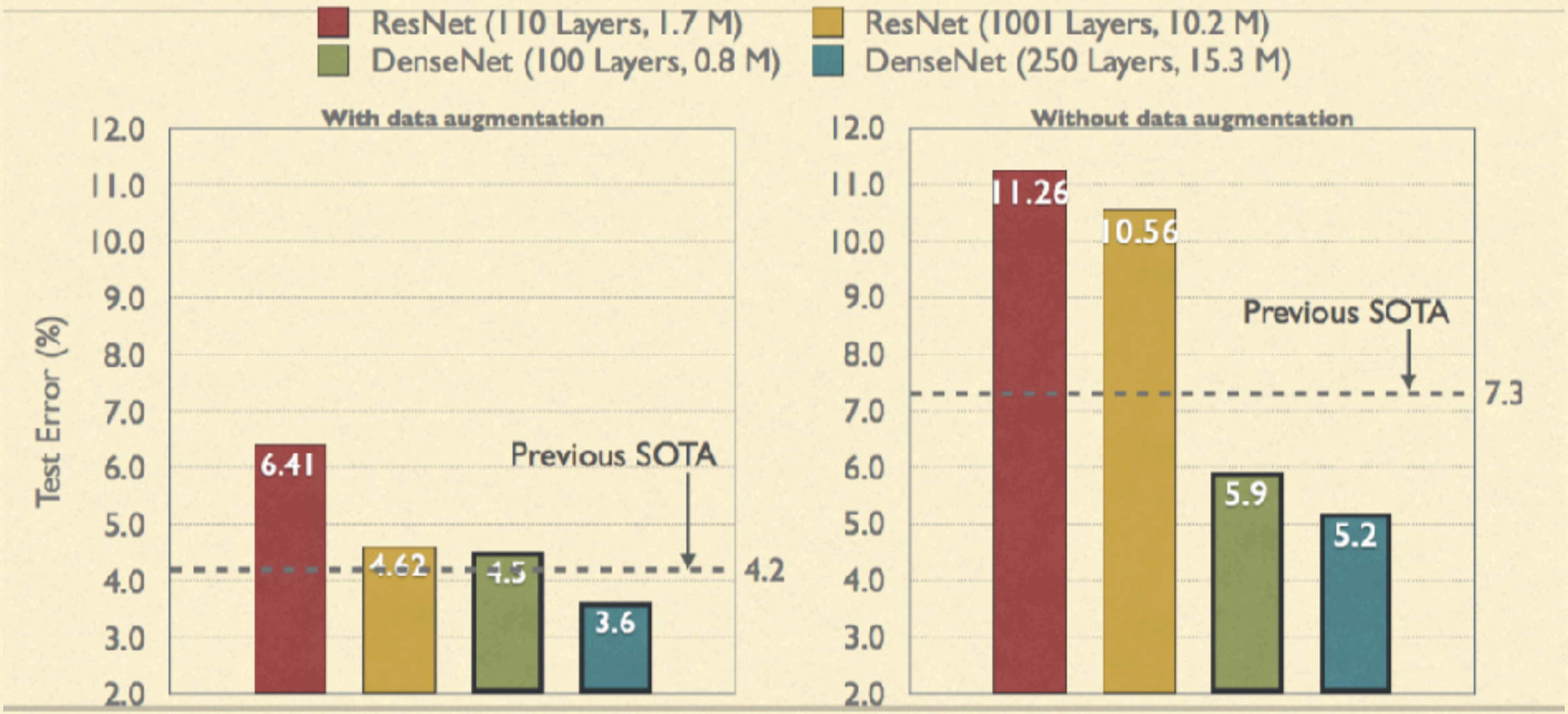
$O(l \times k \times k)$

Growth rate k

ADVANTAGE 3: MAINTAINS LOW COMPLEXITY FEATURES



RESULTS ON **CIFAR-10**



FUTURE WORK

- Memory Using (https://github.com/gpleiss/efficient_densenet_pytorch)
- Feature scale (https://arxiv.org/pdf/1709.01507.pdf) SENet

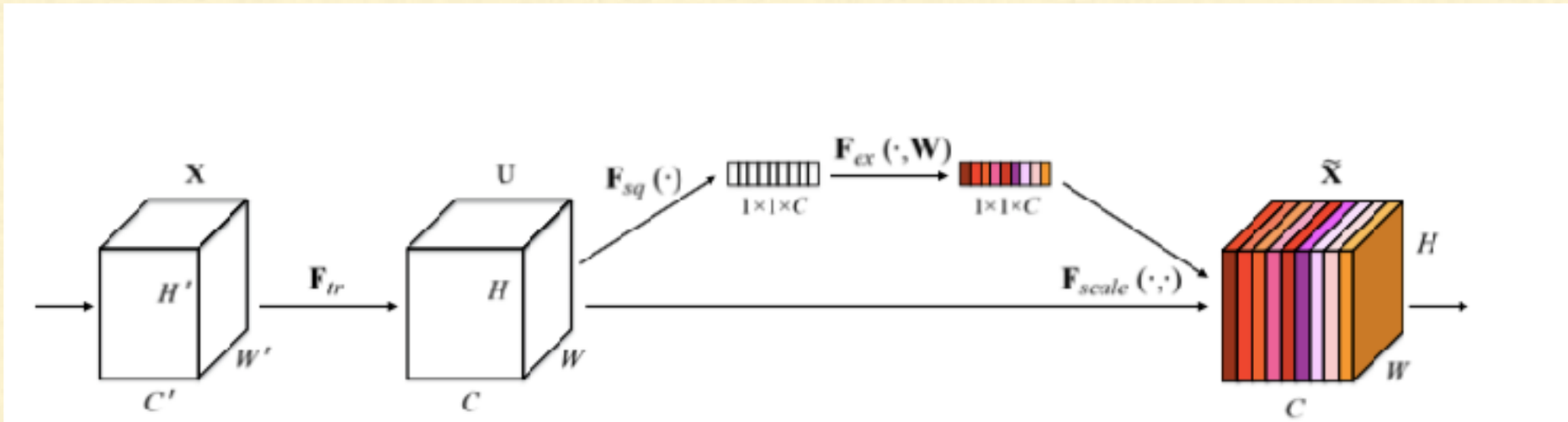


Figure 1. A Squeeze-and-Excitation block.

CODE

- DenseNet(https://arxiv.org/pdf/1608.06993.pdf)
- https://github.com/liuzhuang13/DenseNet(Torch)
- https://github.com/gpleiss/efficient_densenet_pytorch(PyTorch)

Third-party implementations in TensorFlow, Keras, Caffe, MxNet, Chainer, etc

MY IMPLEMENTATION

https://github.com/SherlockLiao/cifar10-gluon