DenseNet

reusing : pr-028,cs231n,DenseNet(CVPR2017 best paper)

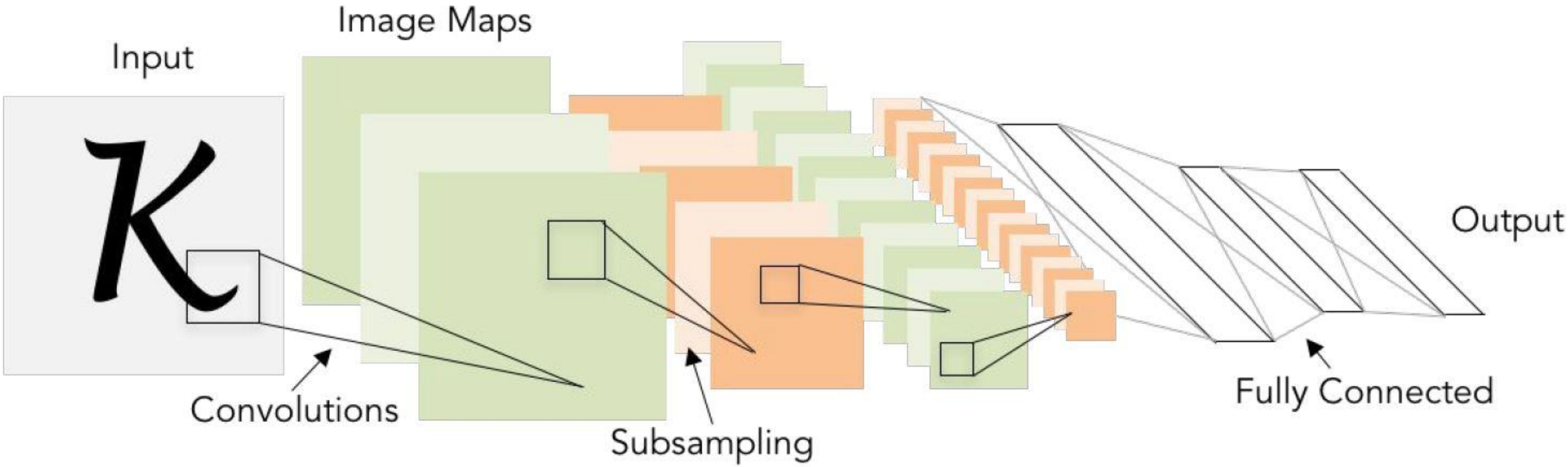
서강대 영상대학원

김태형

Lecture 9:
CNN Architectures

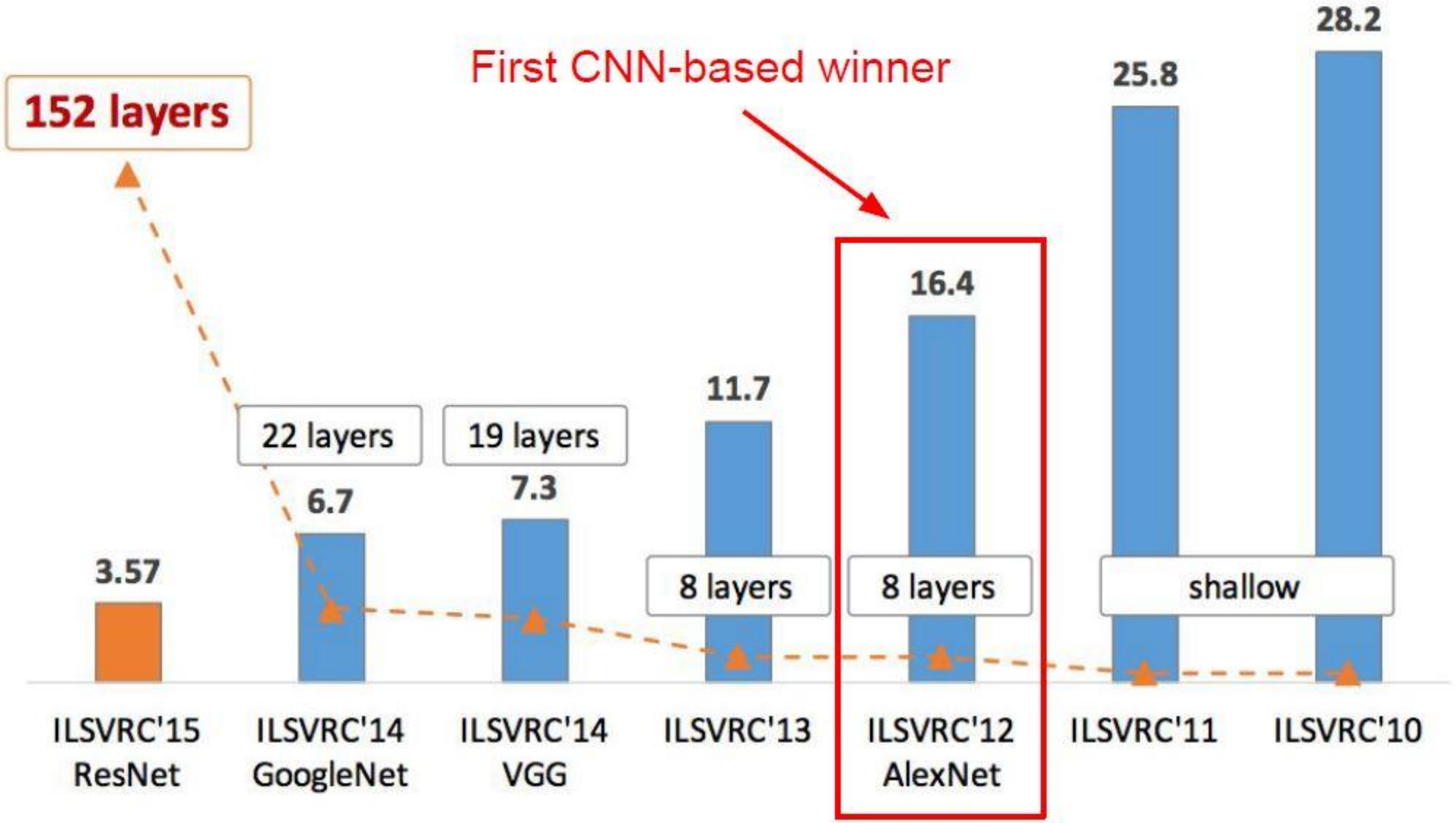
Review: LeNet-5

[LeCun et al., 1998]



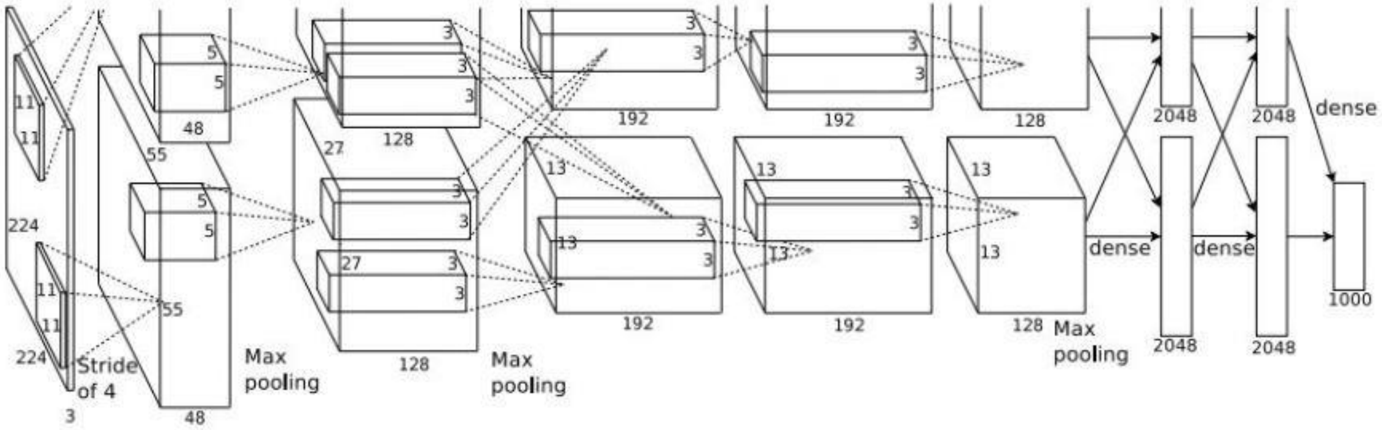
Conv filters were 5x5, applied at stride 1
Subsampling (Pooling) layers were 2x2 applied at stride 2
i.e. architecture is [CONV-POOL-CONV-POOL-FC-FC]

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



Case Study: AlexNet

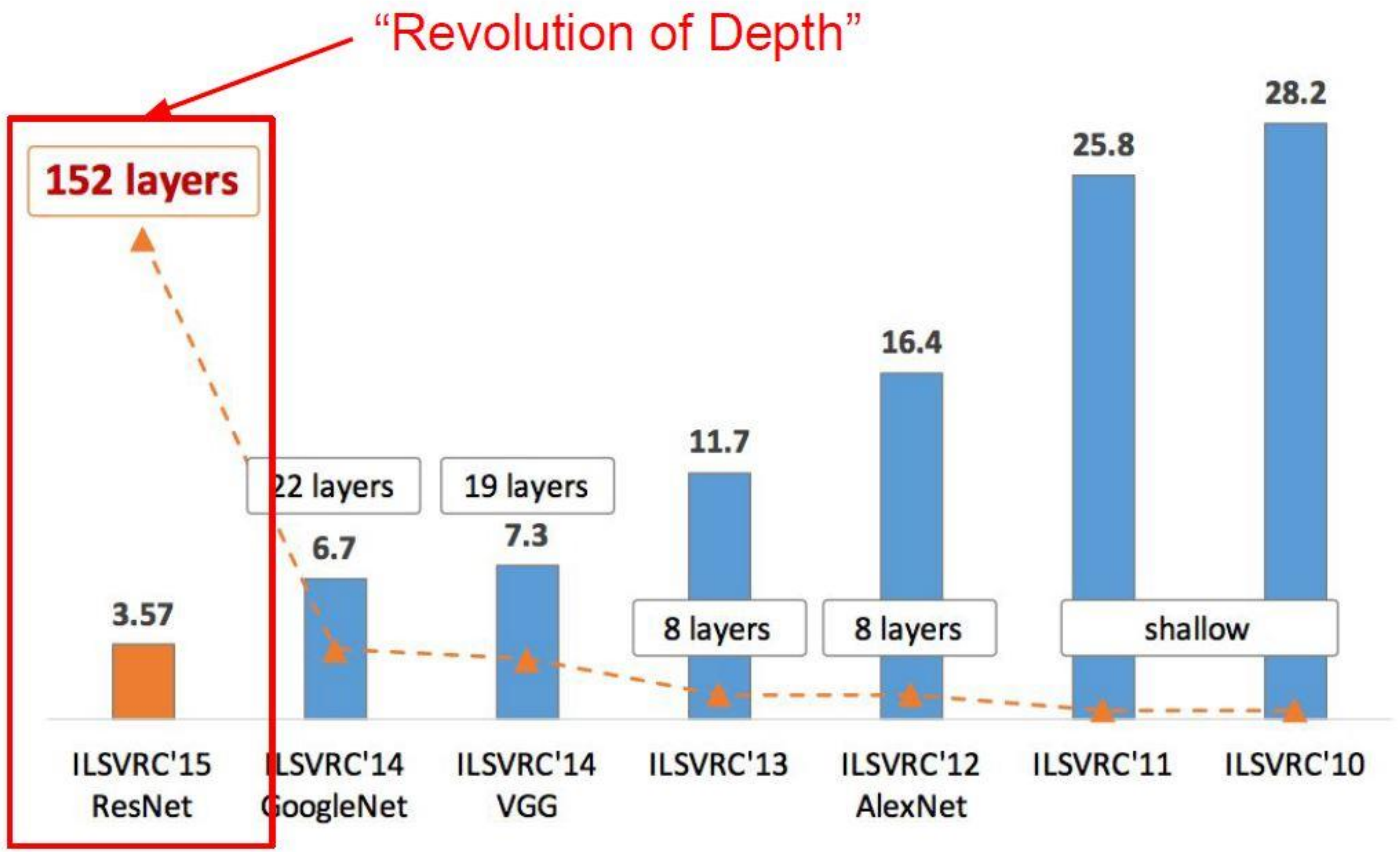
[Krizhevsky et al. 2012]



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

Details/Retrospectives:
- first use of ReLU
- used Norm layers (not common anymore)
- heavy data augmentation
- dropout 0.5
- batch size 128
- SGD Momentum 0.9
- Learning rate 1e-2, reduced by 10 manually when val accuracy plateaus
- L2 weight decay 5e-4
- 7 CNN ensemble: 18.2% -> 15.4%

ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners

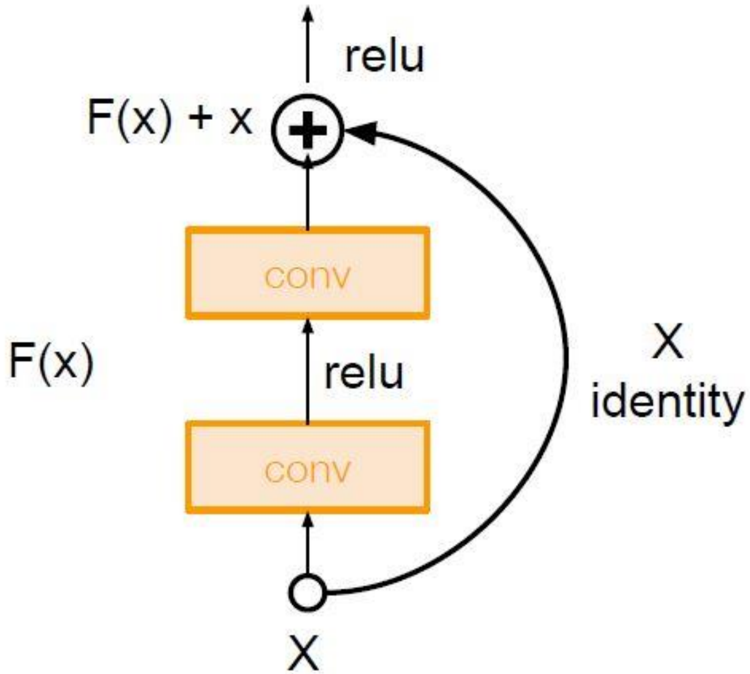


Case Study: ResNet

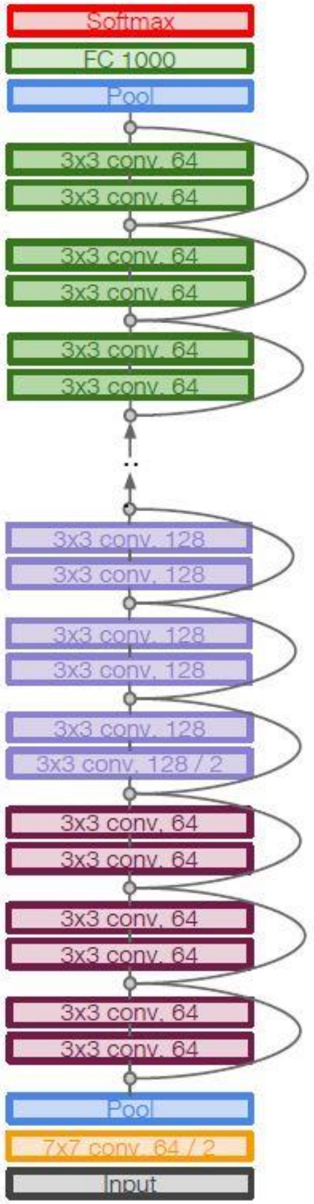
[He et al., 2015]

Very deep networks using residual connections

- 152-layer model for ImageNet
- ILSVRC'15 classification winner (3.57% top 5 error)
- Swept all classification and detection competitions in ILSVRC'15 and COCO'15!



Residual block



Case Study: ResNet

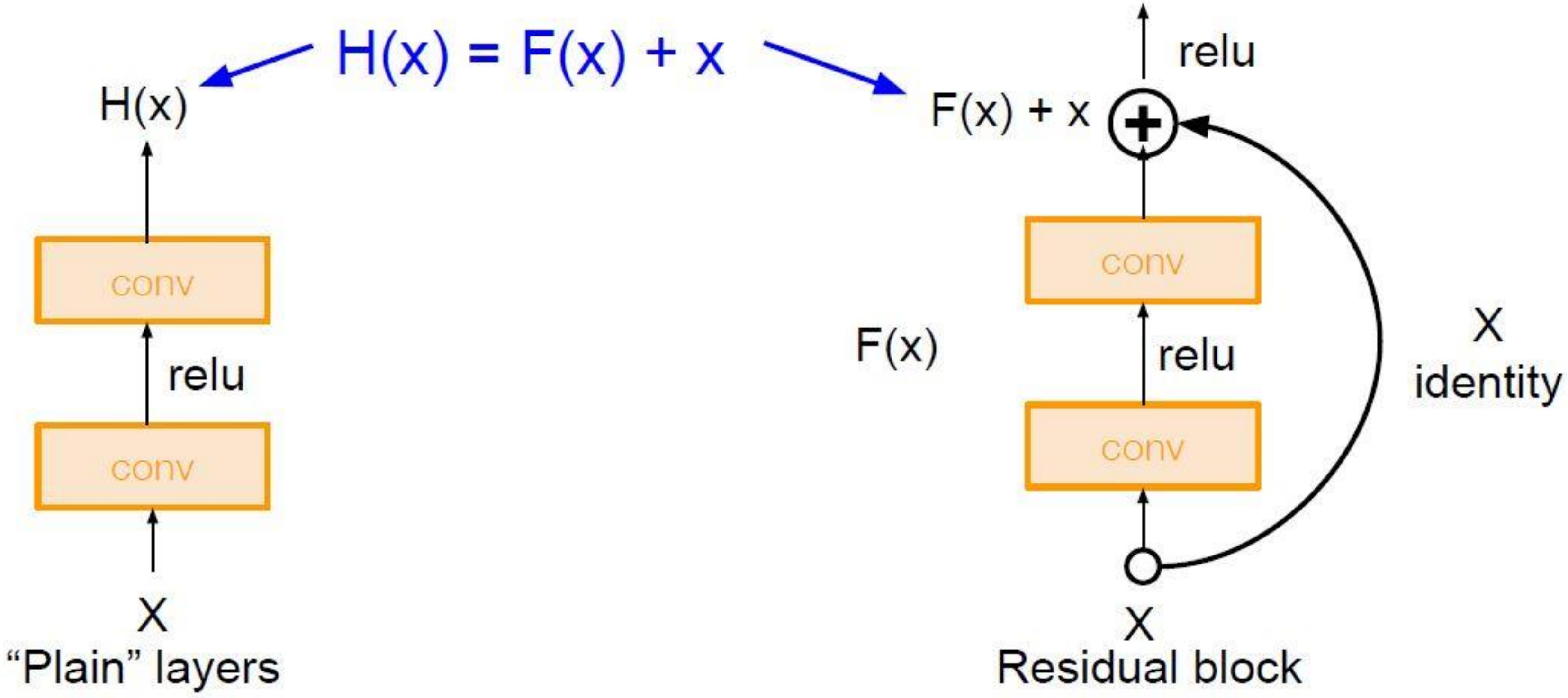
[He et al., 2015]

What happens when we continue stacking deeper layers on a “plain” convolutional neural network?

Case Study: ResNet

[He et al., 2015]

Solution: Use network layers to fit a residual mapping instead of directly trying to fit a desired underlying mapping

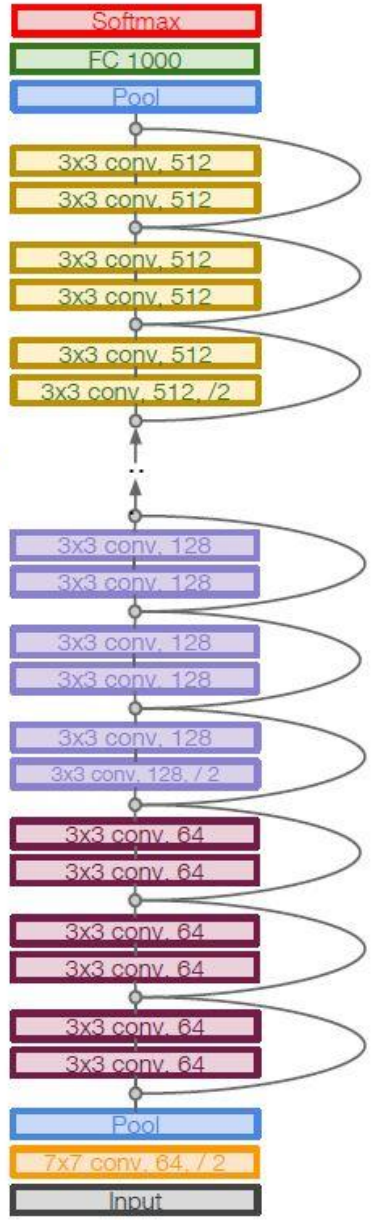


Use layers to fit residual $F(x) = H(x) - x$ instead of $H(x)$ directly

Case Study: ResNet

[He et al., 2015]

Total depths of 34, 50, 101, or 152 layers for ImageNet



Best paper award

DENSELY CONNECTED CONVOLUTIONAL NETWORKS

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

Cornell University

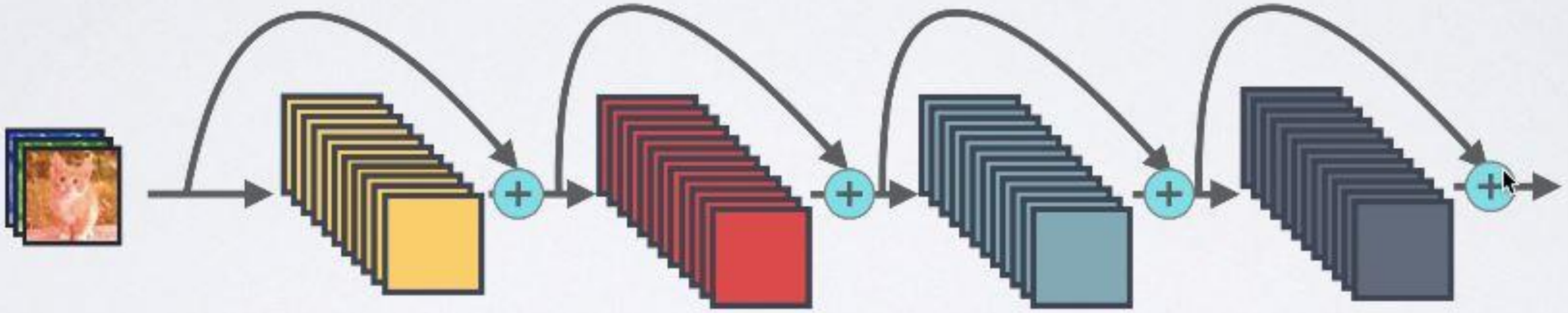
Tsinghua University

Facebook AI Research

CVPR 2017

RESNET CONNECTIVITY

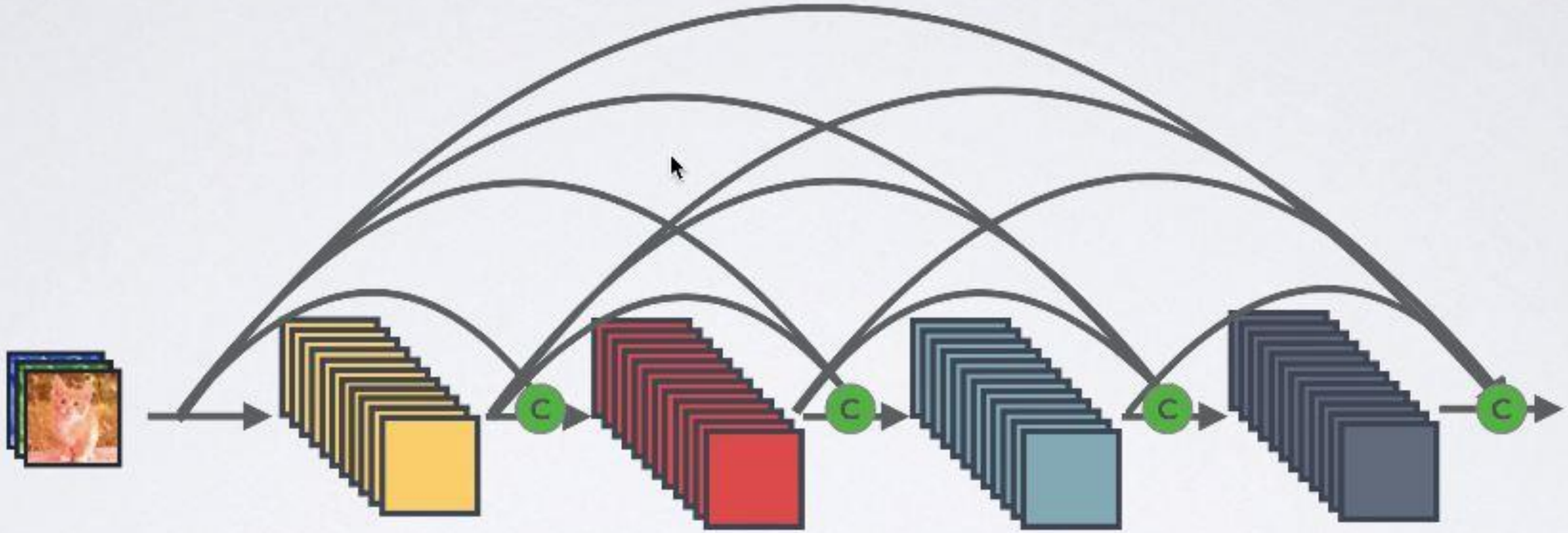
Identity mappings promote gradient propagation.



 : Element-wise addition

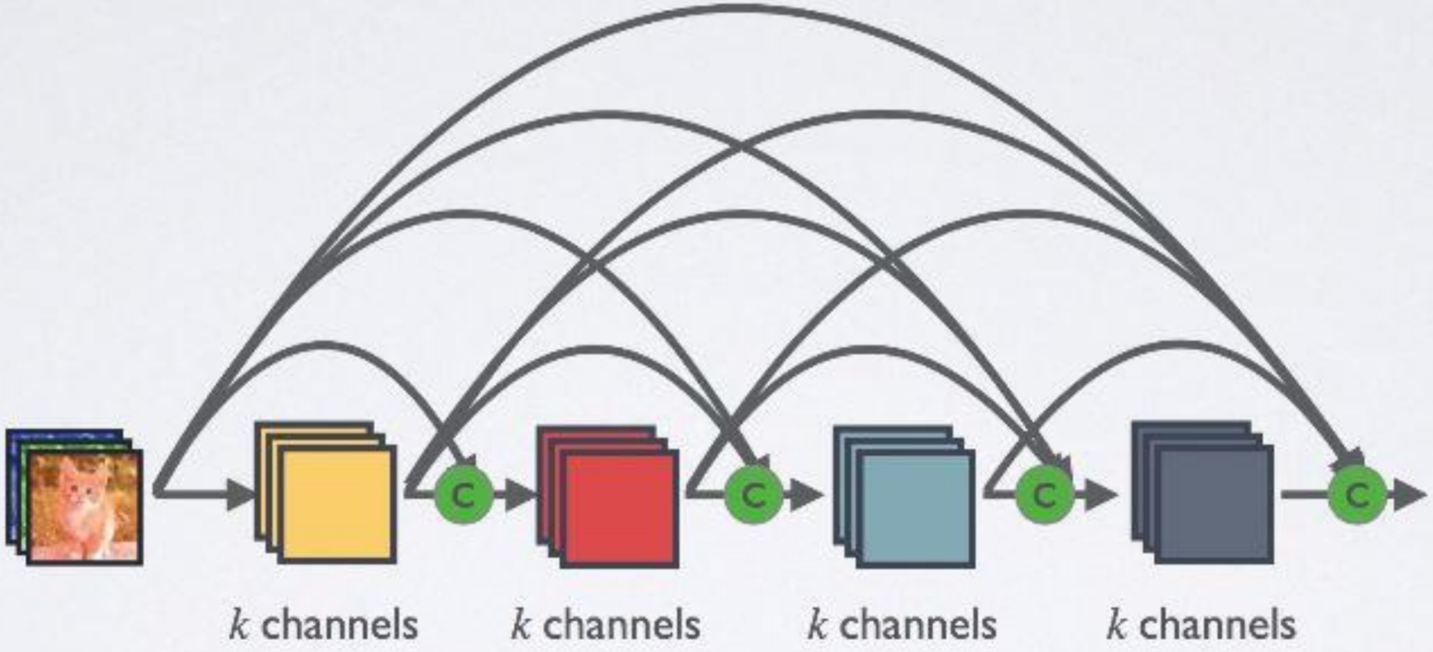
Deep residual learning for image recognition: [He, Zhang, Ren, Sun] (CVPR 2015

DENSE CONNECTIVITY



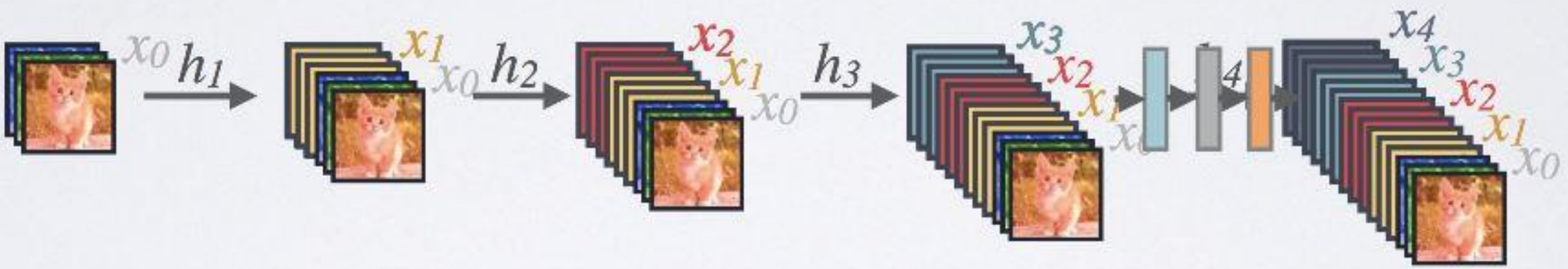
C : Channel-wise concatenation

DENSE AND SLIM

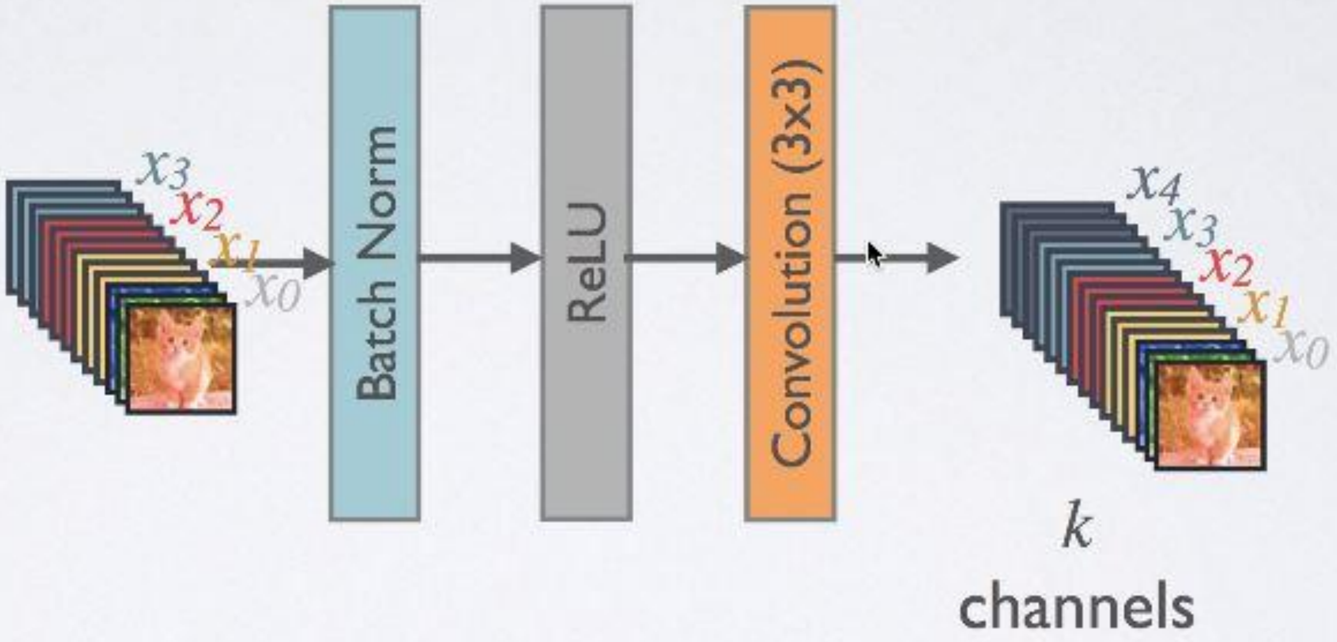


k : Growth Rate

FORWARD PROPAGATION

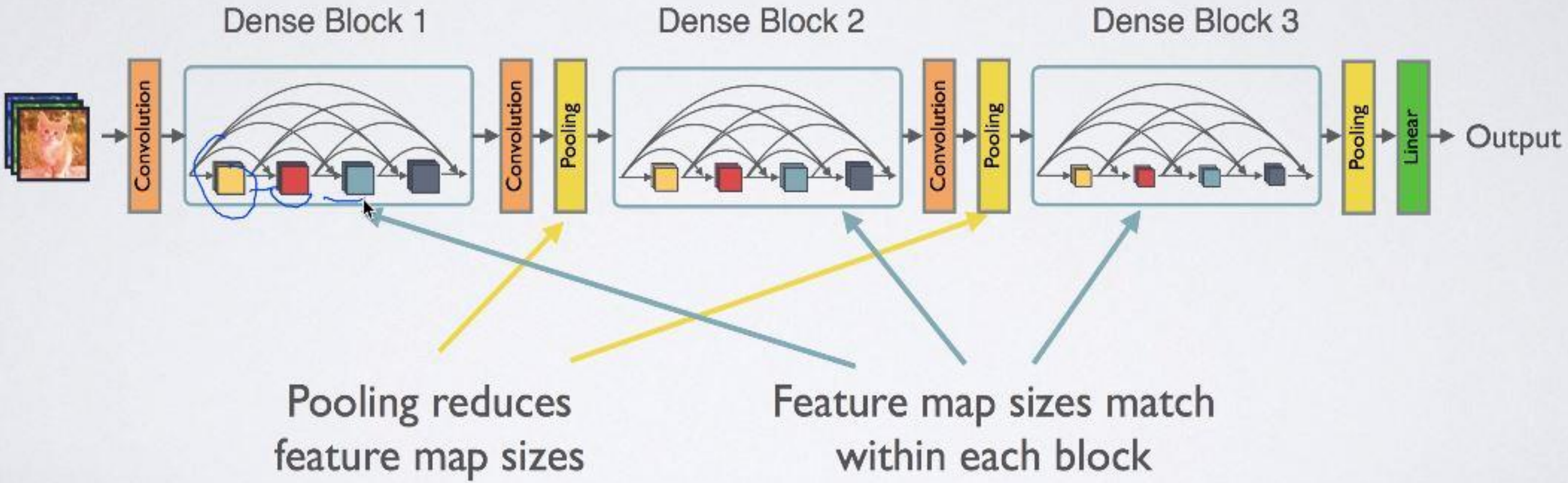


COMPOSITE LAYER IN DENSENET



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

DENSENET

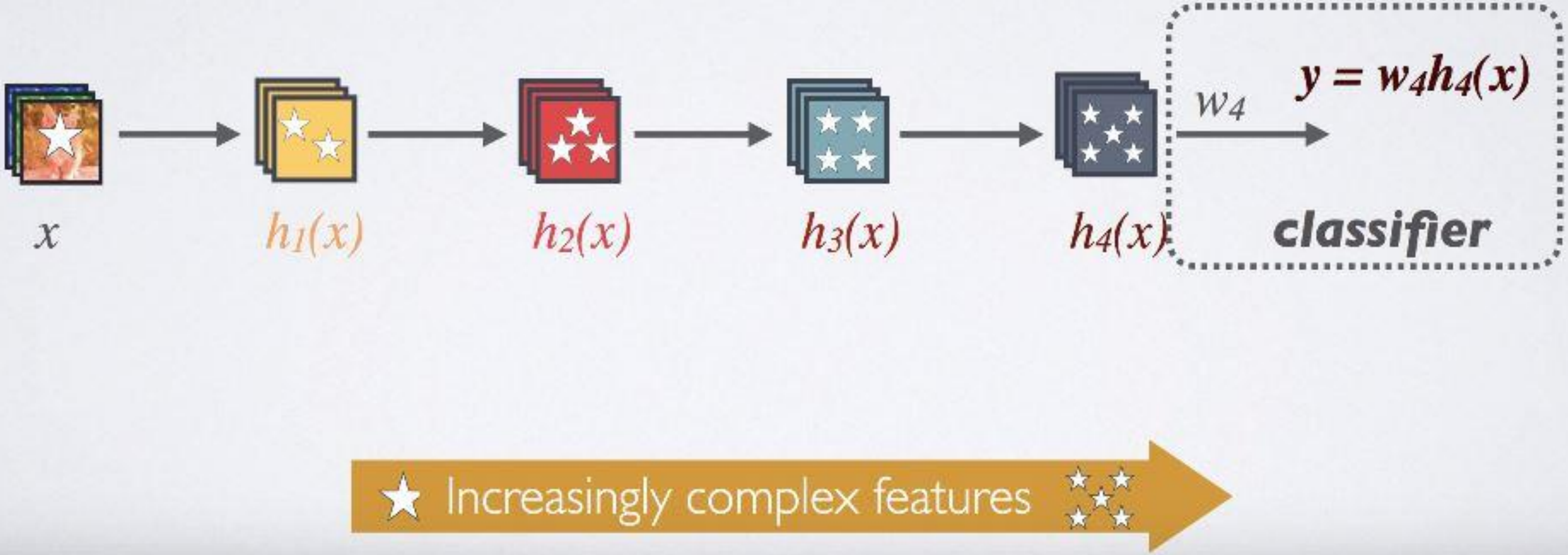


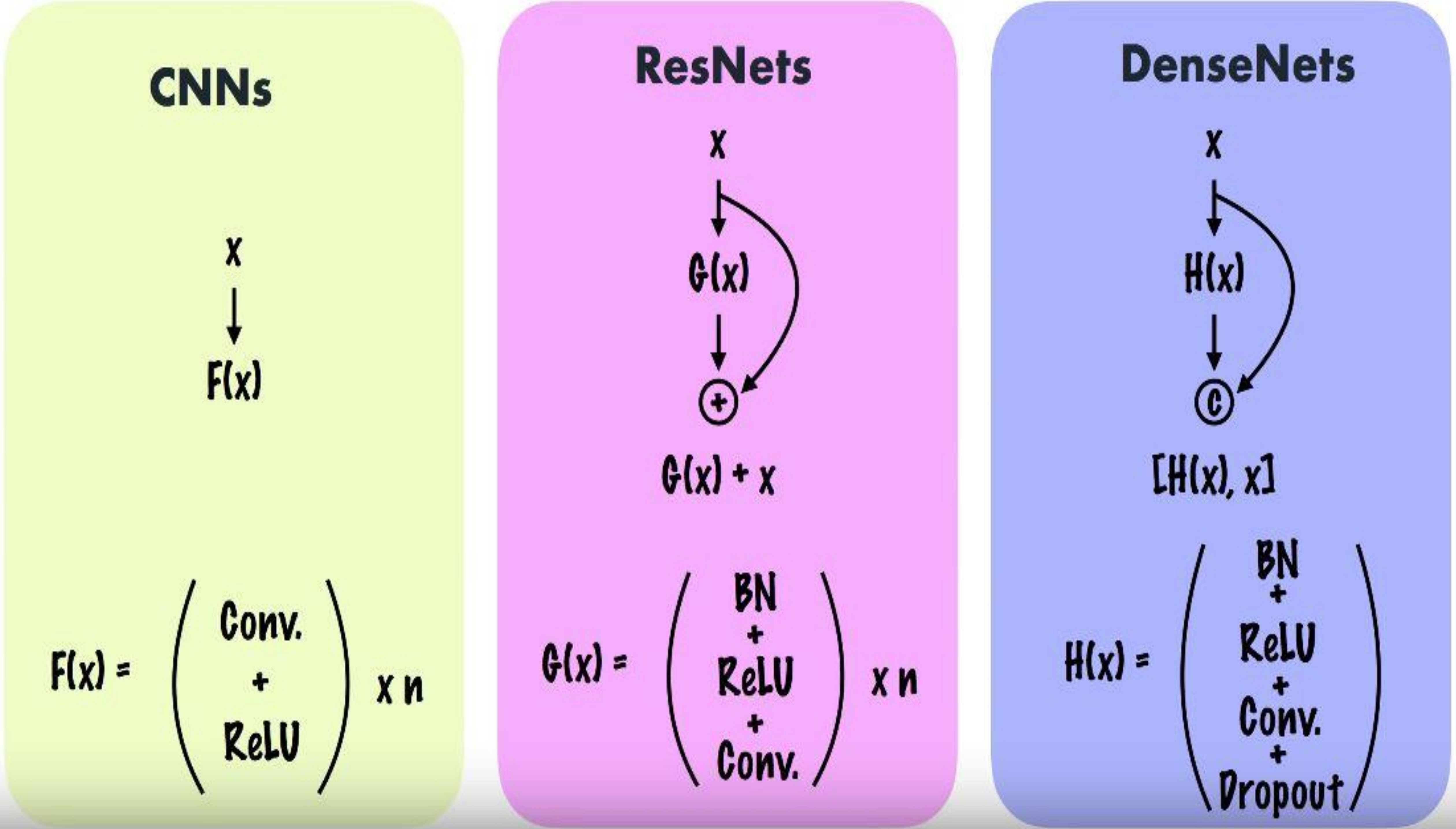
Dense Block 1
Dense Block 2
Dense Block 3
Convolution
Convolution
Pooling
Convolution
Pooling
Pooling
Linear
Output
Pooling reduces feature map sizes
Feature map sizes match within each block

ADVANTAGE 3: MAINTAINS LOW COMPLEXITY FEATURES

Standard Connectivity:

Classifier uses most complex (high level) features




CNNs
x
F(x)
F(x) = (Conv. + ReLU) x n
ResNets
x
G(x)
+
G(x) + x
G(x) = (BN + ReLU + Conv.) x n
DenseNets
x
H(x)
c
[H(x), x]
H(x) = (BN + ReLU + Conv. + Dropout)

Deep Learning for Visual Question Answering

VQA(visual question answering)

What is the mustache made of?

AI System

bananas

VQA Real Image Challenge (Open-Ended) 2017
Organized by: VQA Team

Overview · Evaluation · Phases · Participate · Leaderboard

VQA Real Image Challenge (Open-Ended) 2017

Recent progress in computer vision and natural language processing has demonstrated that lower-level tasks are much closer to being solved. We believe that the time is ripe to pursue higher-level tasks, one of which is Visual Question Answering (VQA), where the goal is to be able to understand the semantics of scenes well enough to be able to answer open-ended, free-form natural language questions (asked by humans) about images.

VQA Challenge 2017 is the 2nd edition of the VQA Challenge on the 2nd edition (v2.0) of the VQA dataset introduced in Goyal et al., CVPR 2017. The 1st edition of the VQA Challenge was organized in CVPR 2016 on the 1st edition (v1.0) of the VQA dataset introduced in Antol et al., ICCV 2015.

VQA v2.0 dataset is a more balanced version of VQA v1.0 which significantly reduces the language biases. VQA v2.0 is about twice the size of VQA v1.0. For almost every question in the VQA v2.0 dataset, there are two similar images that have different answers to the question.

To participate in the challenge, you can find instructions on the VQA website. In particular, please see the overview, download, evaluation and challenge pages for more details. We also provide dataset visualization and browser pages to give everyone a sense of the dataset contents.

Note: All the timings mentioned are local to your timezone.

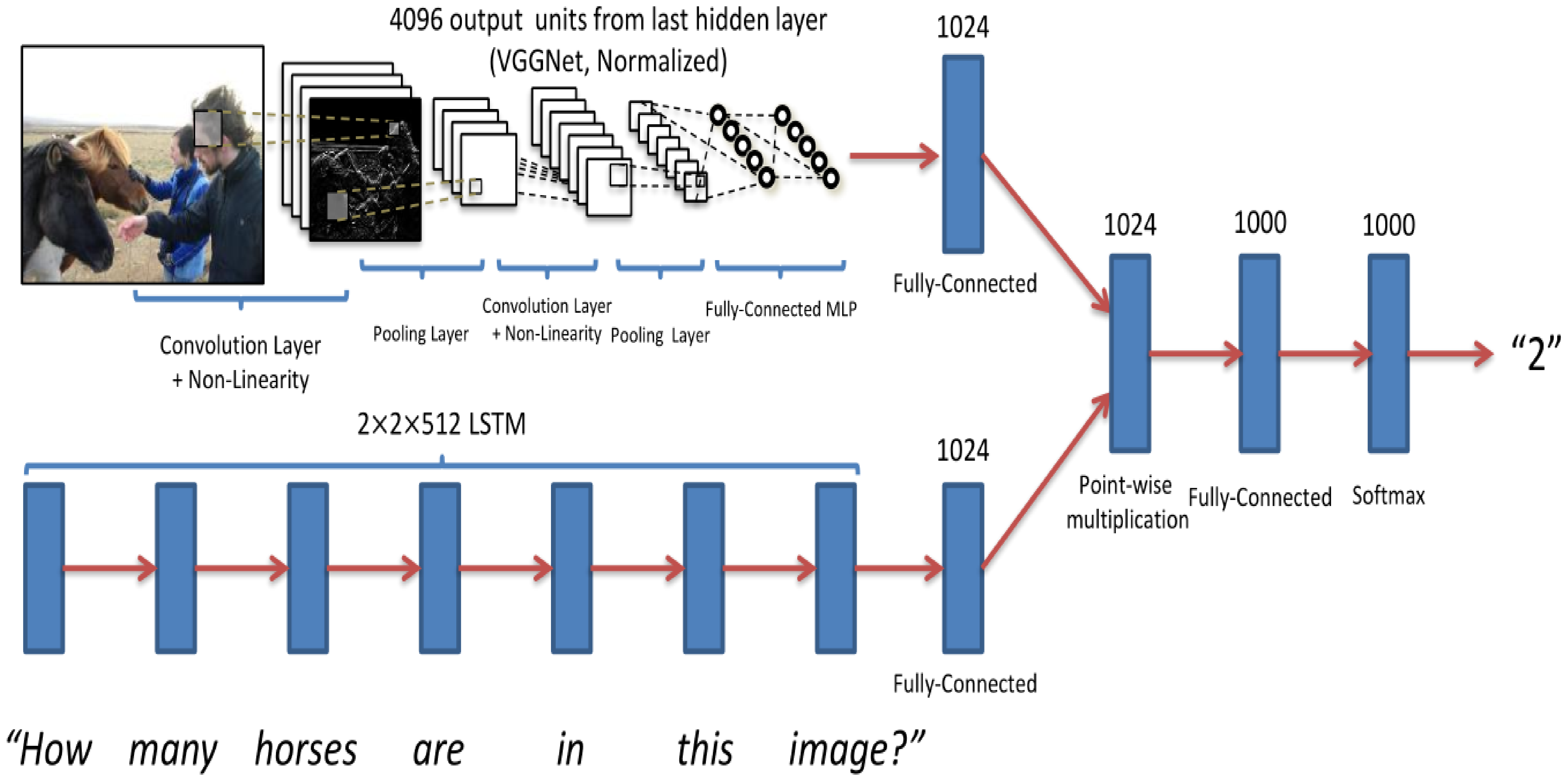
VQA Real Image Challenge (Open-Ended) 2017
Organized by: VQA Team

Overview · Evaluation · Phases · Participate · Leaderboard

Please select from following phases

Real test2017 (oe)

Rank	Participant Team	yes/no	number	other	overall
1	HDU-USYD-UNCC	86.65	51.13	61.75	70.92
2	Adelaide-Teney ACRV MSR	86.60	48.64	61.15	70.34
3	CFM-UESTC	84.08	45.40	56.95	66.94
4	LV_NUS	81.89	46.29	58.30	66.77
5	Adelaide-Teney	83.71	43.77	57.20	66.73
6	Athena	82.88	43.17	57.95	66.67
7	Tohoku CV	83.31	45.77	56.03	66.22
8	UPMC-LIP6	82.07	41.06	57.12	65.71
9	JuneflowerIvaNlpr	81.09	41.56	57.83	65.70
10	ReasonNet_	78.86	41.98	57.39	64.61


4096 output units from last hidden layer
(VGGNet, Normalized)
1024
Convolution Layer
+ Non-Linearity
Pooling Layer
Convolution Layer
+ Non-Linearity
Pooling Layer
Fully-Connected MLP
Fully-Connected
1024
1000
1000
"2"
Point-wise
multiplication
Fully-Connected
Softmax
2×2×512 LSTM
1024
Fully-Connected
"How many horses are in this image?"

https://github.com/avisingh599/visual-qa
https://keras.io/getting-started/functional-api-guide/
https://github.com/iamaaditya/VQA_Demo/blob/master/Visual_Question_Answering_Demo_in_python_notebook.ipynb