

DEEP
LEARNING
WORKSHOP

Dublin City University
21-22 May 2018



#InsightDL2018

Day 1 Lecture 4

Image Classification on ImageNet



Xavier Giro-i-Nieto

xavier.giro@upc.edu

Associate Professor

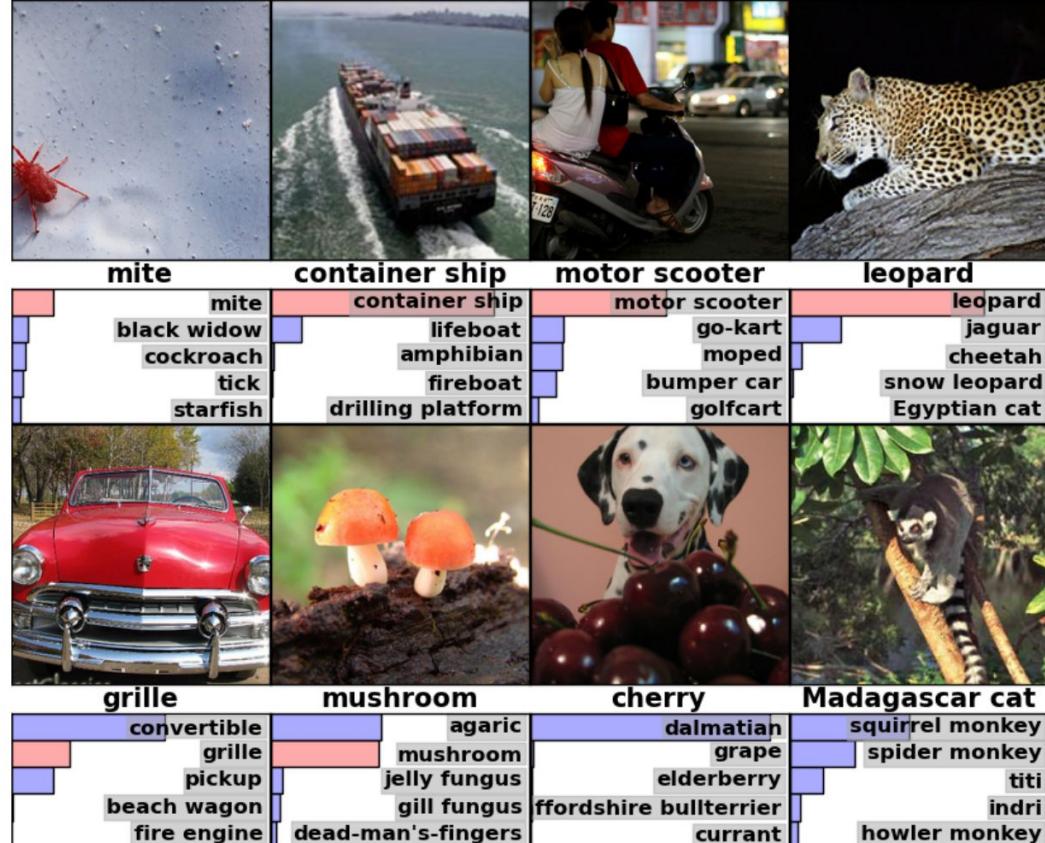
Intelligent Data Science and Artificial Intelligence Center
Universitat Politècnica de Catalunya (UPC)

ImageNet Challenge

bit.ly/InsightDL2018
#InsightDL2018

IMAGENET

- 1,000 object classes (categories).
- Images:
 - 1.2 M train
 - 100k test.





ImageNet Challenge

IMAGENET



Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "[Imagenet large scale visual recognition challenge.](#)" International Journal of Computer Vision 115, no. 3 (2015): 211-252. [\[web\]](#)

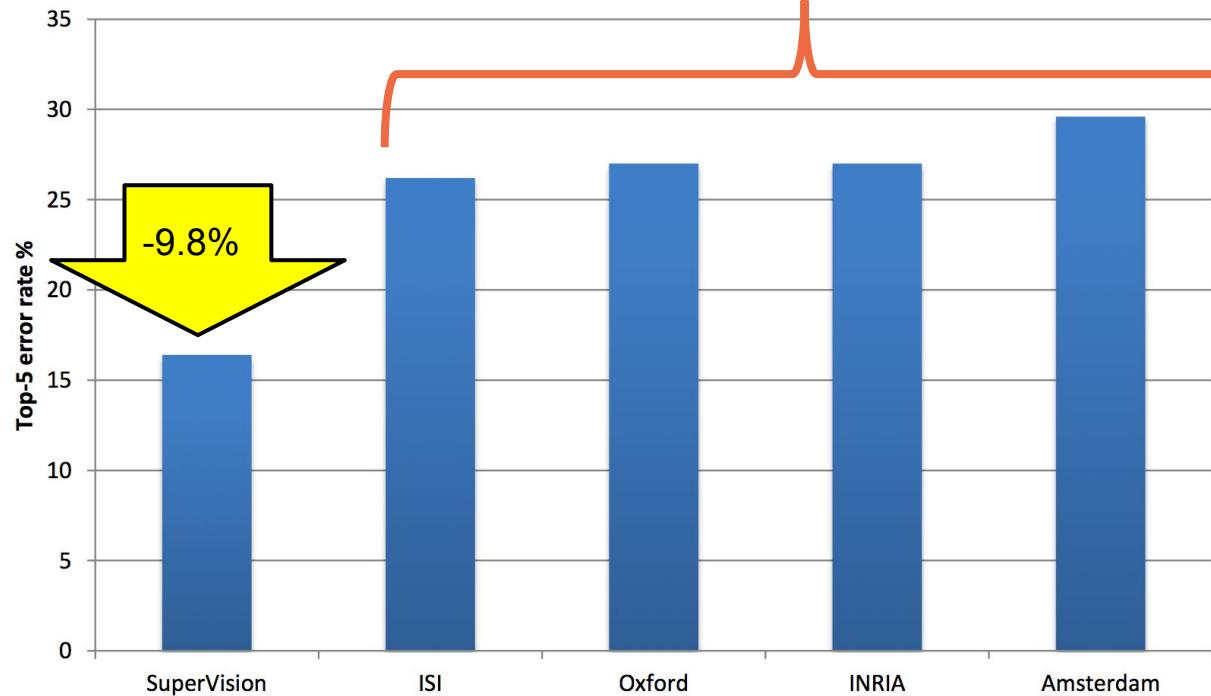


ImageNet Challenge: 2012



Slide credit:
[Rob Fergus](#) (NYU)

Based on SIFT + Fisher Vectors



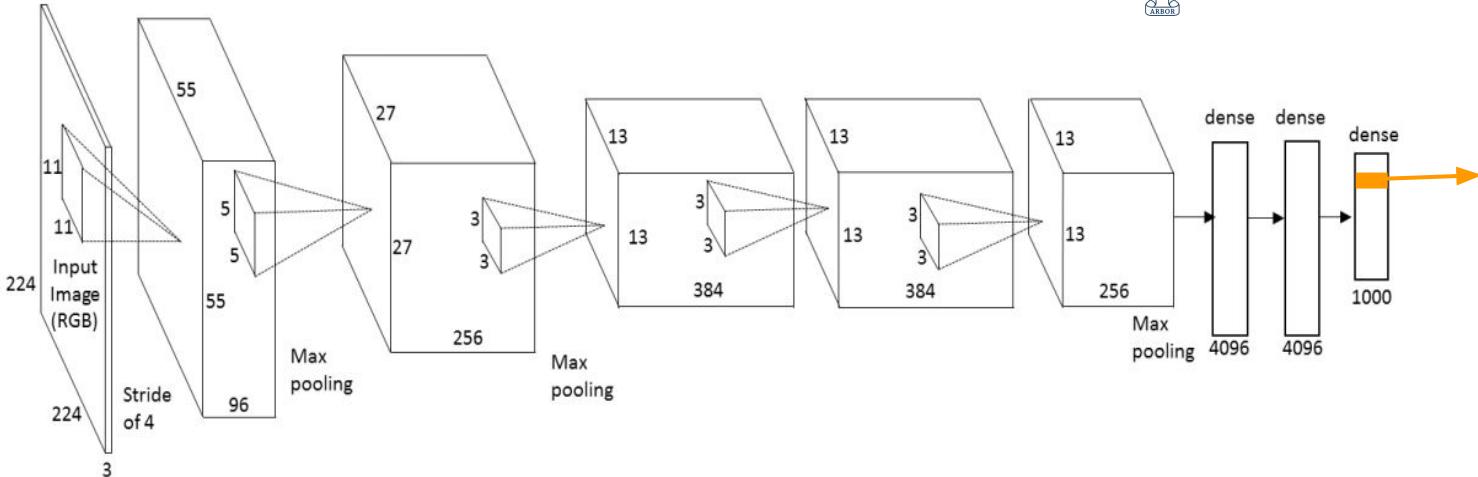
Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "[Imagenet large scale visual recognition challenge.](#)" International Journal of Computer Vision 115, no. 3 (2015): 211-252. [\[web\]](#)



AlexNet (SuperVision)



UNIVERSITY OF
TORONTO

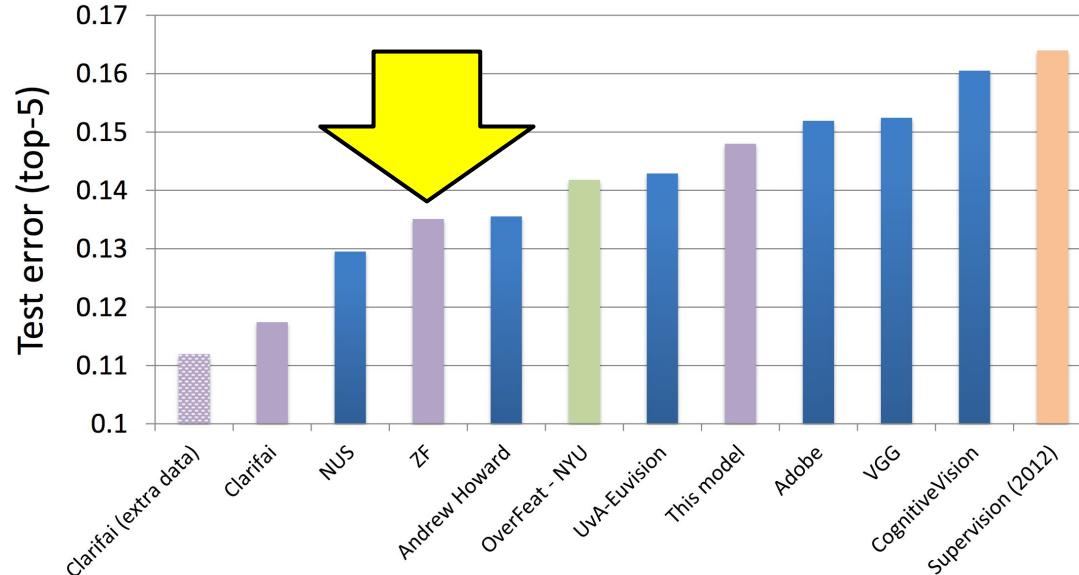


Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "[Imagenet classification with deep convolutional neural networks.](#)" NIPS 2012

ImageNet Challenge: 2013



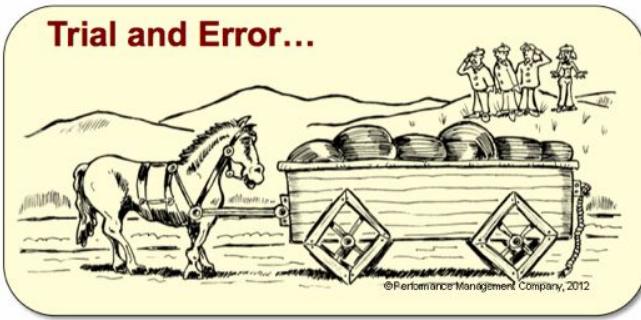
ImageNet Classification 2013



Slide credit:
[Rob Fergus](#) (NYU)

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "[Imagenet large scale visual recognition challenge.](#)" International Journal of Computer Vision 115, no. 3 (2015): 211-252. [\[web\]](#)

Zeiler-Fergus (ZF)



The development of better convnets is reduced to trial-and-error.



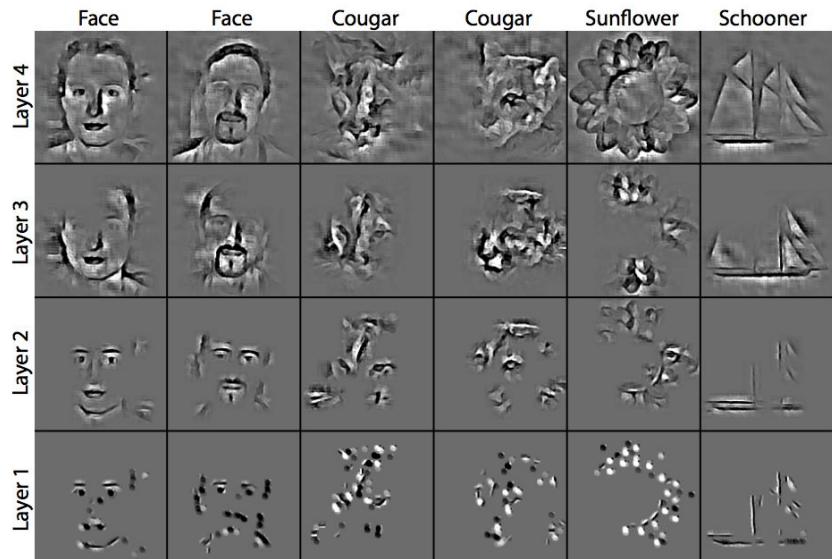
Visualization can help in proposing better architectures.

Zeiler-Fergus (ZF)

“A convnet model that uses the same components (filtering, pooling) but in reverse, so instead of mapping pixels to features does the opposite.”

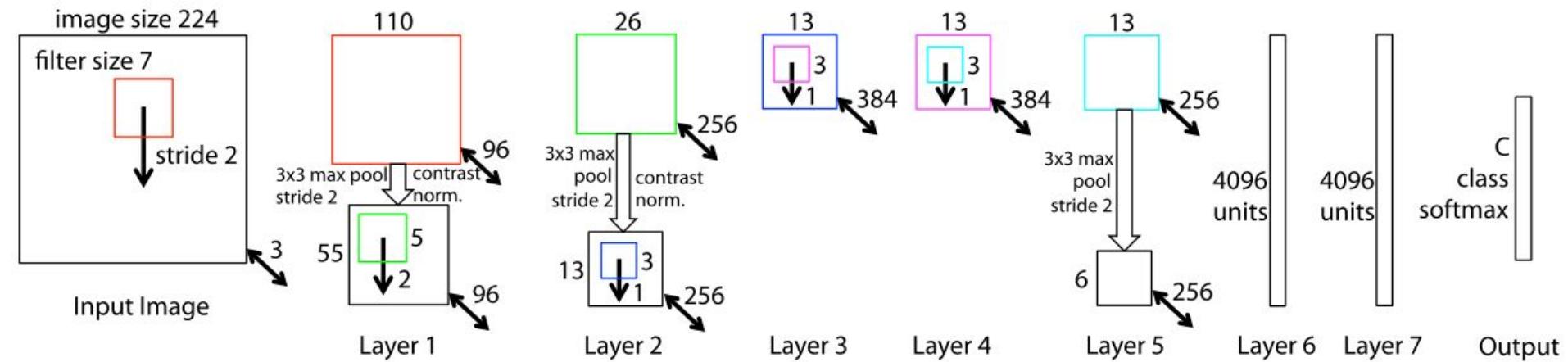


NEW YORK UNIVERSITY





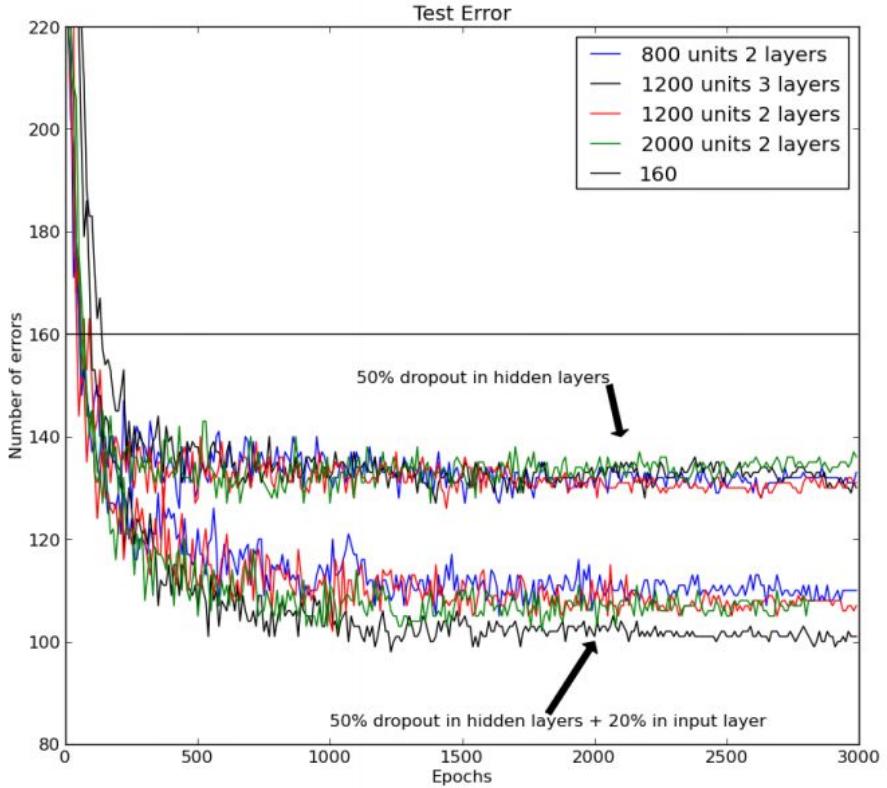
Zeiler-Fergus (ZF)





Zeiler-Fergus (ZF)

Regularization with more dropout: introduced in the input layer.

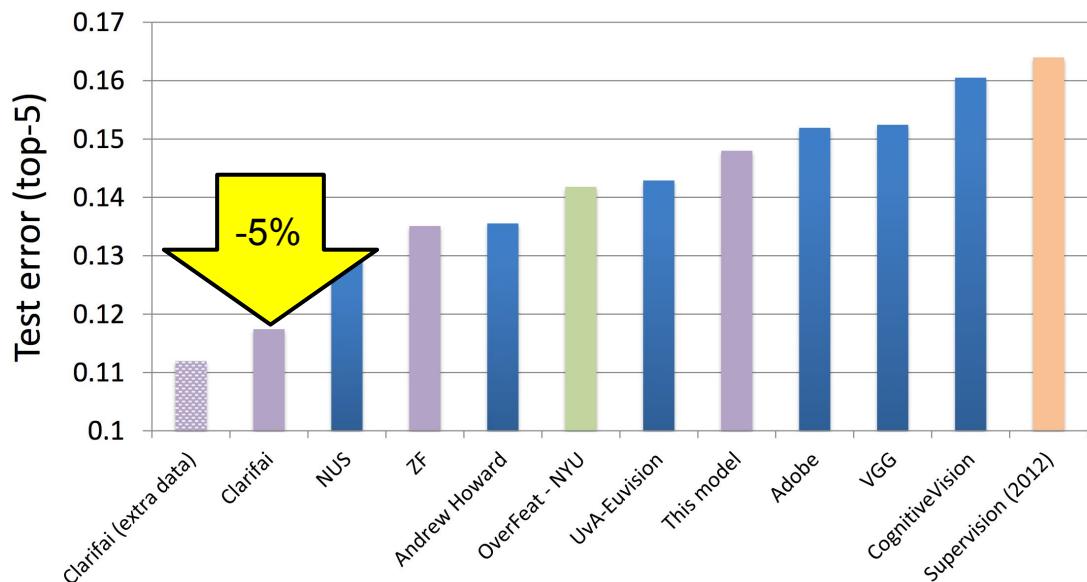


ImageNet Challenge: 2013



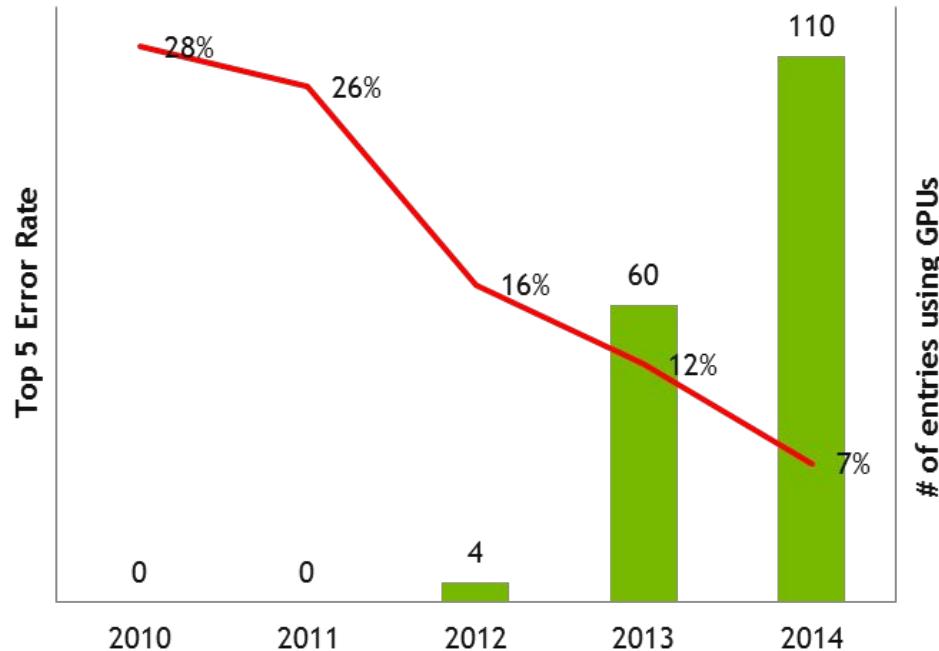
clarifai

ImageNet Classification 2013





ImageNet Challenge: 2014



ImageNet Challenge: 2014

image

2018

conv-64

2018

conv-64

maxpool

conv-128

conv-128

maxpool

conv-256

conv-256

conv-256

conv-256

maxpool

conv-512

conv-512

conv-512

conv-512

maxpool

conv-512

conv-512

conv-512

maxpool

FC-4096

FC-4096

FC-1000

softmax

AlexNet

image

conv-64

conv-192

conv-384

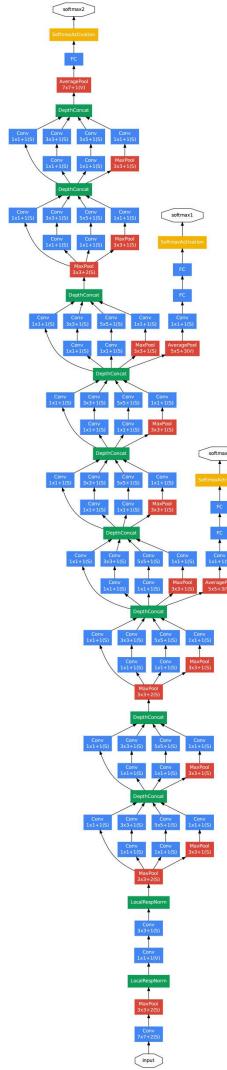
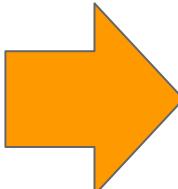
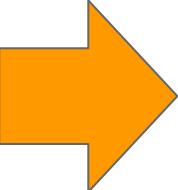
conv-256

conv-256

FC-4096

FC-4096

FC-1000



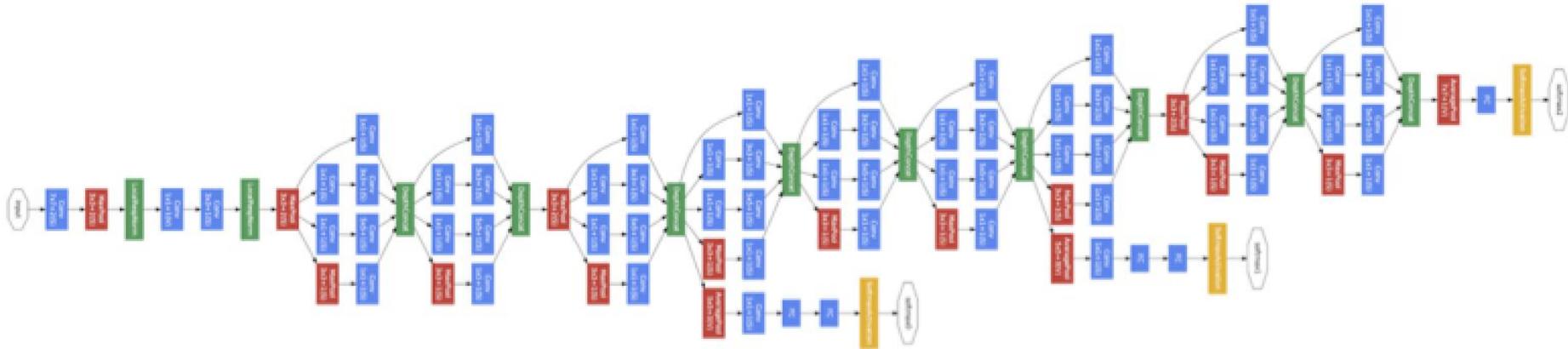
GoogleNet (Inception)





GoogleNet (Inception)

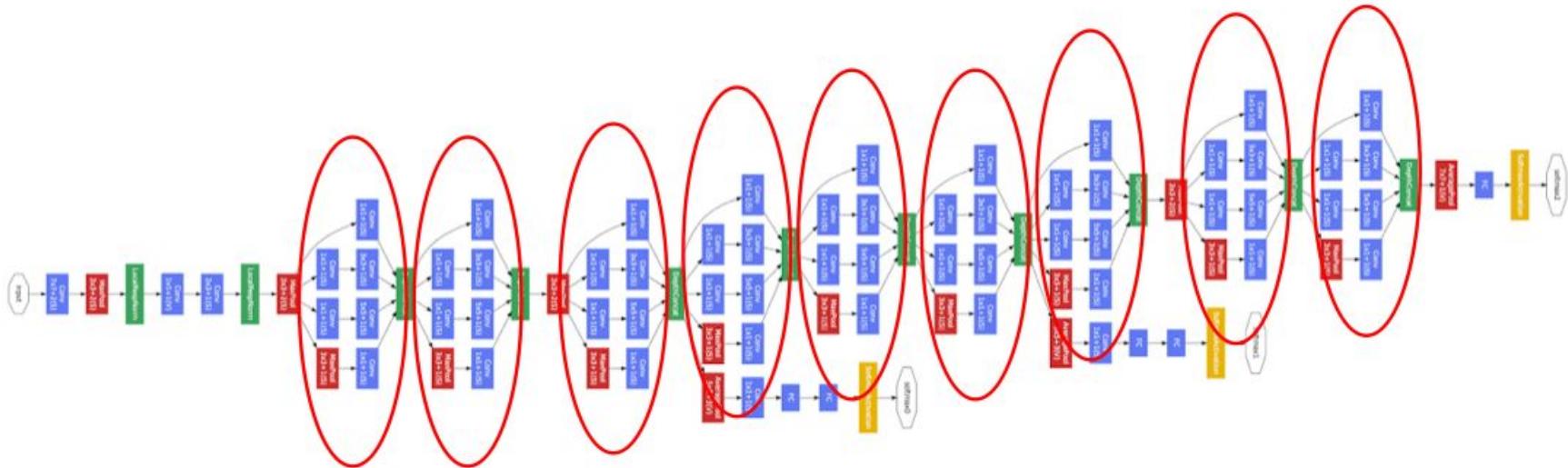
22 layers !



Convolution
Pooling
Softmax
Other

Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "[Going deeper with convolutions.](#)" CVPR 2015. [\[video\]](#) [\[slides\]](#) [\[poster\]](#)

GoogleNet (Inception)



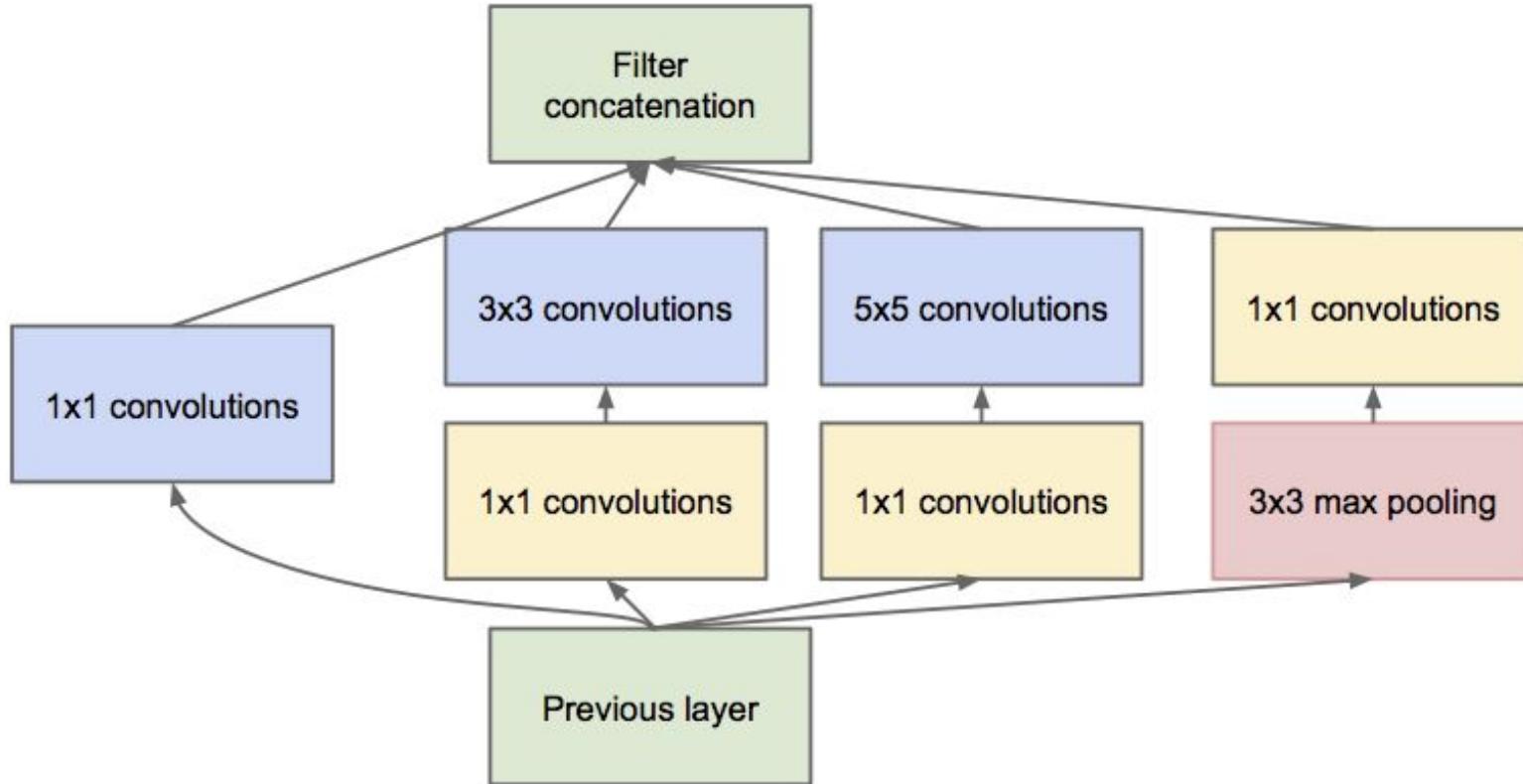
9 Inception modules

Network in a network in a network...

Convolution
Pooling
Softmax
Other

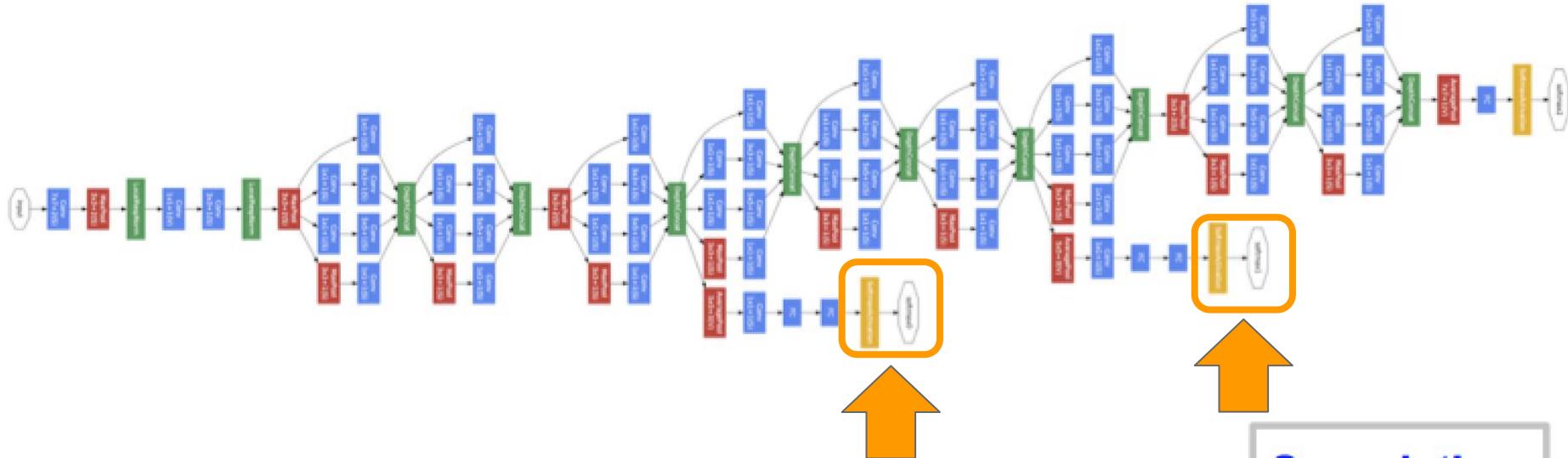


GoogleNet (Inception)



GoogleNet (Inception)

Two Softmax Classifiers at intermediate layers combat the vanishing gradient while providing regularization at training time.

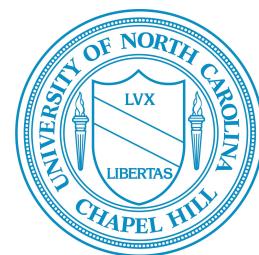


...and no fully connected layers needed
(12 times fewer parameters than AlexNet. !)

Convolution
Pooling
Softmax
Other



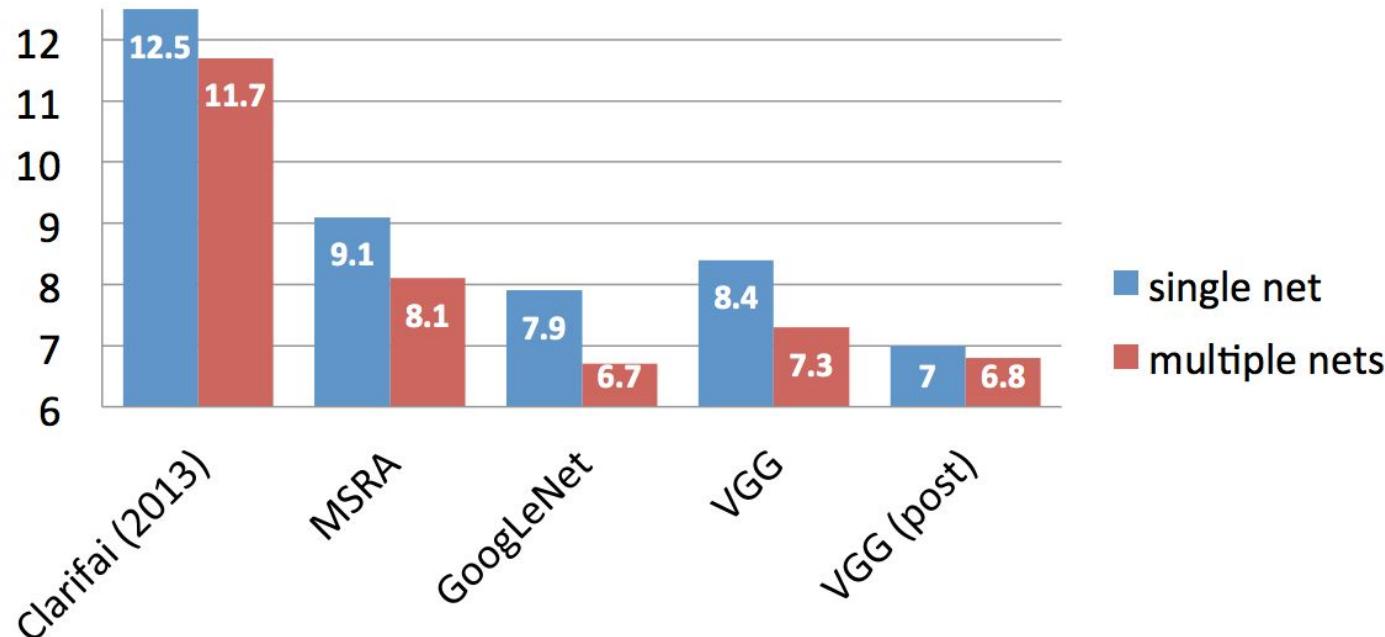
GoogleNet (Inception)





Simonyan, Karen, and Andrew Zisserman. "[Very deep convolutional networks for large-scale image recognition.](#)" ICLR 2015. [\[video\]](#) [\[slides\]](#) [\[project\]](#)

Top-5 Classification Error (Test Set)

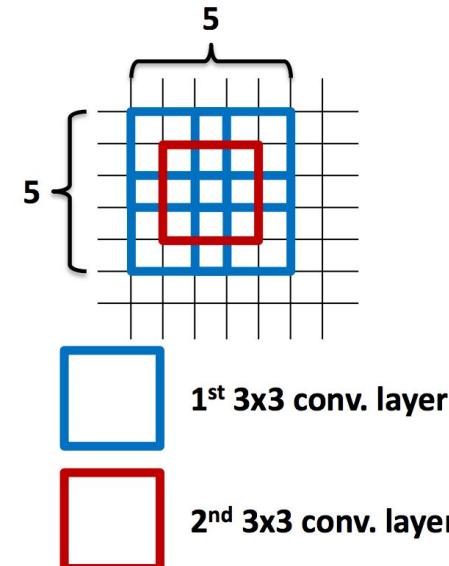


Simonyan, Karen, and Andrew Zisserman. "[Very deep convolutional networks for large-scale image recognition.](#)" ICLR 2015. [\[video\]](#) [\[slides\]](#) [\[project\]](#)

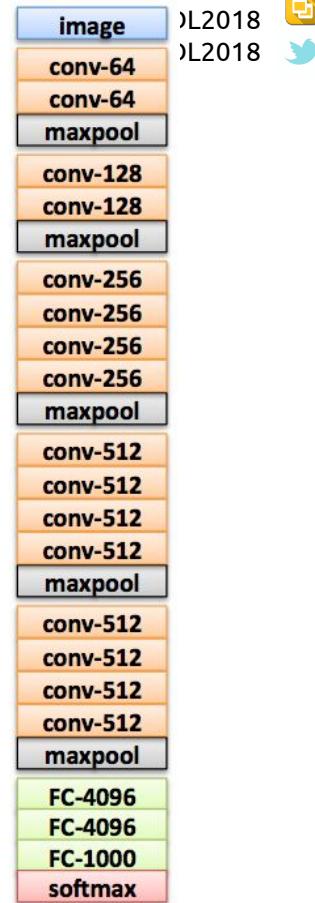
VGG: Stacked 3x3 convolutions

Why 3x3 layers?

- Stacked conv. layers have a large receptive field
 - two 3x3 layers – 5x5 receptive field
 - three 3x3 layers – 7x7 receptive field
- More non-linearity
- Less parameters to learn
 - ~140M per net



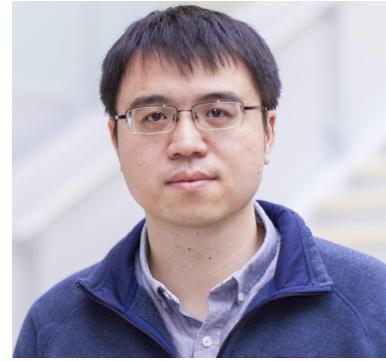
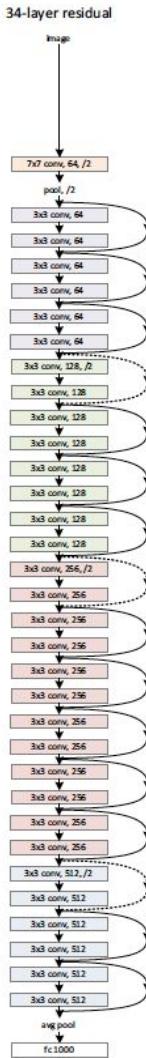
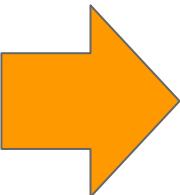
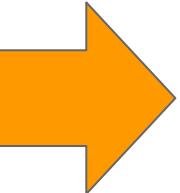
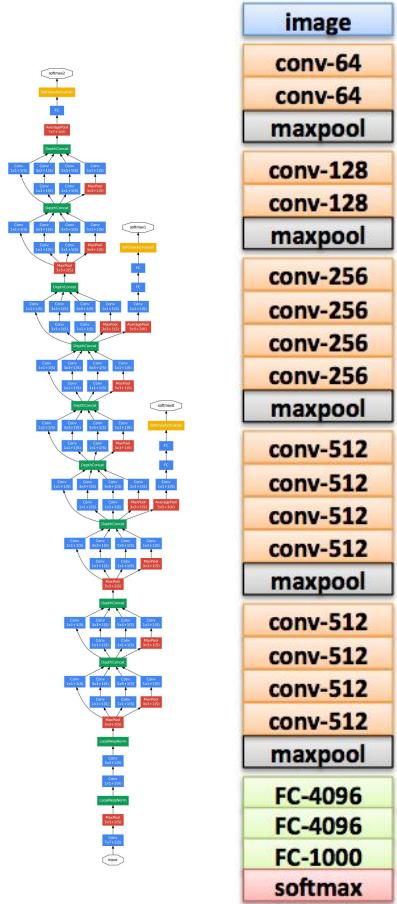
VGG: Stacked 3x3 convolutions



- No poolings between some convolutional layers.
- Convolution strides of 1 (no skipping).

Simonyan, Karen, and Andrew Zisserman. "[Very deep convolutional networks for large-scale image recognition.](#)" ICLR 2015. [\[video\]](#) [\[slides\]](#) [\[project\]](#)

ImageNet Challenge: 2015

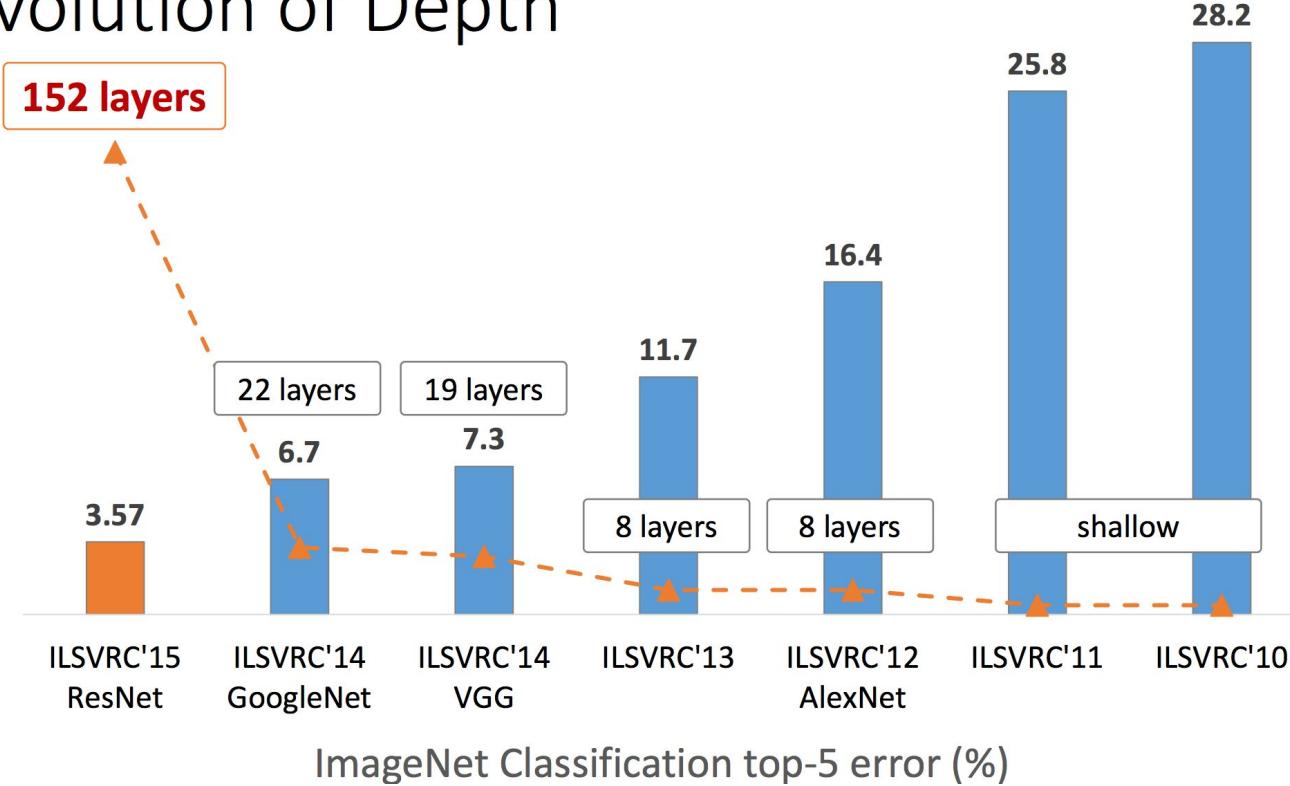


Microsoft
Research

3.6% top 5 error...
with 152 layers !!

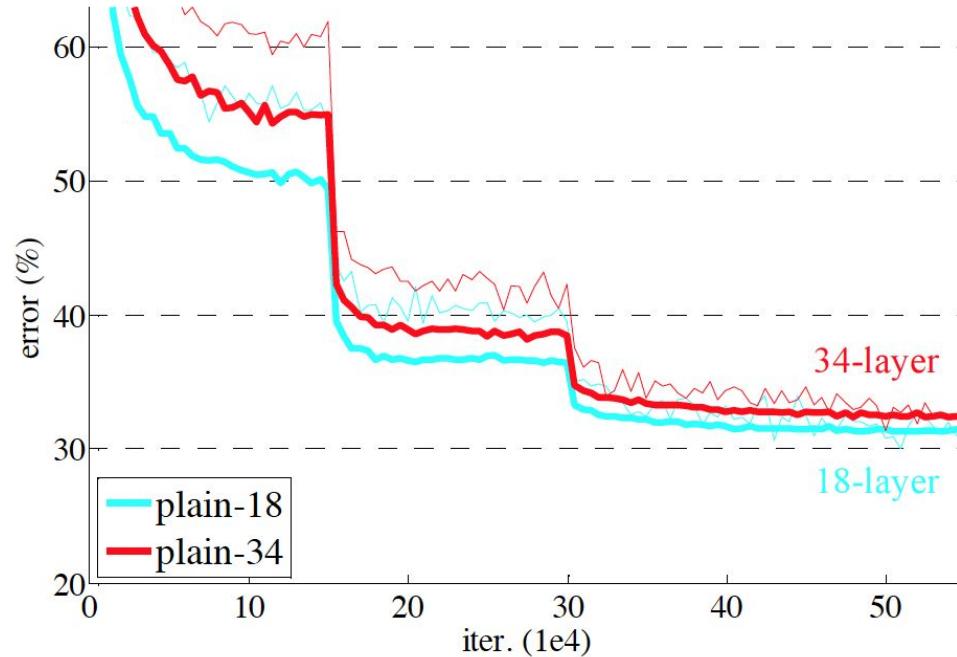
ResNet

Revolution of Depth



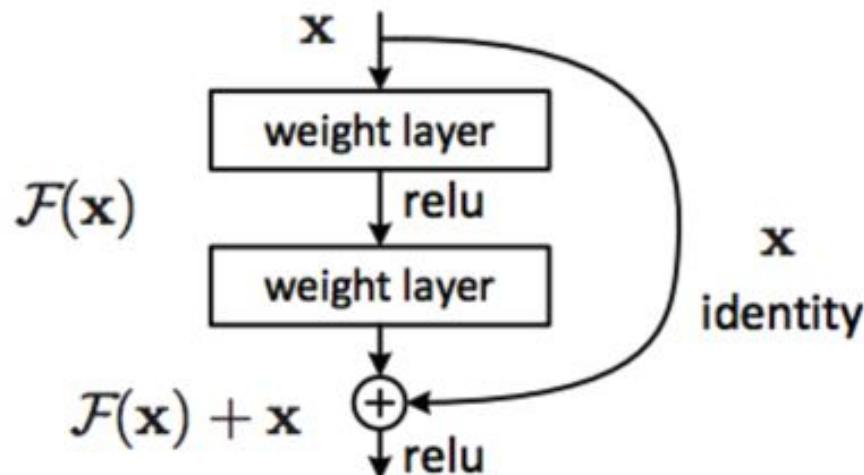
ResNet

Deeper networks (34 is deeper than 18) are more difficult to train.



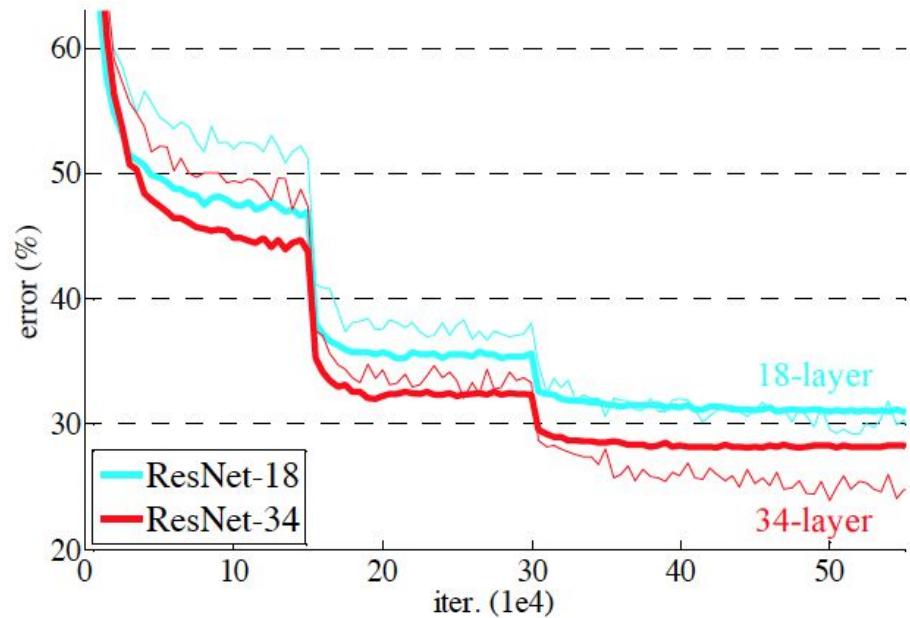
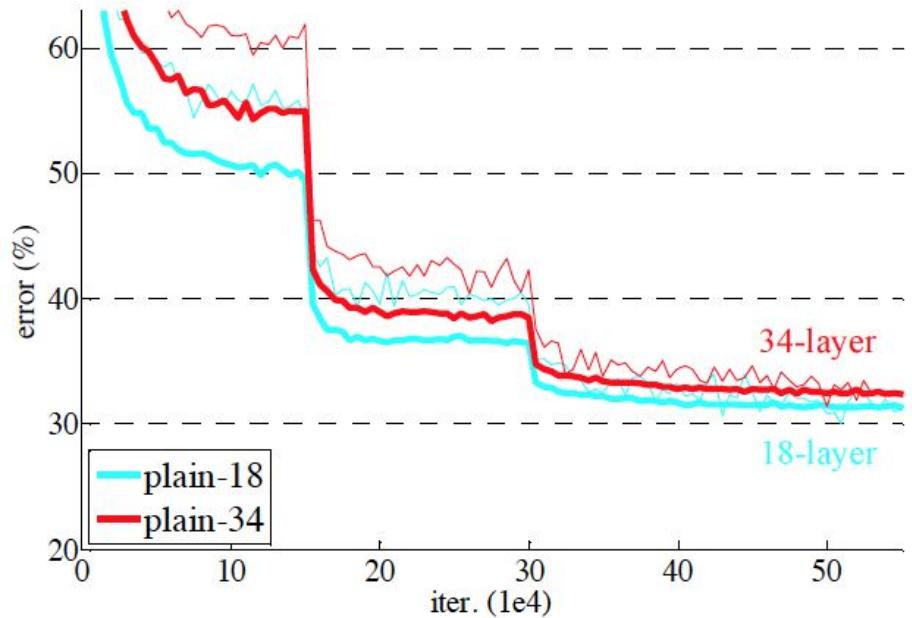
ResNet

Residual learning: reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions



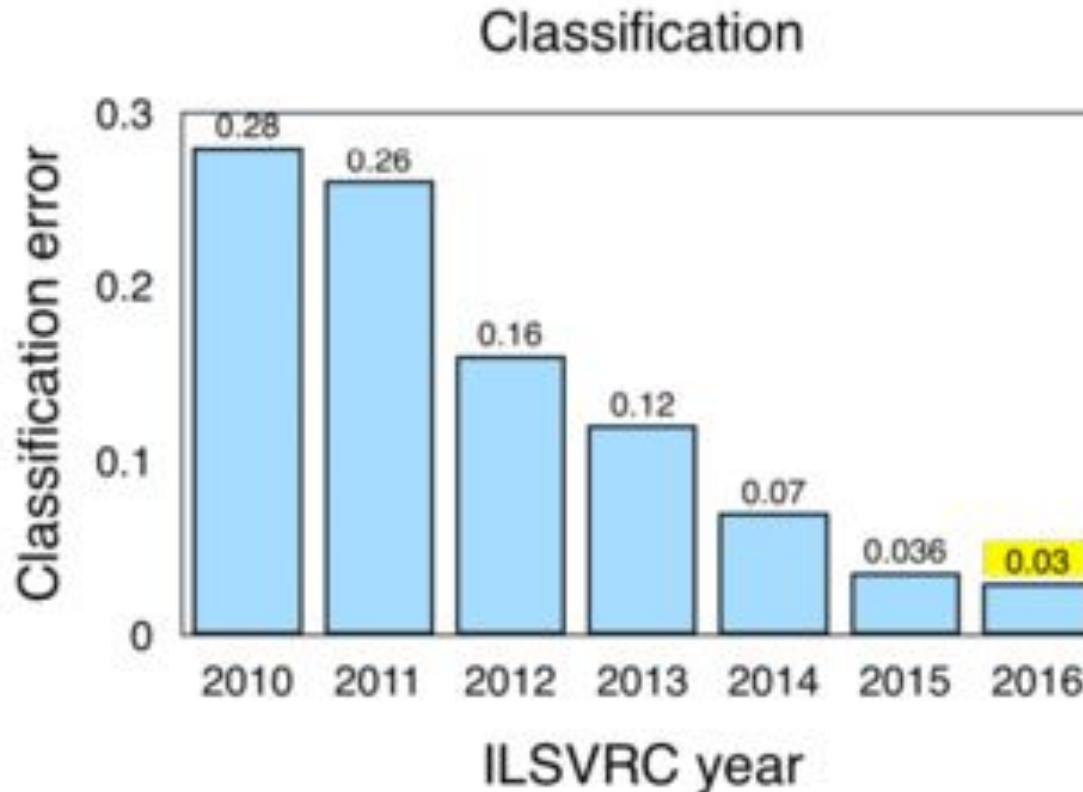


ResNet





ImageNet Challenge: 2016





ImageNet Challenge: 2016

This is an archived post. You won't be able to vote or comment.

Large Scale Visual Recognition Challenge 2016 - Results finally available (image-net.org)
enviat fa 1 any per DrPharael
28 comentaris comarteix desa amaga give gold reporta crosspost

tots els 28 comentaris
ordenats per: millor ▼

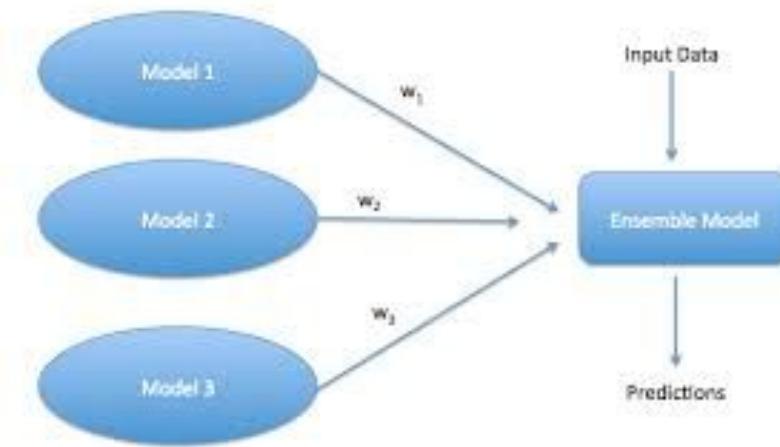
[+] BeatLeJuce 39 punts fa 1 any
TL;DR:

- No big new technologies or revolutionary architectures
- everyone uses Deep Learning
- none of the big companies care anymore (no Google, MSRA, Facebook, Baidu, ...)
- almost all competitors are from Asian organizations

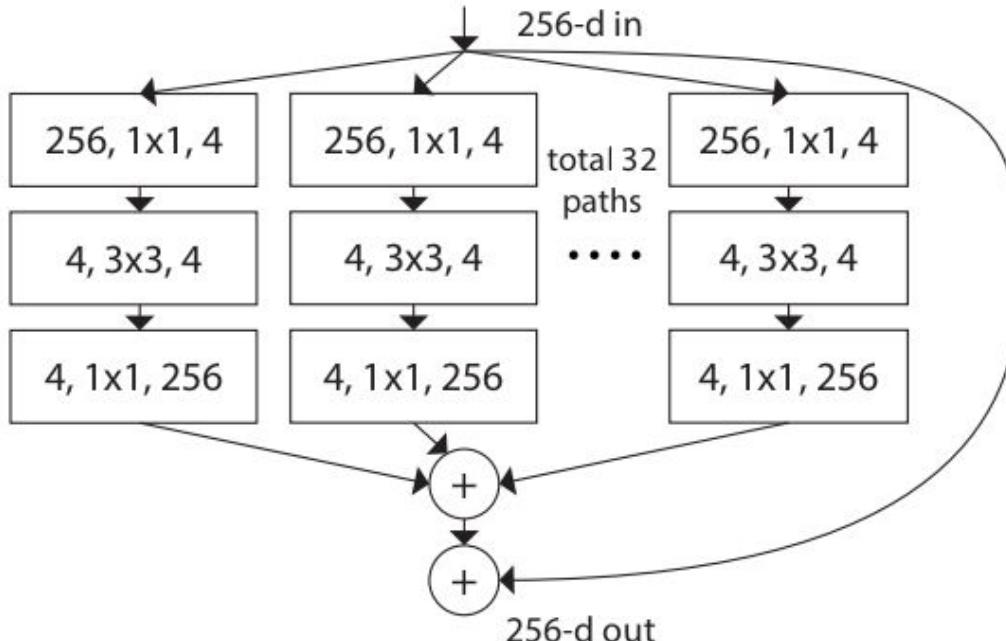
Seems to me like ImageNet is mostly dead

Ensembles of Models (Hikivision)

- More than 20 models, including VGG, Inception, ResNet and variations of it.
- Novel data augmentation.
- Novel learning rate policy.
- ...and “some small tricks”



ResNext = ResNet + Inception



facebook
research

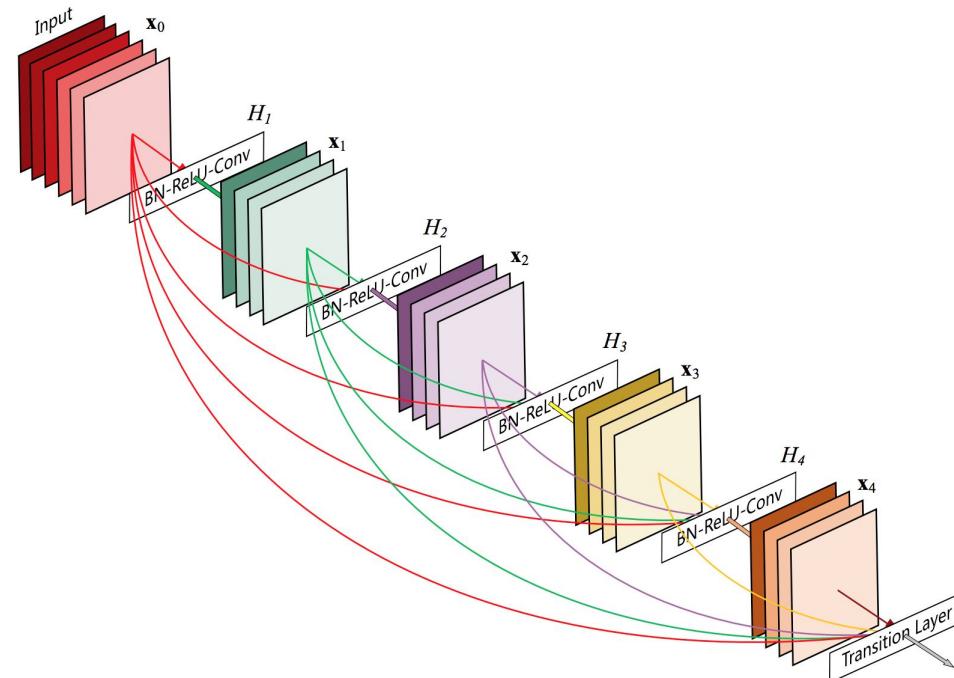
Xie, Saining, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. "[Aggregated residual transformations for deep neural networks.](#)" CVPR 2017 [\[code\]](#)



DenseNet

Connect every layer to every other layer of the same filter size.

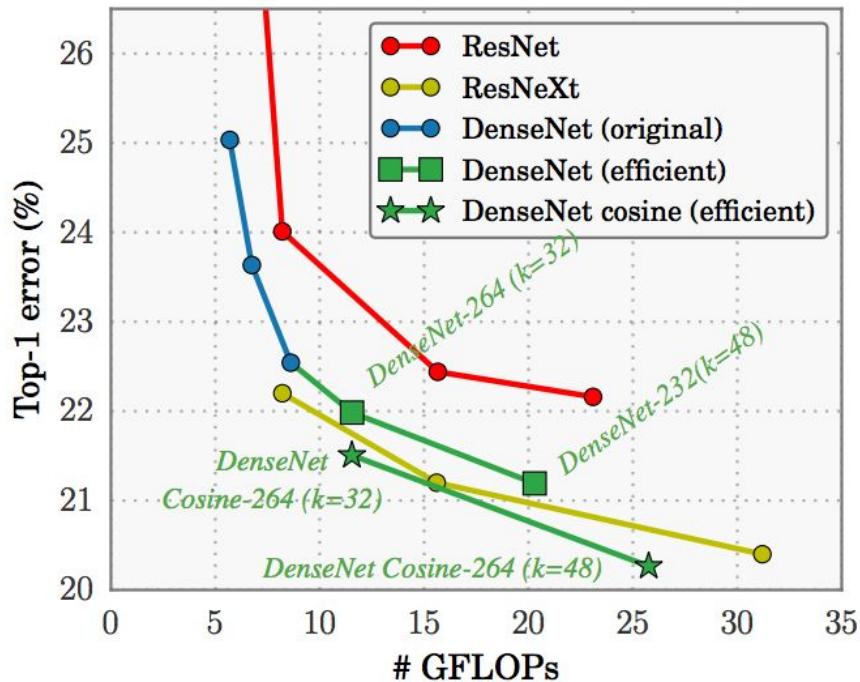
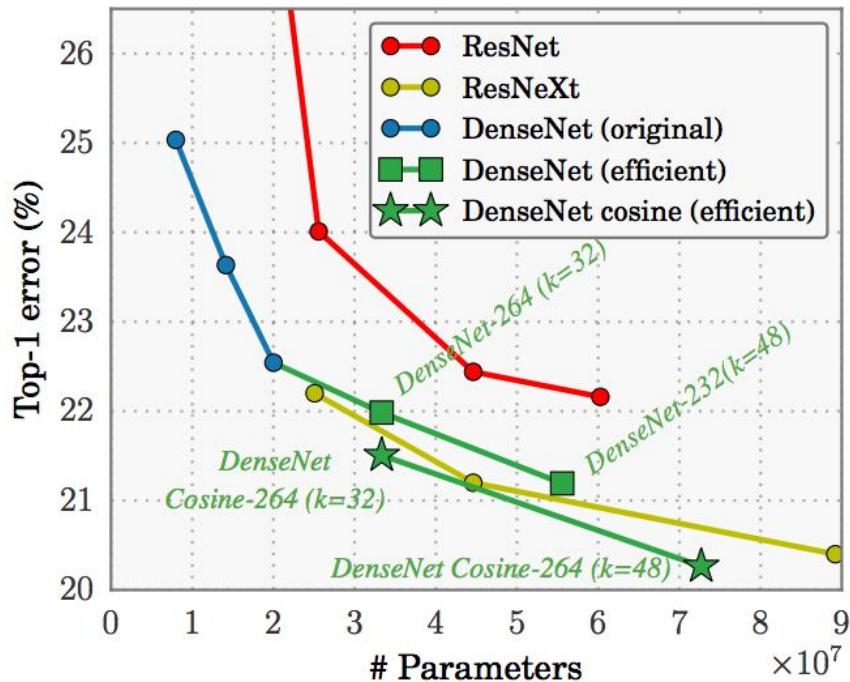
Dense Block of
5-layers with a
growth rate of $k=4$

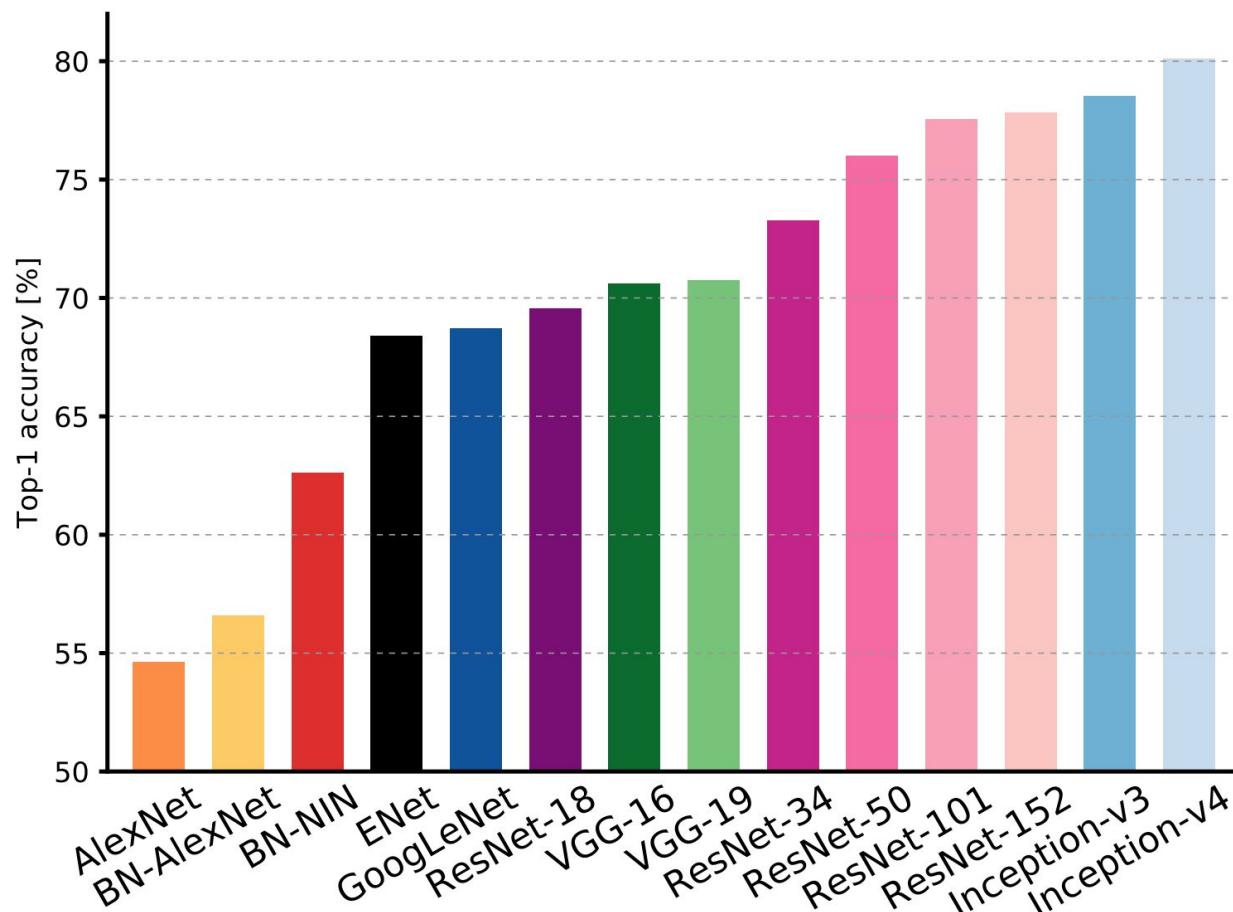




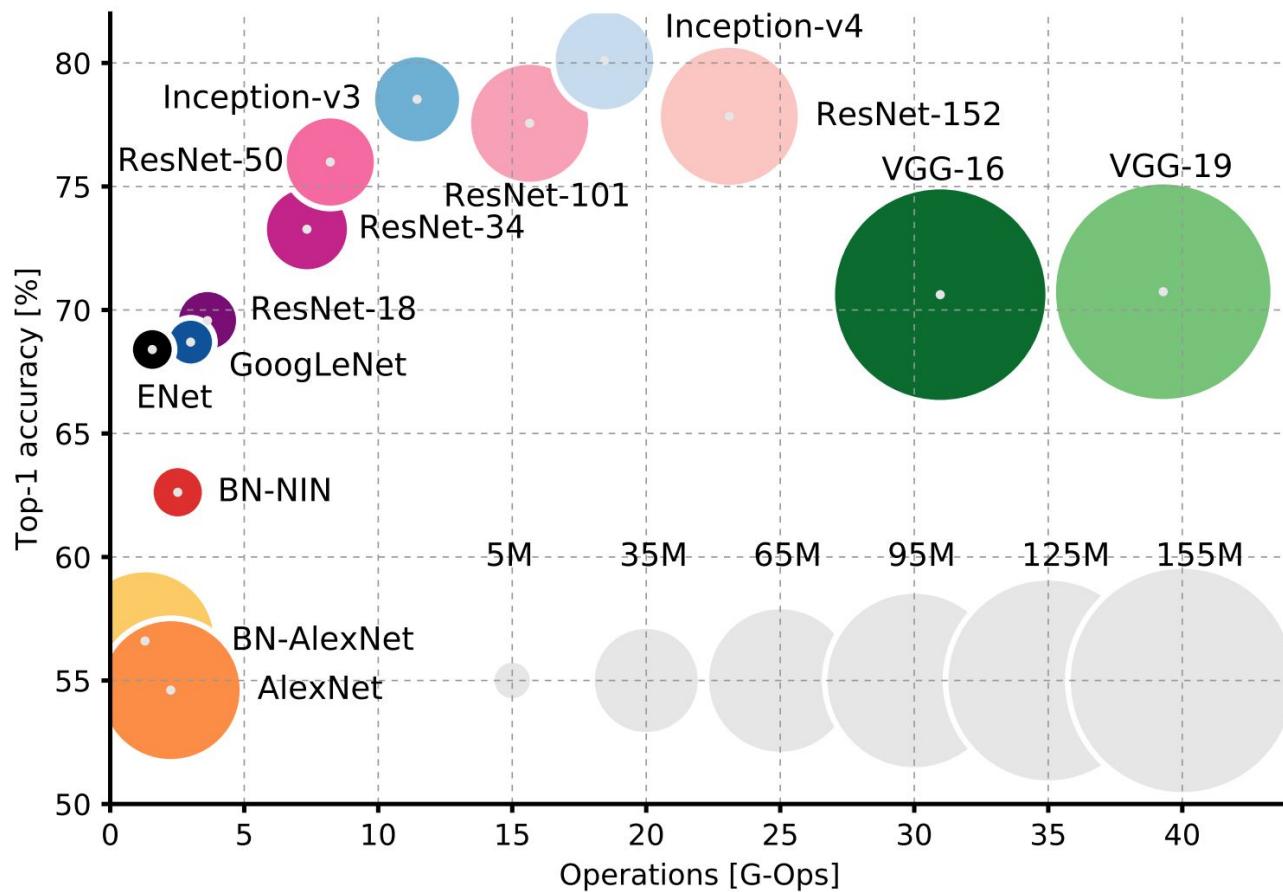
Beyond ImageNet Challenge

Results on ImageNet



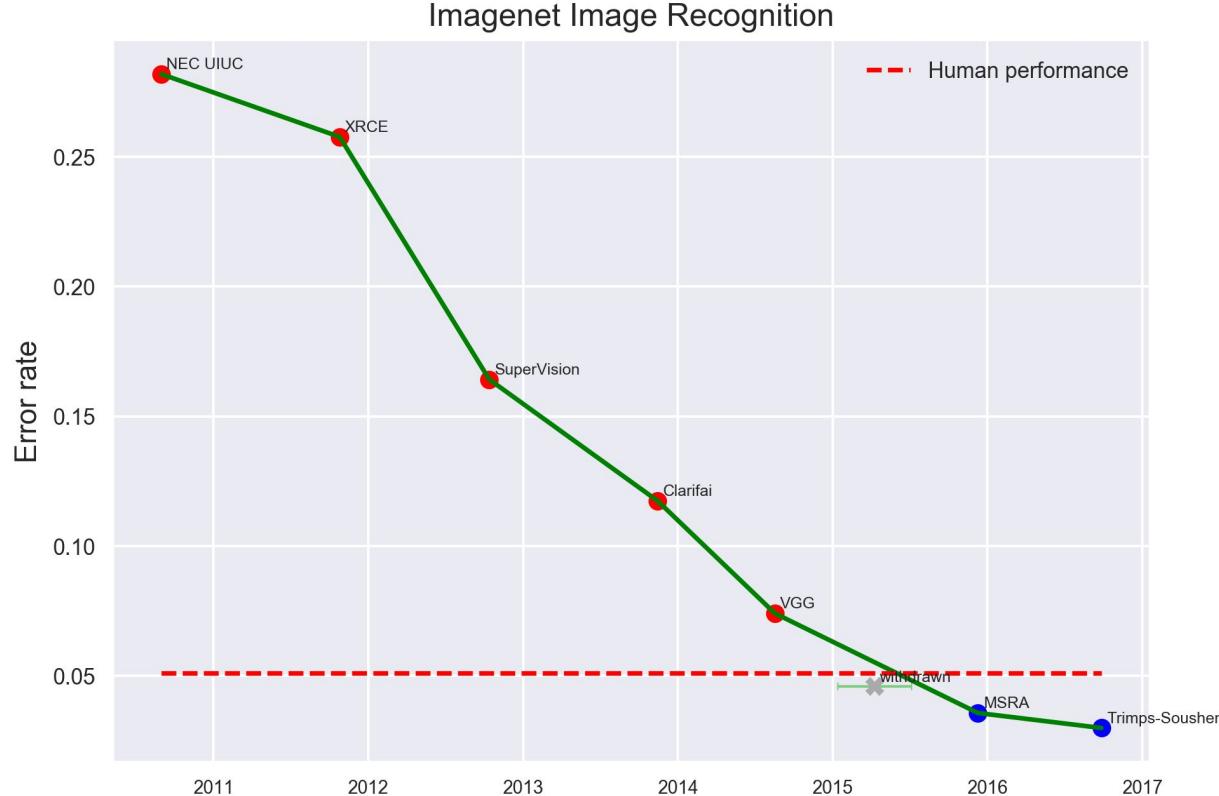


Canziani, Alfredo, Adam Paszke, and Eugenio Culurciello. "[An analysis of deep neural network models for practical applications.](#)" arXiv preprint arXiv:1605.07678 (2016).





The end of the challenge





The end of the challenge



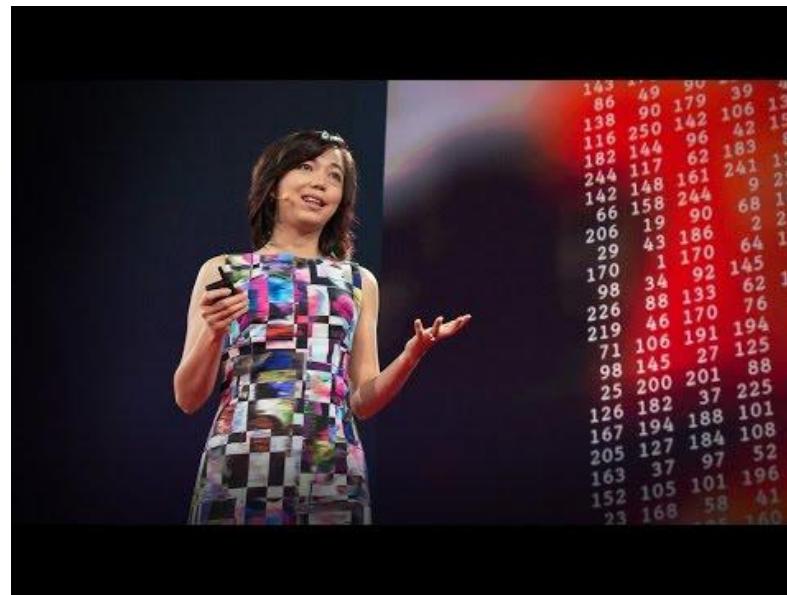
Beyond ImageNet Large Scale Visual Recognition Challenge

July 26th in conjunction with CVPR 2017

http://image-net.org/challenges/beyond_ilsvrc

Learn more

Li Fei-Fei, ["How we're teaching computers to understand pictures"](#) TEDTalks 2014.



Questions?