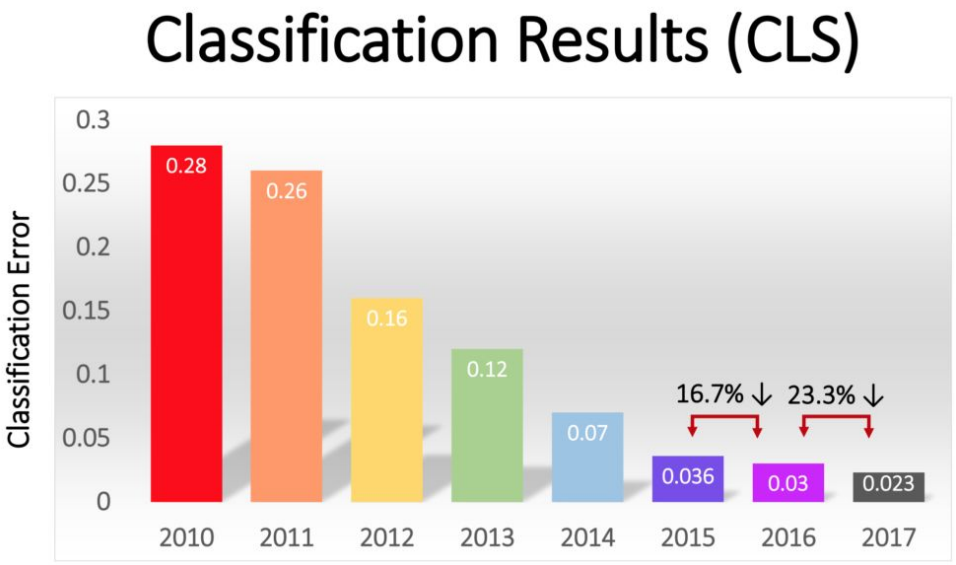
ImageNet Winning CNN Architectures (ILSVRC)



ImageNet Dataset

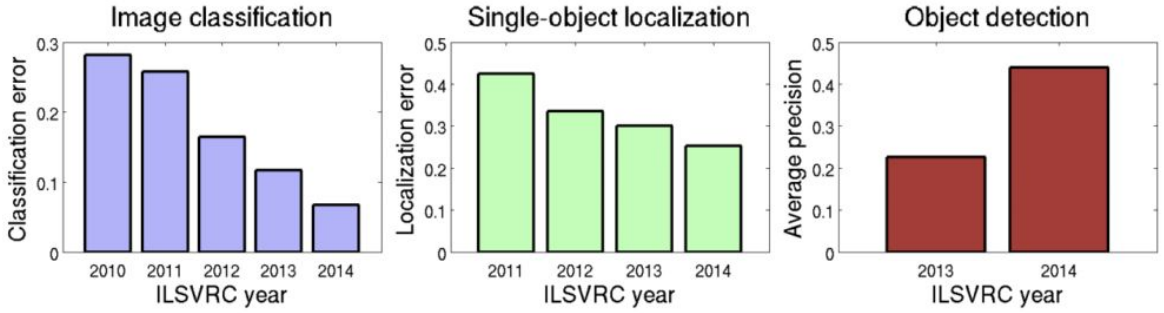
ImageNet is a dataset of over 15 million labeled high-resolution images belonging to roughly 22,000 categories. The images were collected from the web and labeled by human labelers using Amazon's Mechanical Turk crowd-sourcing tool. Starting in 2010, as part of the Pascal Visual Object Challenge, an annual competition called the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) has been held. ILSVRC uses a subset of ImageNet with roughly 1000 images in each of 1000 categories. In all, there are roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images.

On ImageNet, it is customary to report two error rates: top-1 and top-5, where the top-5 error rate is the fraction of test images for which the correct label is not among the five labels considered most probable by the model. ImageNet consists of variable-resolution images, while our system requires a constant input dimensionality.

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

The general challenge tasks for most years are as follows:

- Image classification: Predict the classes of objects present in an image.
- Single-object localization: Image classification + draw a bounding box around one example of each object present.
- Object detection: Image classification + draw a bounding box around each object present.



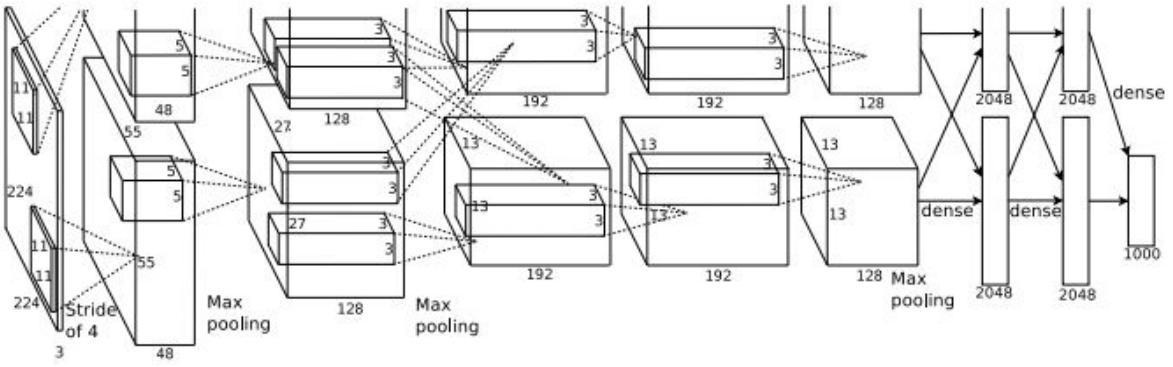
Summary of the Improvement on ILSVRC Tasks Over the First Five Years of the Competition. Taken from ImageNet Large Scale Visual Recognition Challenge, 2015

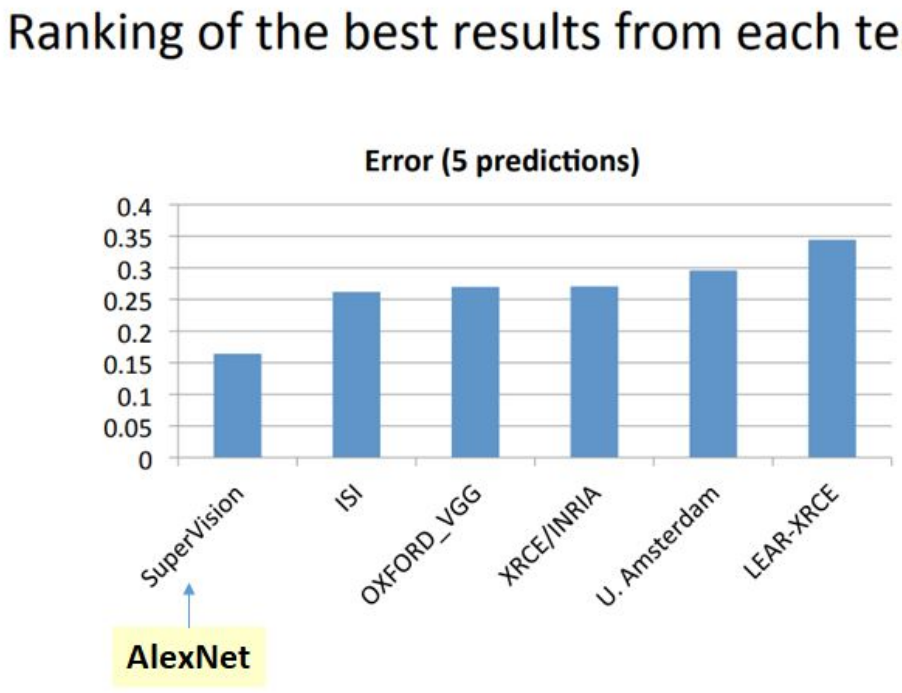
Deep Learning Milestones From ILSVRC

The pace of improvement in the first five years of the ILSVRC was dramatic, perhaps even shocking to the field of computer vision. Success has primarily been achieved by large (deep) convolutional neural networks (CNNs) on graphical processing unit (GPU) hardware, which sparked an interest in deep learning that extended beyond the field out into the mainstream.

ILSVRC-2012

AlexNet (SuperVision)



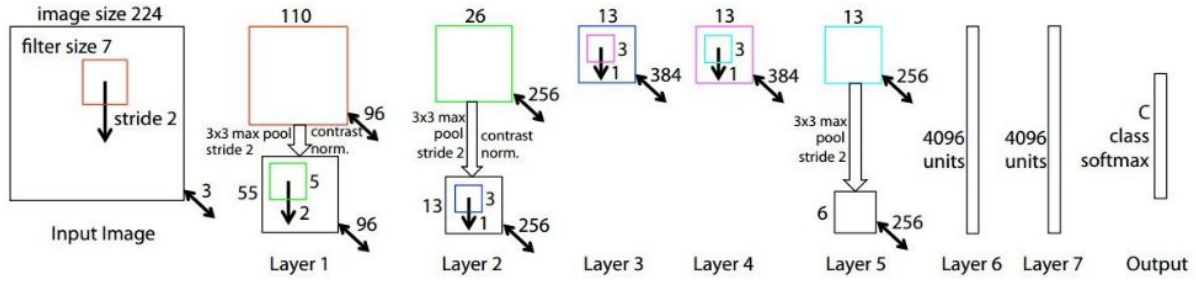


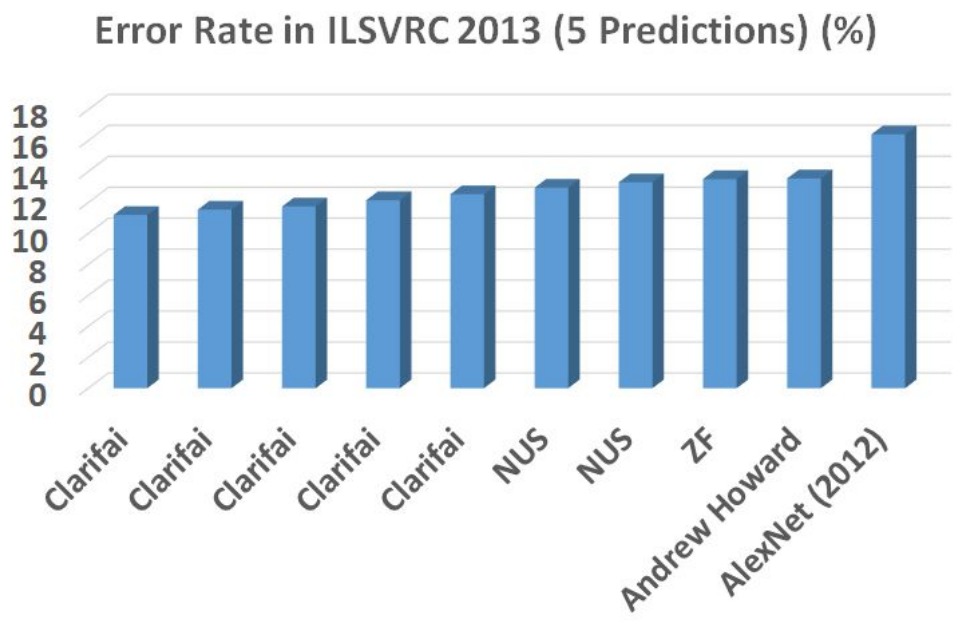
On 30 September 2012, a convolutional neural network (CNN) called AlexNet achieved a top-5 error of 15.3% in the ImageNet 2012 Challenge, more than 10.8 percentage points lower than that of the runner up. This was made feasible due to the use of Graphics processing units (GPUs) during training, an essential ingredient of the deep learning revolution. According to The Economist, "Suddenly people started to pay attention, not just within the AI community but across the technology industry as a whole.

- ImageNet Classification with Deep Convolutional Neural Networks, 2012. (Authors: Alex Krizhevsky, Ilya Sutskever, Geoffrey Hinton. University of Toronto, Canada.)
- With 60M parameters, AlexNet has 8 layers — 5 convolutional and 3 fully-connected.
- They were the first to implement Rectified Linear Units (ReLUs) as activation functions.

ILSVRC-2013

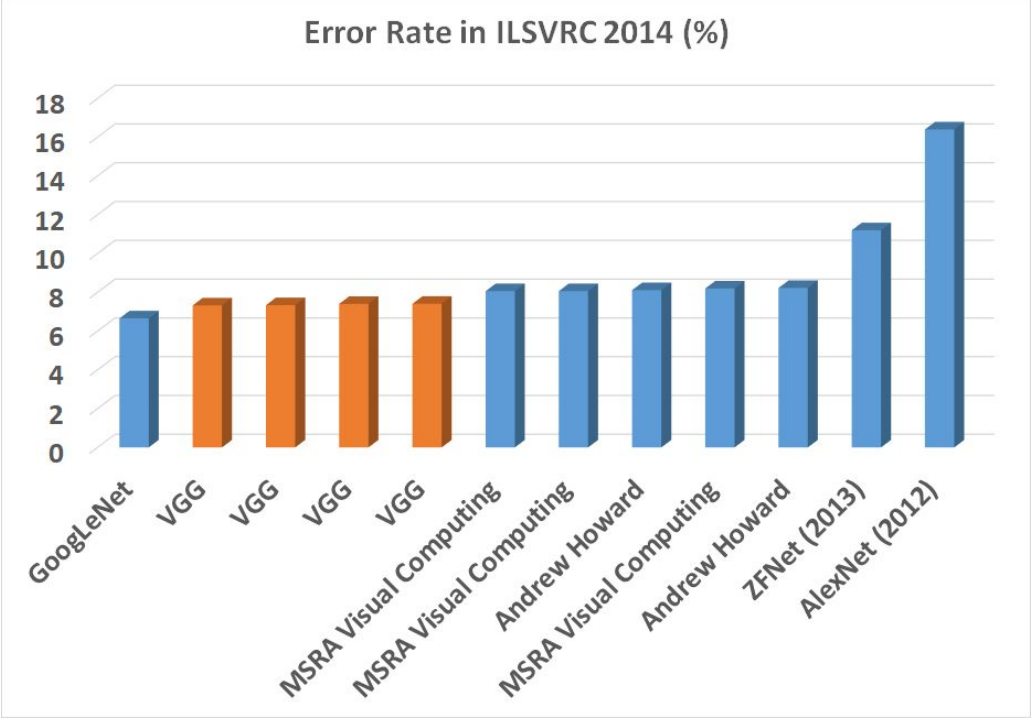
ZFNet (Clarifai)



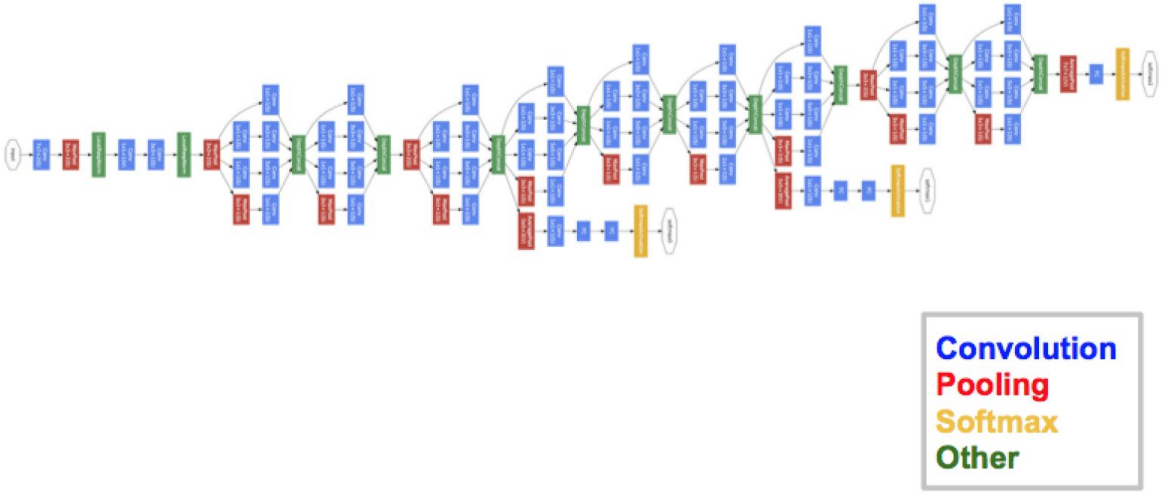


Matthew Zeiler and Rob Fergus propose a variation of AlexNet generally referred to as ZFNet in their 2013 paper titled "Visualizing and Understanding Convolutional Networks," a variation of which won the ILSVRC-2013 image classification task.

ILSVRC-2014



Inception (GoogLeNet)



Christian Szegedy, et al. from Google achieved top results for object detection with their GoogLeNet model that made use of the inception module and architecture. This approach was described in their 2014 paper titled “Going Deeper with Convolutions.”

Introduced the Inception Module, which emphasized that the layers of a CNN doesn't always have to be stacked up sequentially. The winner of ILSVRC 2014 with an error rate of 6.7%.

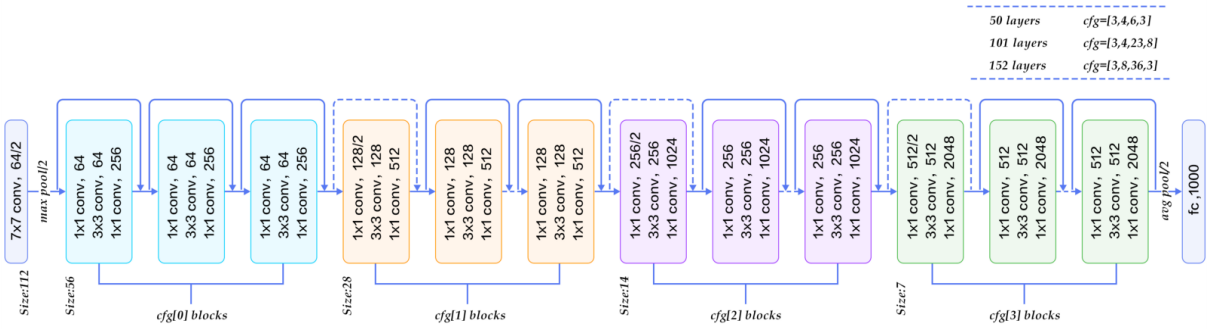
VGG

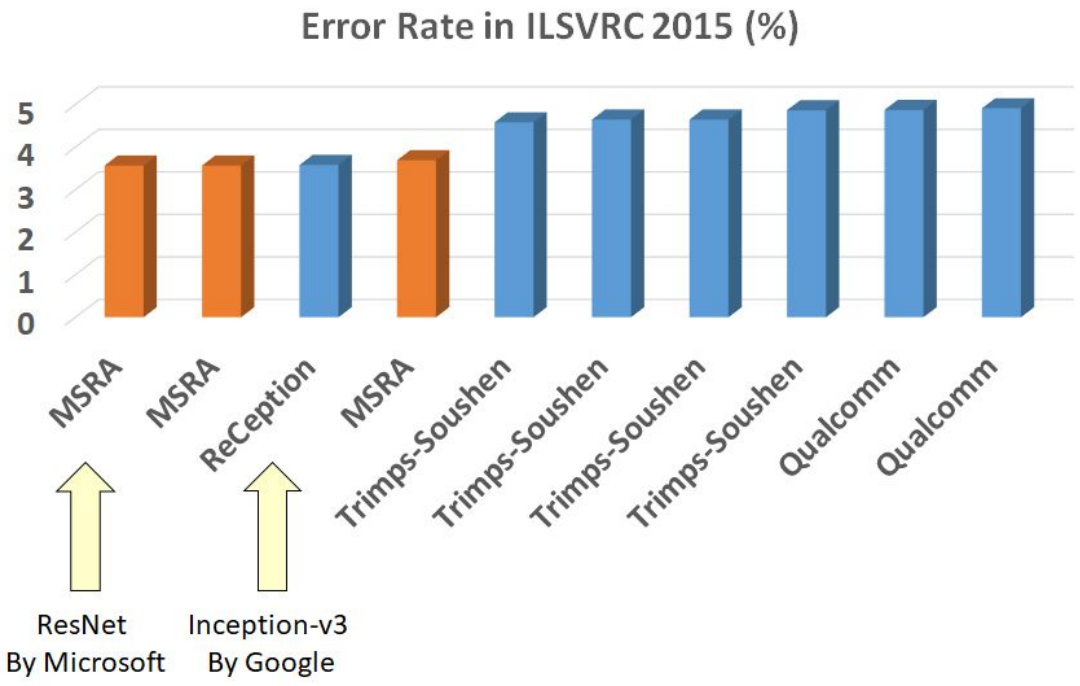
Karen Simonyan and Andrew Zisserman from the Oxford Vision Geometry Group (VGG) achieved top results for image classification and localization with their VGG model. Their approach is described in their 2015 paper titled “Very Deep Convolutional Networks for Large-Scale Image Recognition.”.

The folks at Visual Geometry Group (VGG) invented the VGG-16 which has 13 convolutional and 3 fully-connected layers, carrying with them the ReLU tradition from AlexNet. This network stacks more layers onto AlexNet, and use smaller size filters (2×2 and 3×3). It consists of 138M parameters and takes up about 500MB of storage space They also designed a deeper variant, VGG-19.

ILSVRC-2015

ResNet (MSRA)



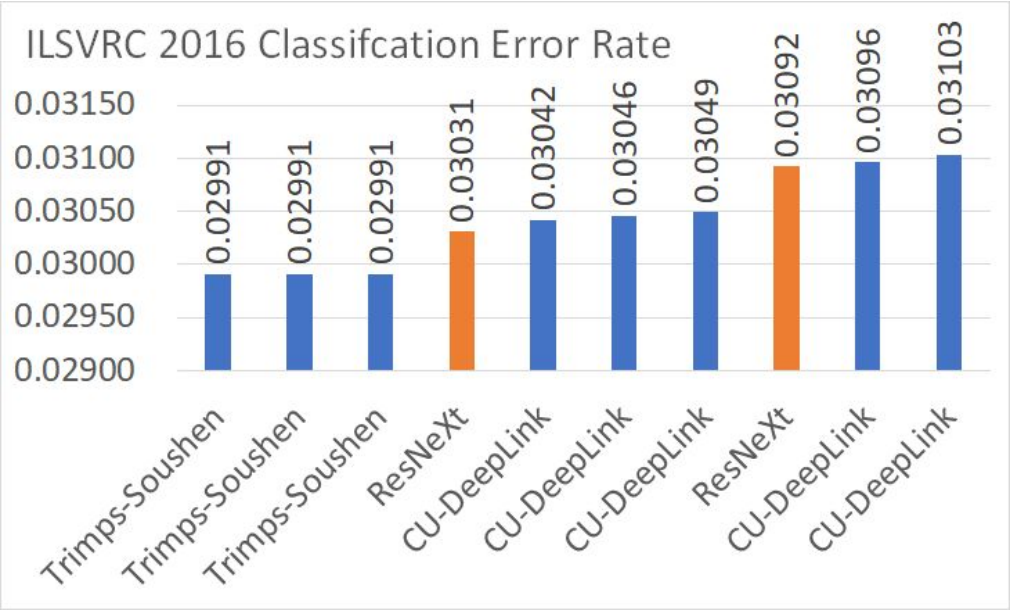


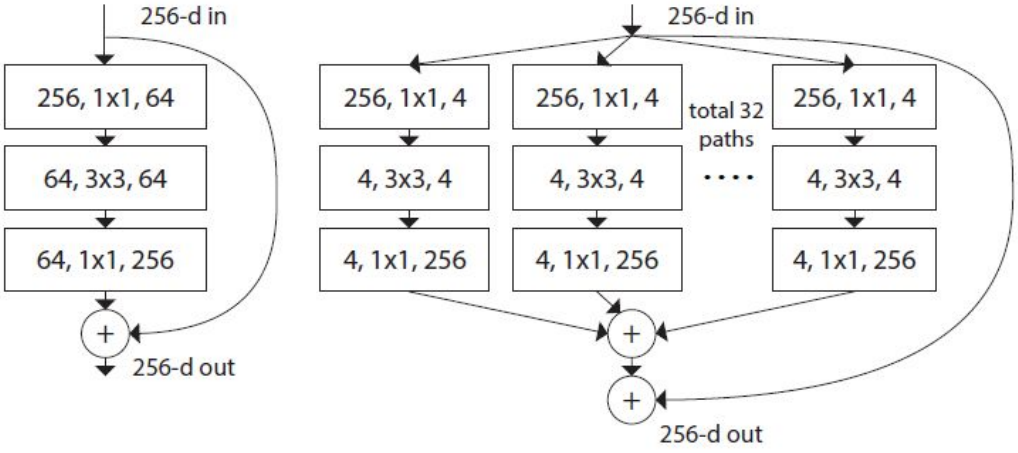
Kaiming He, et al. from Microsoft Research achieved top results for object detection and object detection with localization tasks with their Residual Network or ResNet described in their 2015 paper titled “Deep Residual Learning for Image Recognition.”

An ensemble of these residual nets achieves 3.57% error on the ImageNet test set. Ultra-deep (quoting the authors) architecture with 152 layers. Introduced the Residual Block, to reduce overfitting

ILSVRC-2016

ResNeXt

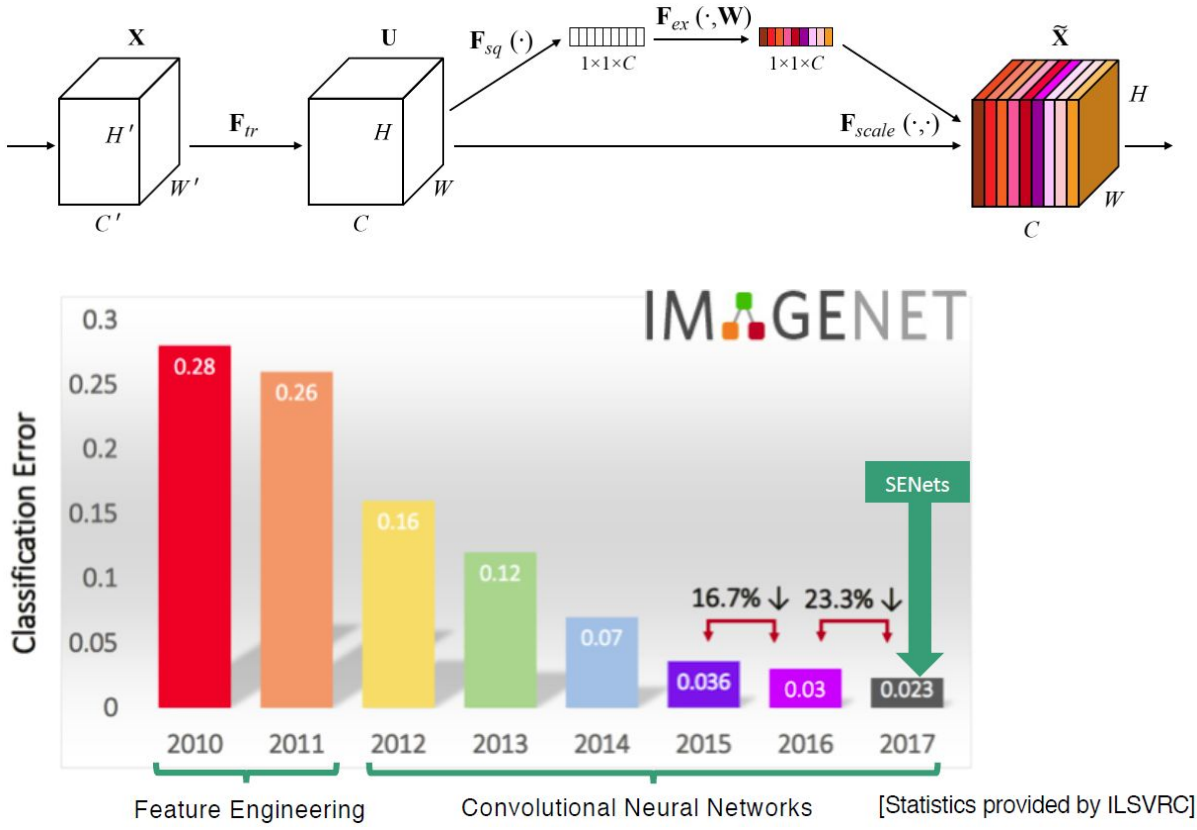




The model name, ResNeXt, contains Next. It means the next dimension, on top of the ResNet. This next dimension is called the “cardinality” dimension. And ResNeXt becomes the 1st Runner Up of ILSVRC classification task.

ILSVRC-2017

SENet



With “Squeeze-and-Excitation” (SE) block that adaptively recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels, SENet is constructed. And it won the first place in ILSVRC 2017 classification challenge with top-5 error to 2.251% which has about 25% relative improvement over the winning entry of 2016. And this is a paper in 2018 CVPR with more than 600 citations. Recently, it is also published in 2019 TPAMI.

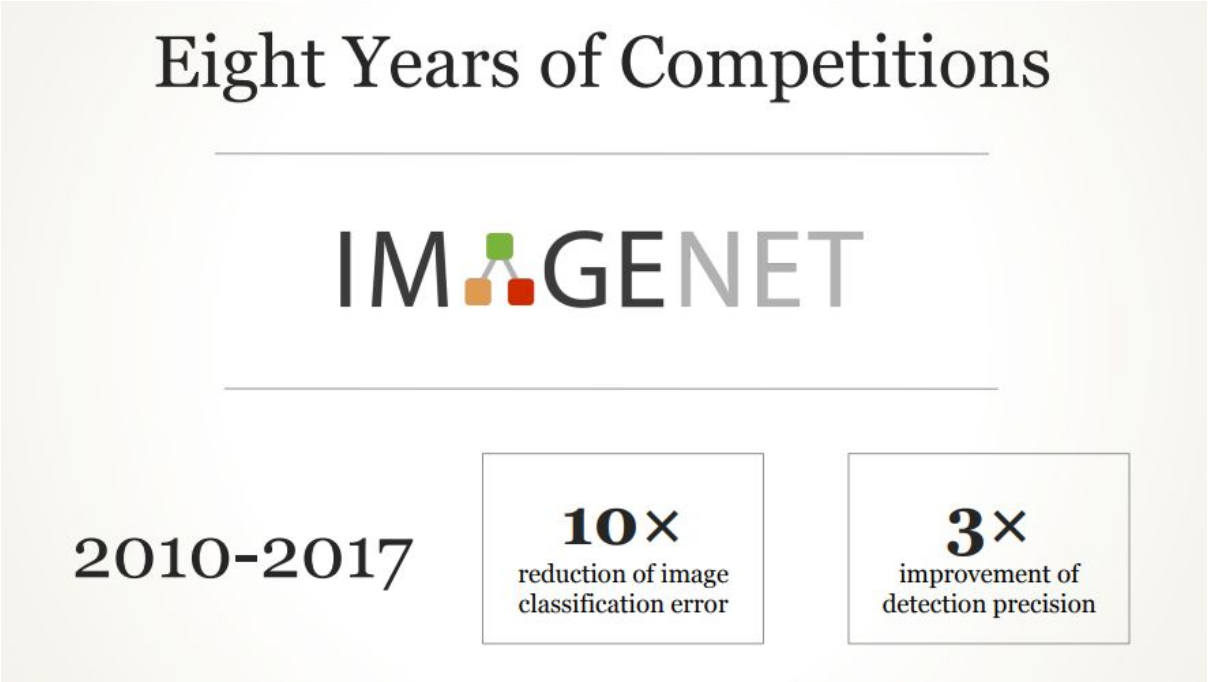
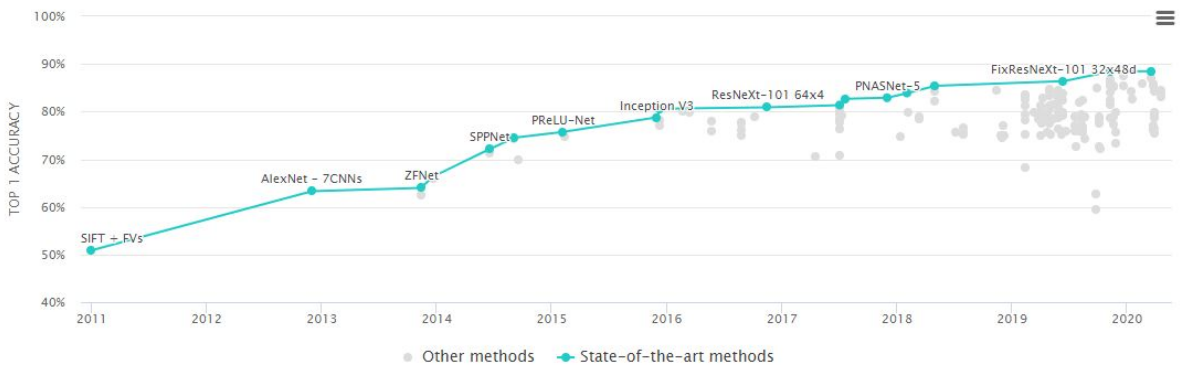


Image Classification on ImageNet

This link contains the codes and all details of all architectures https://www.paperswithcode.com/sota/image-classification-on-imagenet



Thank you so much if you have read so far. I have always wondered how ImageNet is progressing. I hope it benefits everyone who reads.

References:

- http://www.cs.toronto.edu/~hinton/absps/imagenet.pdf
- https://machinelearningmastery.com/introduction-to-the-imagenet-large-scale-visual-recognition-challenge-ilsvrc/
- https://arxiv.org/pdf/1409.0575.pdf
- http://image-net.org/challenges/LSVRC/2017/index
- https://en.wikipedia.org/wiki/ImageNet
- https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d
- https://www.codesofinterest.com/2017/07/milestones-of-deep-learning.html
- http://image-net.org/challenges/LSVRC/2016/results
- https://towardsdatascience.com/review-trimps-soushen-winner-in-ilsvrc-2016-image-classification-dfbc423111dd
- https://towardsdatascience.com/review-resnext-1st-runner-up-of-ilsvrc-2016-image-classification-15d7f17b42ac
- http://image-net.org/challenges/LSVRC/2017/results
- https://towardsdatascience.com/review-senet-squeeze-and-excitation-network-winner-of-ilsvrc-2017-image-classification-a887b98b2883
- https://medium.com/coinmonks/paper-review-of-alexnet-caffenet-winner-in-ilsvrc-2012-image-classification-b93598314160
- https://mc.ai/paper-review-of-zfnet-the-winner-of-ilsvlc-2013-image-classification/
- http://image-net.org/challenges/talks_2017/imagenet_ilsvrc2017_v1.0.pdf
- https://iq.opengenus.org/evolution-of-cnn-architectures/