

Lecture Notes for **Machine Learning in Python**



Professor Eric Larson
An Ongoing History of Convolutional Networks

Class logistics and Agenda

- Wide/Deep Lab due soon!
- Agenda:
 - CNN Demo
 - History of CNNs
 - with Modern CNN Architectures
- Next Time:
 - More Advanced CNN Demo

Class Overview, by topic

Table Data
Visualization

Numpy, Pandas, Seaborn
Overviews with some in-depth discussion

Dimension
Reduction and
Image Processing

Scikit-learn, Scikit Image,
Intuition only, Some mathematics

Linear and
Logistic
Regression

Numpy, Recreate API for Scikit-learn
Detailed mathematics for simple optimization
intuition for advanced optimization

Neural Networks
and Back Prop.

Numpy
Detailed mathematics for NN operations

Wide and Deep
Networks

Convolutional
Networks

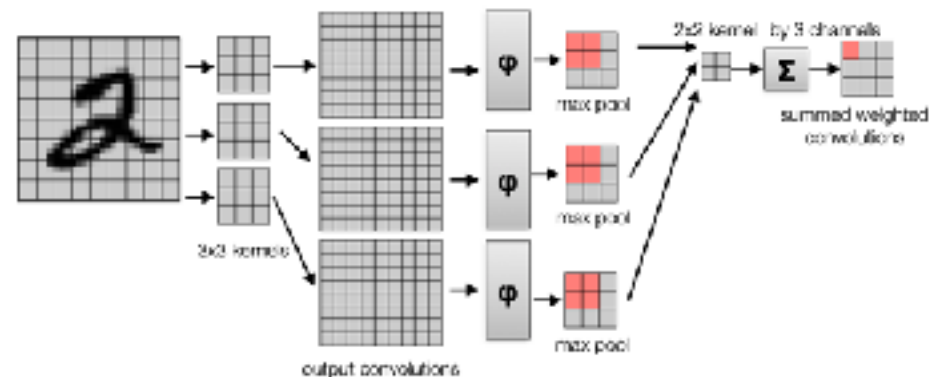
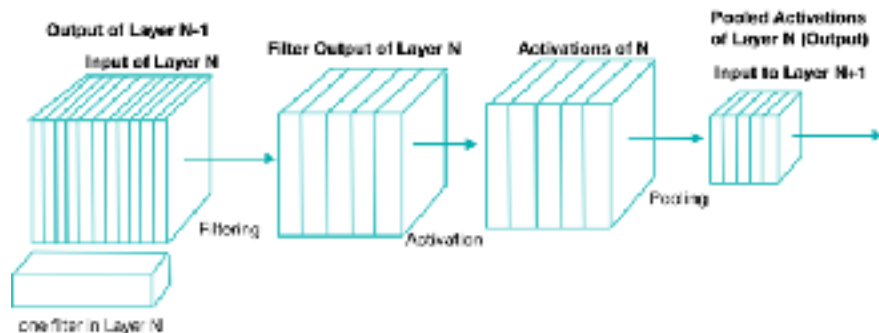
Recurrent
Networks

Keras, Tensorflow
Intuition, Detailed implement.

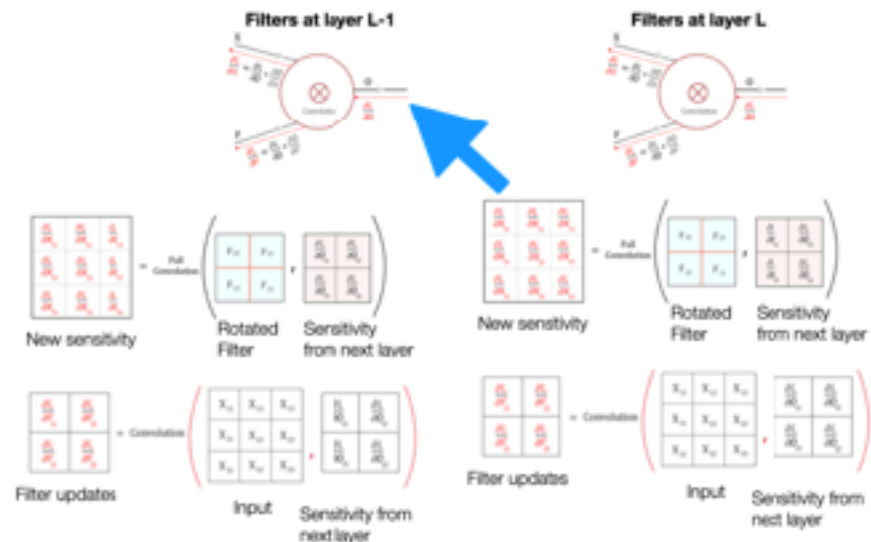
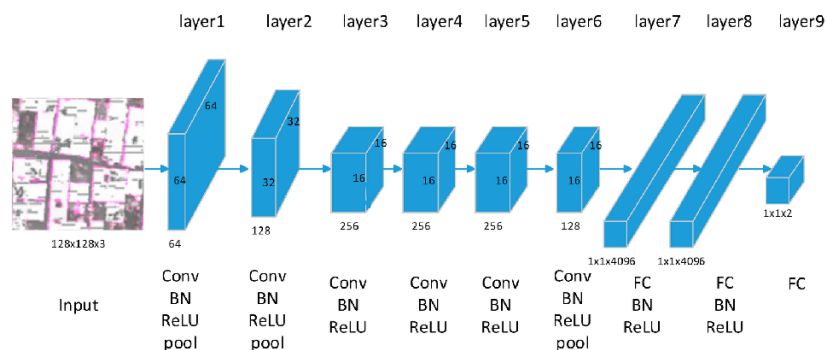
Ethics in
Language Models

ConceptNet
Case studies

Last Time:



Structure of Each Tensor: Channels x Rows x Columns



TensorFlow and Basic CNNs

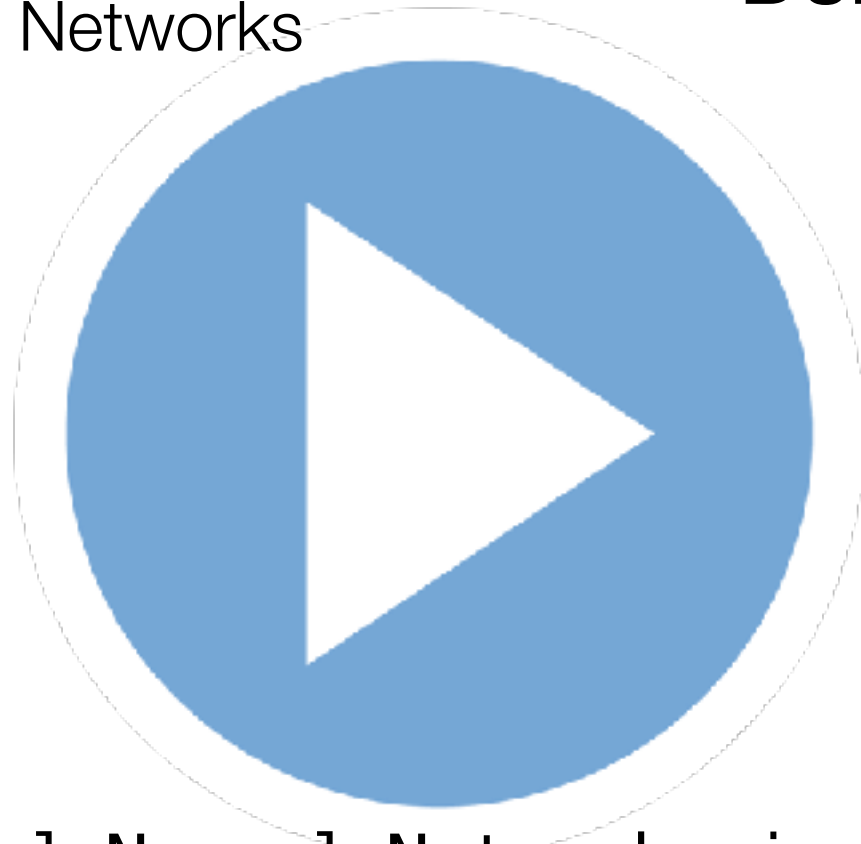
**Last
Time!**

Convolutional Neural Networks
in TensorFlow
with Keras

with Sequential API!

If needed:

**Finish
Demo**



11. Convolutional Neural Networks.ipynb

History of Convolutional Neural Networks



Thanks to machine-learning algorithms,
the robot apocalypse was short-lived.

Types of CNN, 1988-1998

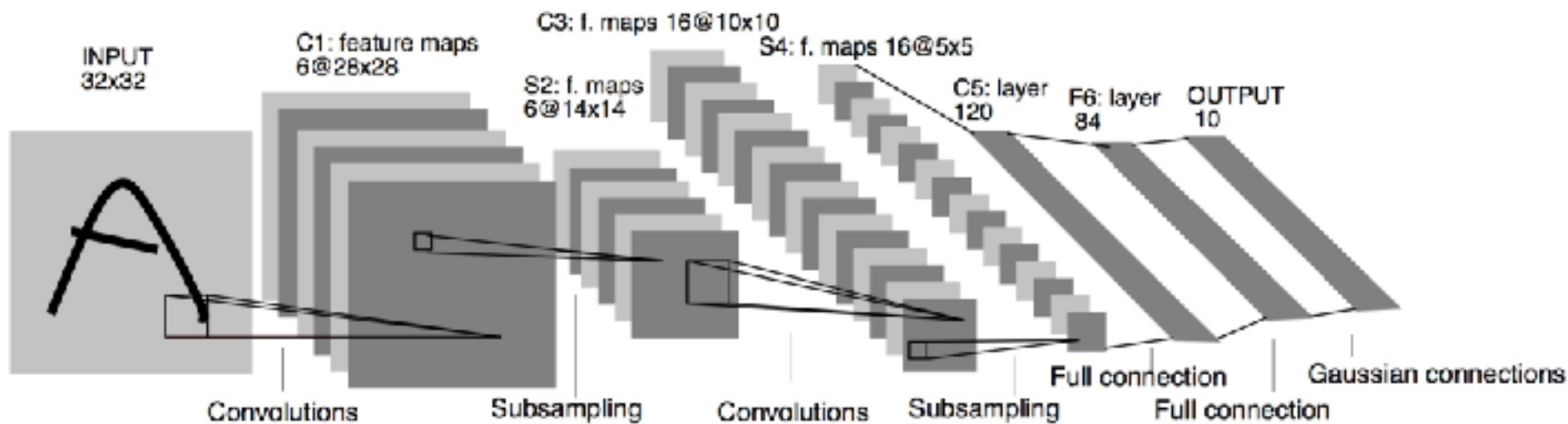


Yann LeCun
Heads Facebook AI Team

- **LeNet-1** (1988)
 - ~2600 params, not many layers
- **LeNet-5** (1998)
 - 7 layers, gets excellent MNIST performance
- Major contribution, general structure:
 - conv=>pool=>non-linearity=> ...=>MLP

avg

tanh or sigmoid



CNN History

- List of major breakthroughs from 1998 through 2010 in convolutional networks:



- 2010



Types of CNN, 2010



Dan Ciresan

AI Researcher
IDSA, Switzerland

- **Ciresan Net**
- Publishes code for running CNN via GPU
 - Subsequently wins 5 international competitions
 - from stop signs => cancer detection
- Major contribution: NVIDIA parallelized training algorithms

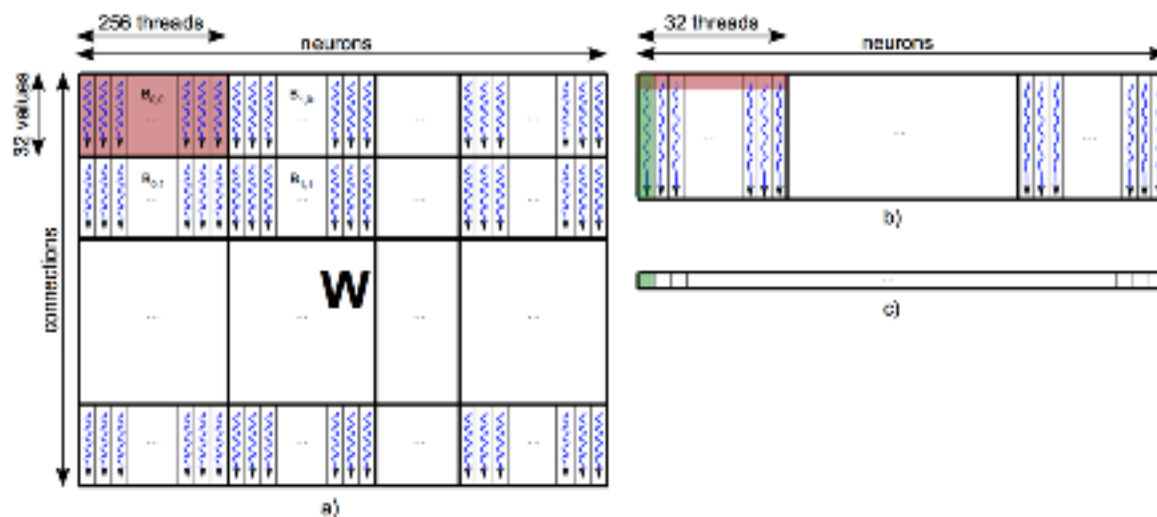
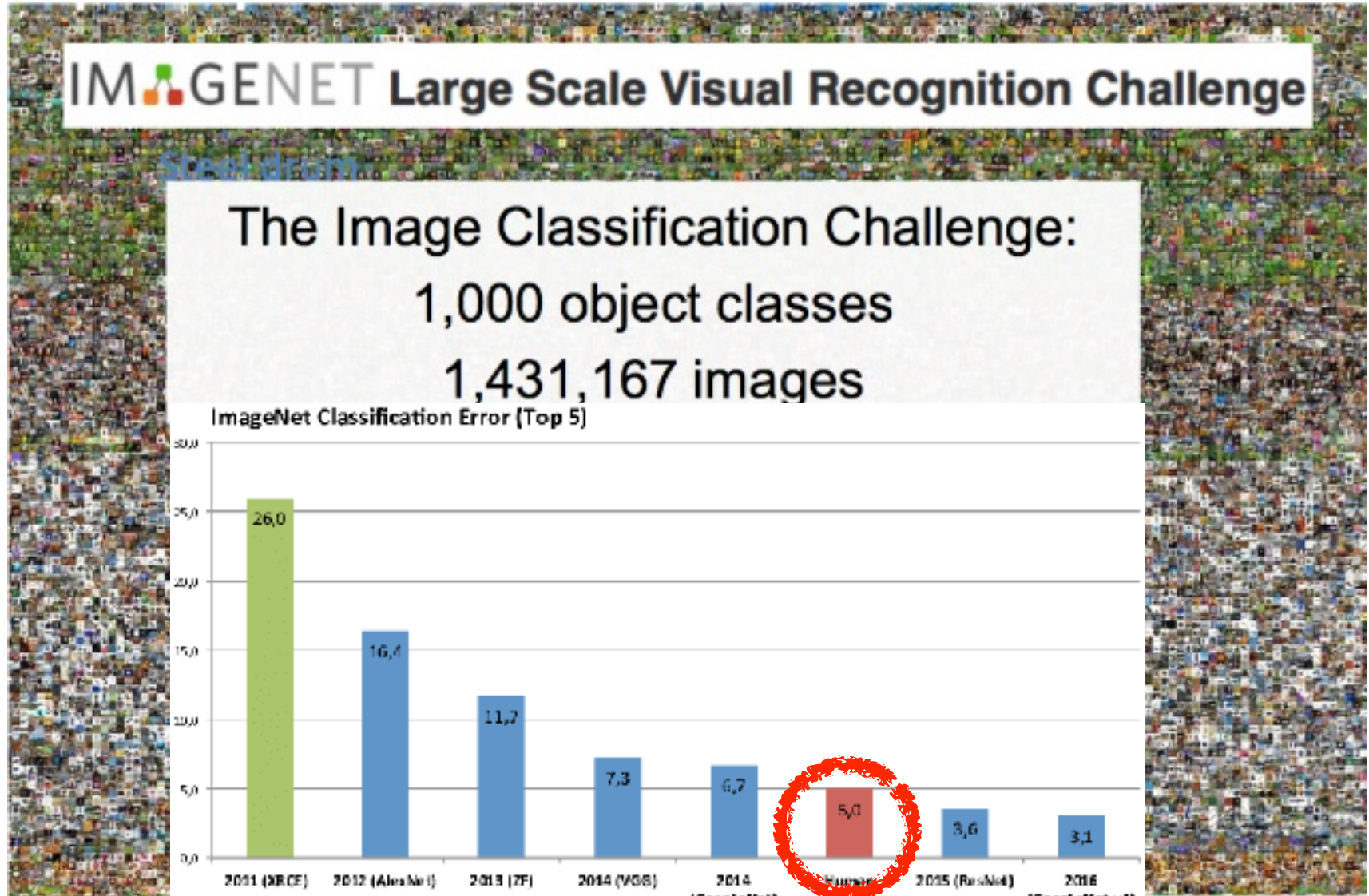


Figure 2: Forward propagation: a) mapping of kernel 1 grid onto the padded weight matrix; b) mapping the kernel 2 grid onto the partial dot products matrix; c) output of forward propagation.

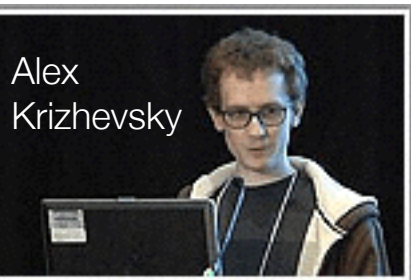
ImageNet Competition (2010-2016)



https://www.researchgate.net/figure/Winner-results-of-the-ImageNet-large-scale-visual-recognition-challenge-LSVRC-of-the_fig7_324476862

<https://www.slideshare.net/nmhkahn/case-study-of-convolutional-neural-network-61556303>

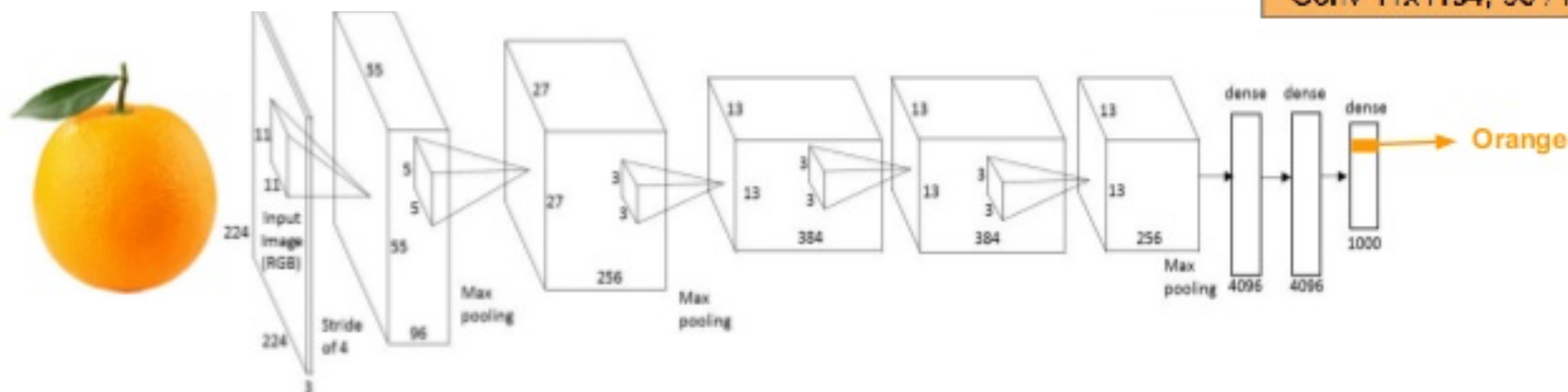
Types of CNN, 2012



Alex Krizhevsky

Google

- **AlexNet**, Hinton is mentor
 - wins ImageNet competition
- Major contributions:
 - dropout for regularization
 - systematic use of ReLU
 - data expansion
 - ***overlapping max pool***



AlexNet

FC 1000

FC 4096 / ReLU

FC 4096 / ReLU

Max Pool 3x3s2

Conv 3x3s1, 256 / ReLU

Conv 3x3s1, 384 / ReLU

Conv 3x3s1, 384 / ReLU

Max Pool 3x3s2

Local Response Norm

Conv 5x5s1, 256 / ReLU

Max Pool 3x3s2

Local Response Norm

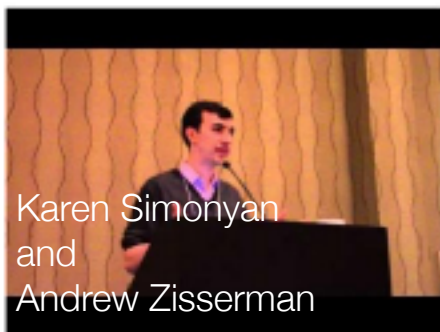
Conv 11x11s4, 96 / ReLU

Warning



WeKnowMemes

Types of CNN, 2013



Karen Simonyan
and
Andrew Zisserman

- Oxford **VGG Net** (Visual Geometry Group)
- Major contributions:
 - small cascaded kernels
 - way more layers (19 versus ~7)
 - “emulates” biology “better”
 - trained on NVIDIA GPUs for 2-3 weeks

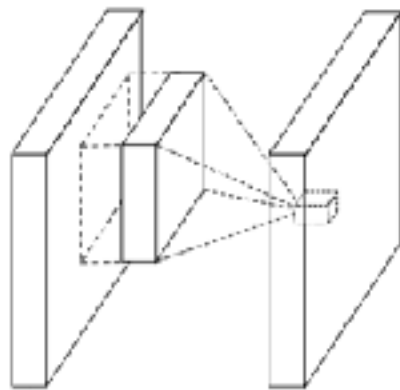
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Table 2: Number of parameters (in millions).

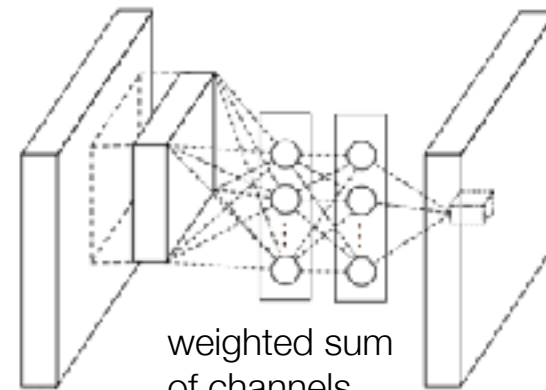
Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

- Network in Network **NiN**
 - or MLPConv

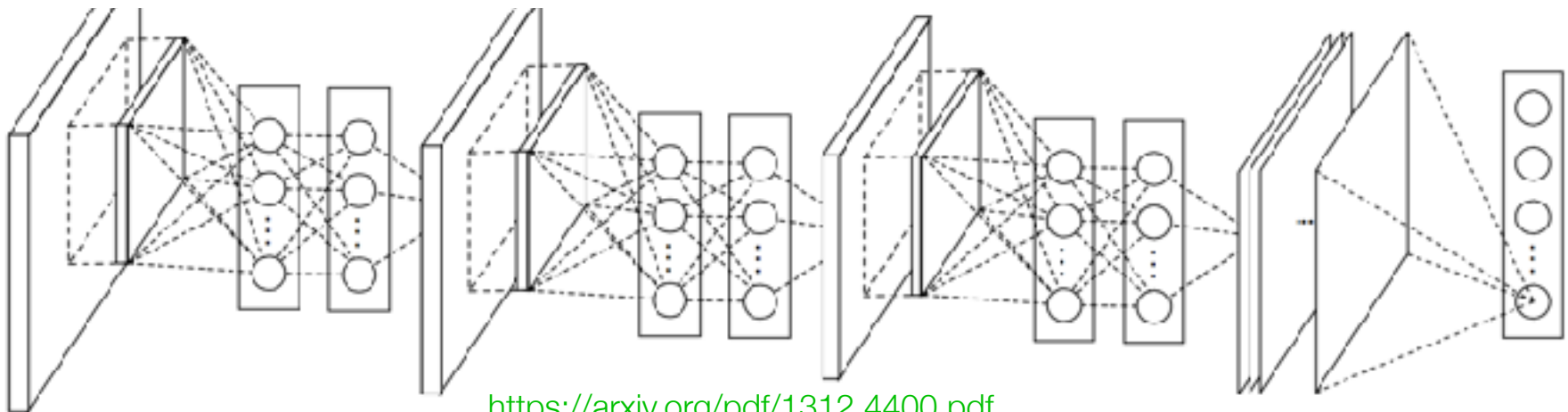
Min Lin^{1,2}, Qiang Chen², Shuicheng Yan²
¹Graduate School for Integrative Sciences and Engineering
²Department of Electronic & Computer Engineering
National University of Singapore, Singapore
{linmin, chenqiang, eleyans}@nus.edu.sg



(a) Linear convolution layer



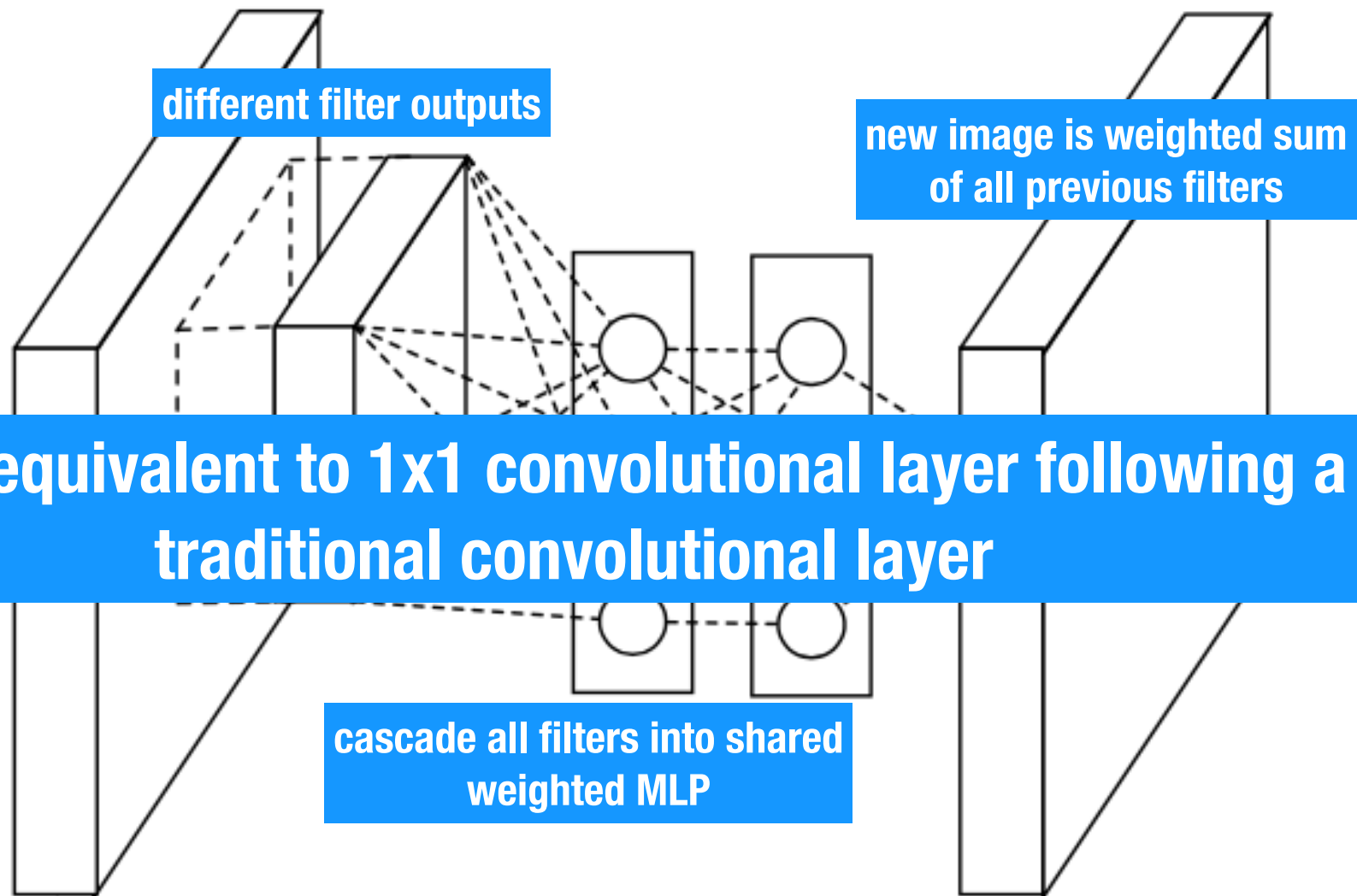
(b) Mlpconv layer



<https://arxiv.org/pdf/1312.4400.pdf>

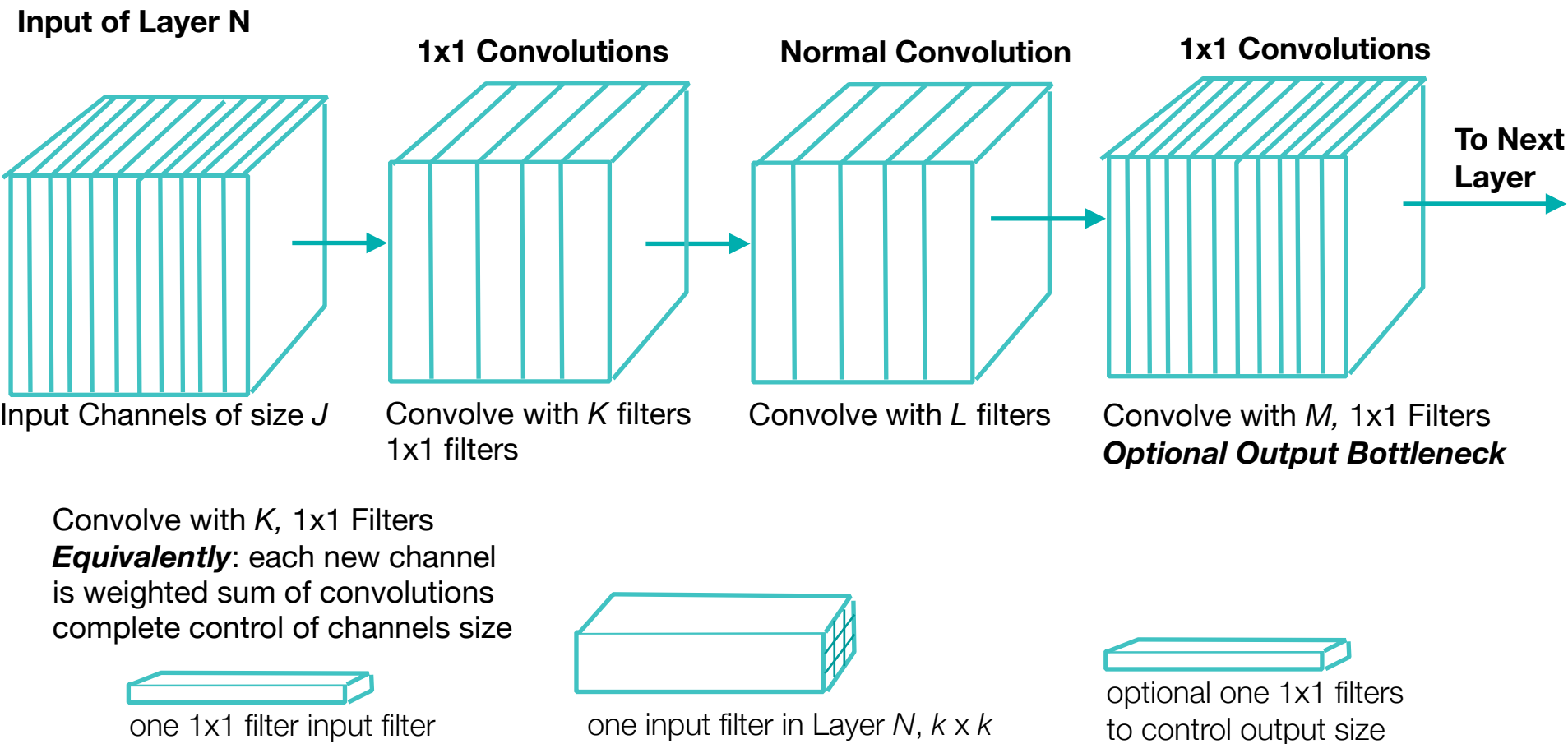
Types of CNN, 2014

- Network in Network



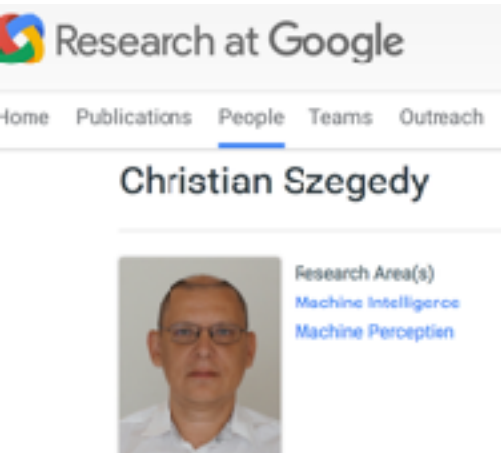
NiN, expanded view

J and $M \gg K$ and L

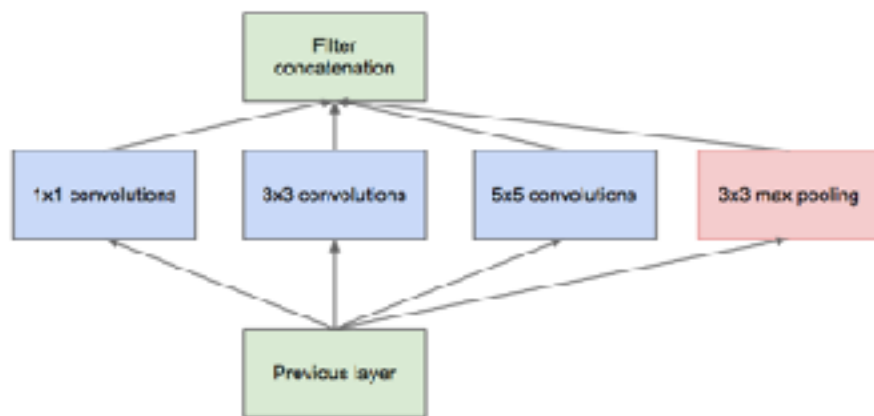


Structure of Each Tensor: Channels x Rows x Columns

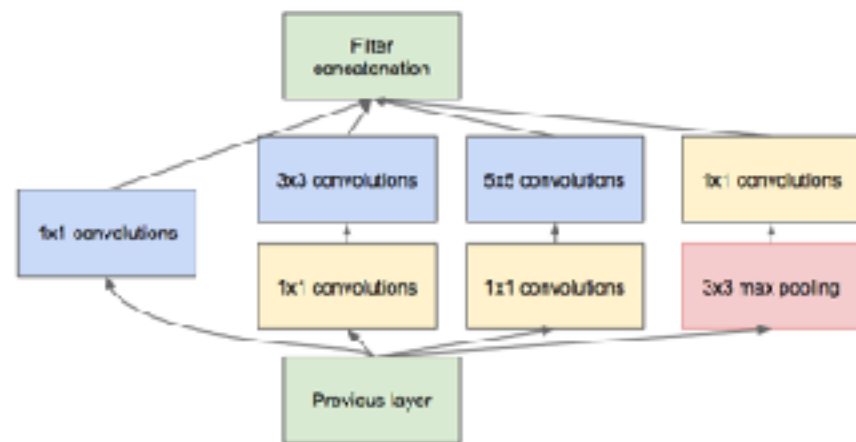
Types of CNN, 2014



- **GoogLeNet**
 - or **Inception V1**
- Major contribution:
 - bottleneck layering
 - parallel NiN



(a) Inception module, naïve version



(b) Inception module with dimension reductions

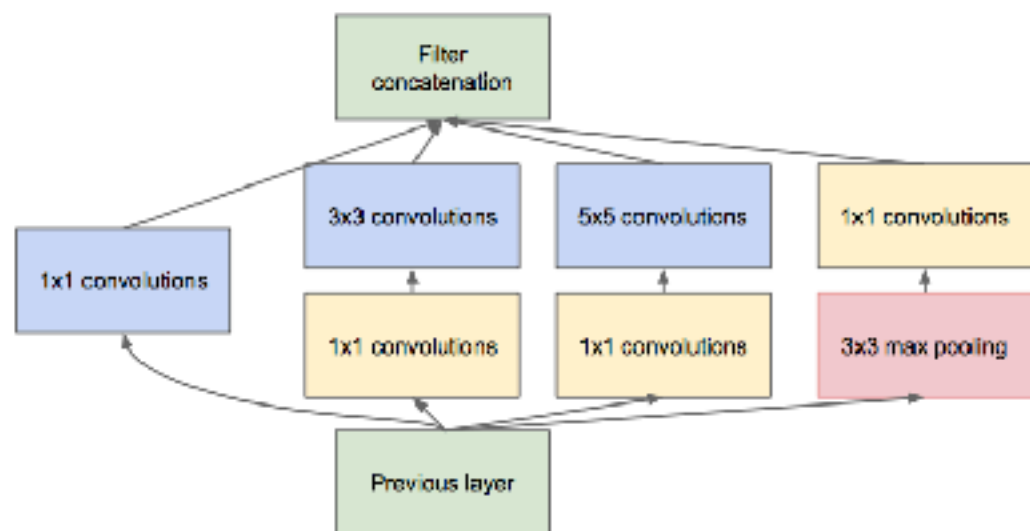
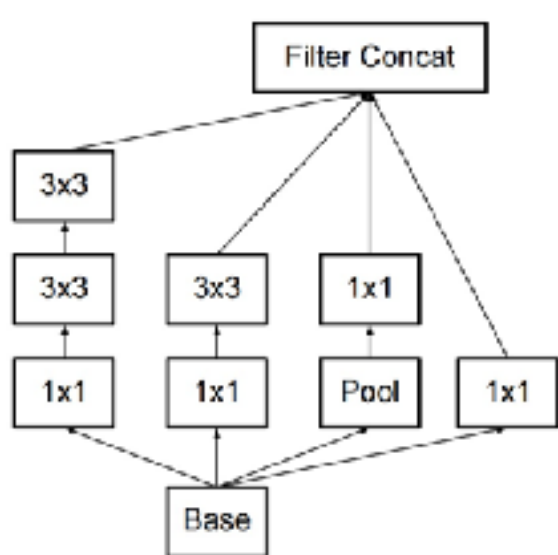


Figure 2: Inception module

Types of CNN, 2015 February and December



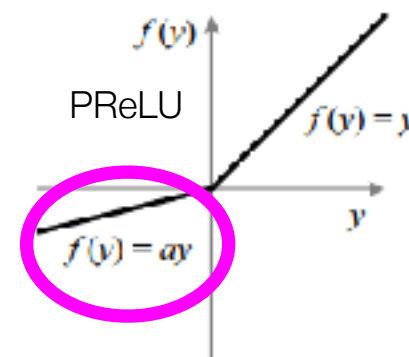
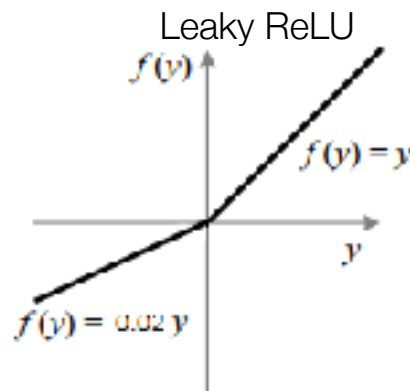
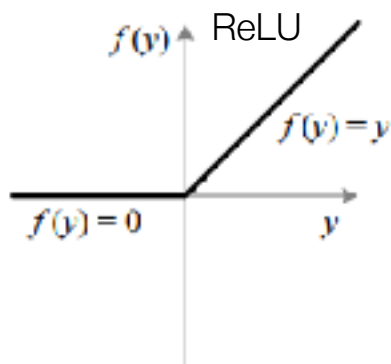
- **Inception V2**, Inception V1 with batch normalization
- **Inception V3:**
 - replace 5x5 with multiple 3x3



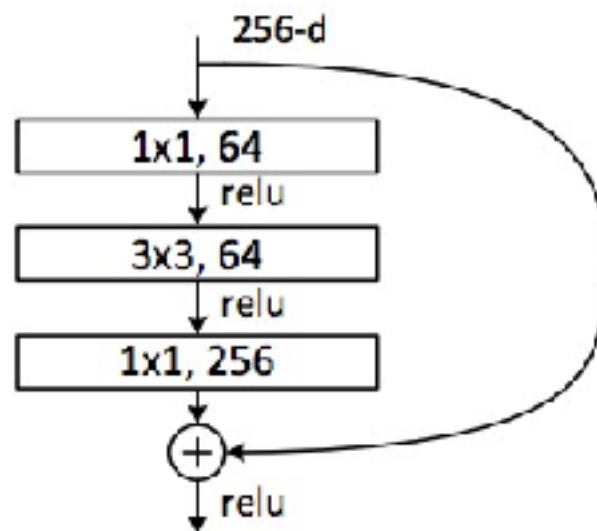
- Major Contributions:
 - “ensembles” not strictly sequential
 - “bio-plausible” with feedback

- ResNet**

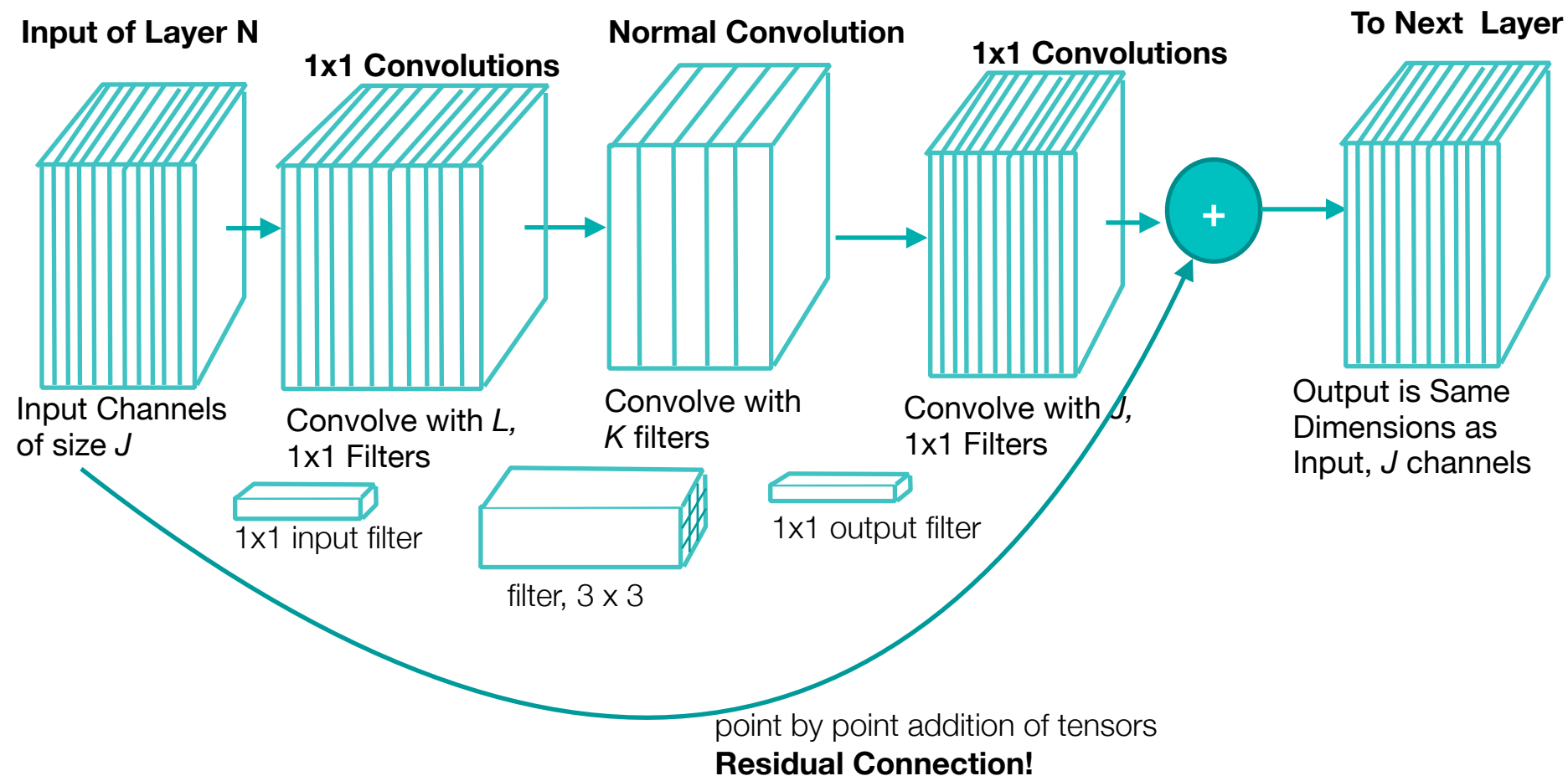
- Parametric ReLU
- PReLU: adaptive trained slope



- NiN: triple bypass layer
 - similar to bottleneck



Residual Connection, expanded view



Back Propagation: Two paths, including one without ANY operations that cause the gradient to vanish...

Types of CNN, 2017

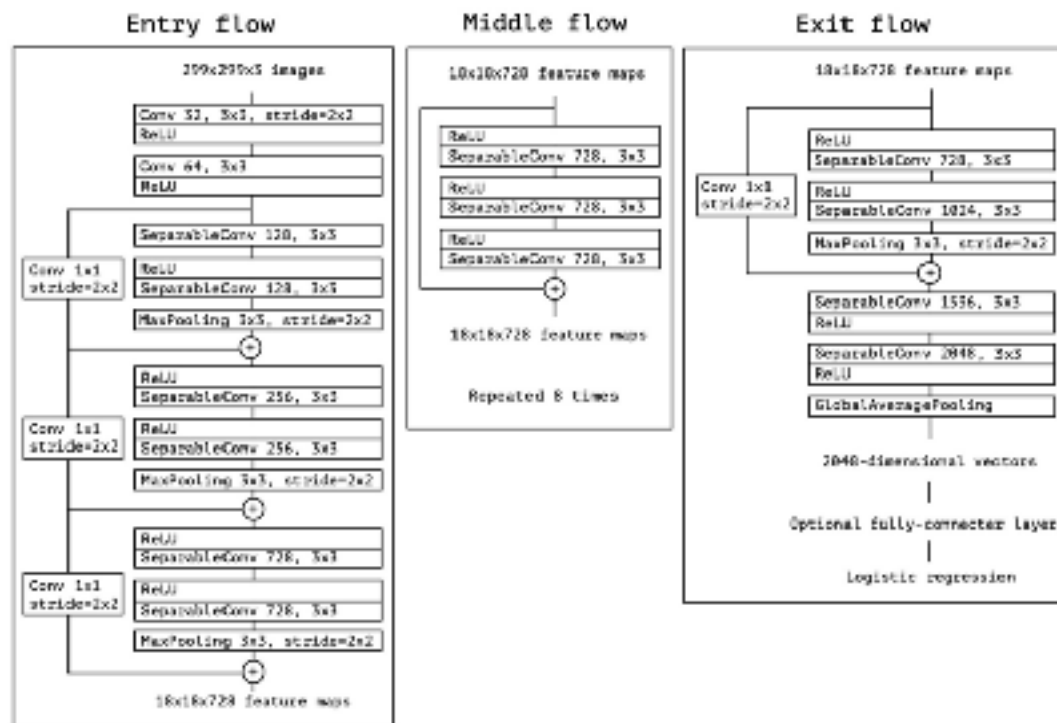
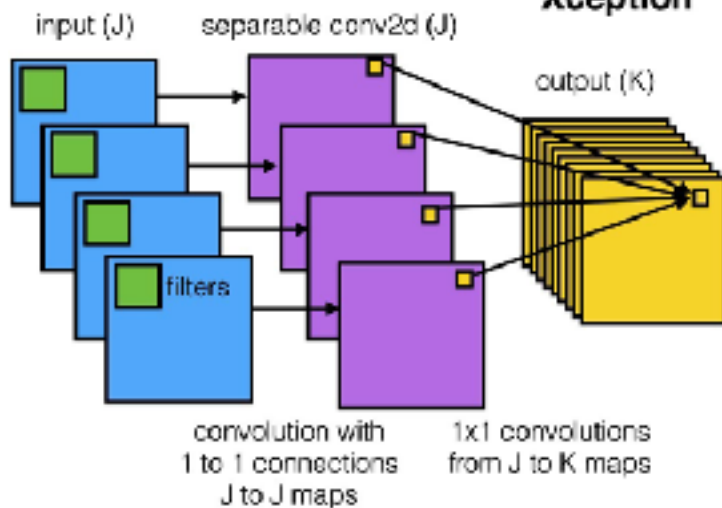
Xception

- Major Contributions:
 - combining branching / residual blocks
 - separable convolutions



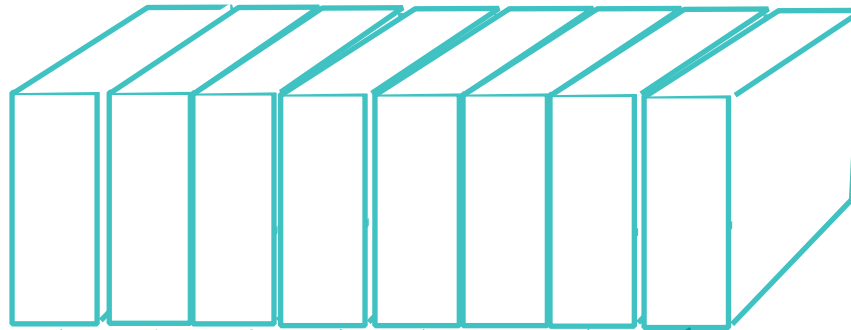
Francois Chollet
Google

Xception

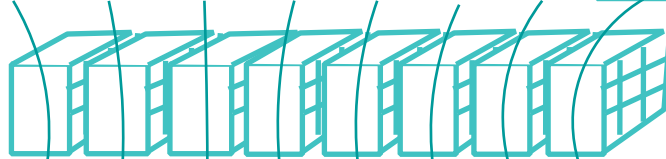


<https://arxiv.org/pdf/1610.02357.pdf> 57

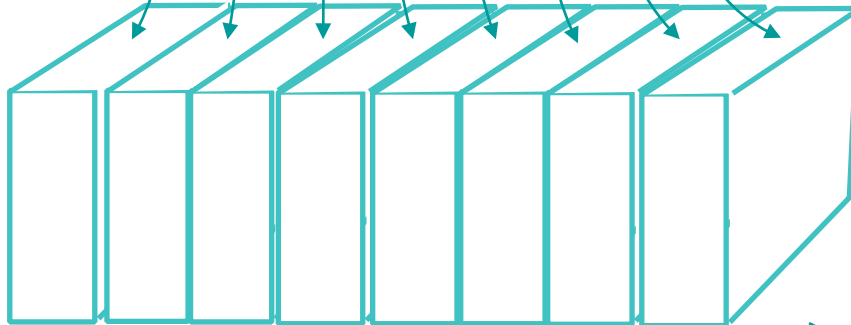
Separable Convolution Primer



Inputs, From Layer N-1
Num Channels = J



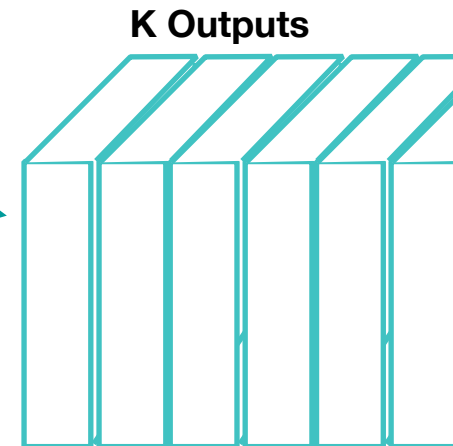
Filters, Layer N
Convolve Each Channel Separately
Num filters = J



Concat Channel Outputs
Num Intermediate Channels = J

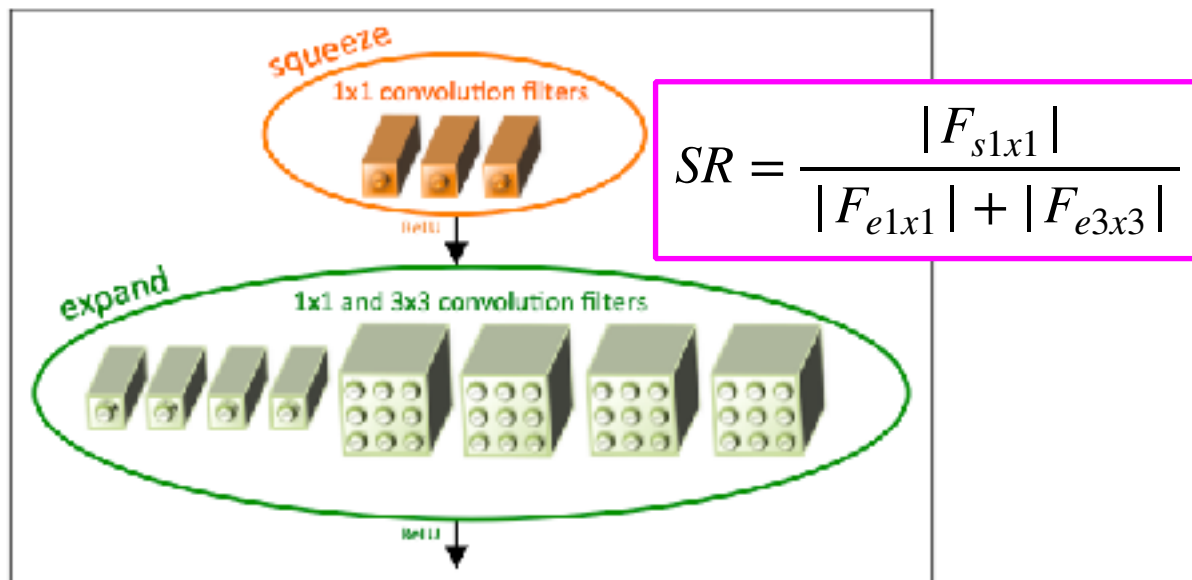


Perform K 1x1
Traditional Convolutions
to get K Outputs



SqueezeNet (2018)

- Idea: squeeze and expand in each layer
 - Use mostly 1x1 filters,
 - downsample later in network,
 - reduce channels before 3x3 filters



SQUEEZENET: ALEXNET-LEVEL ACCURACY WITH 50X FEWER PARAMETERS AND <0.5MB MODEL SIZE

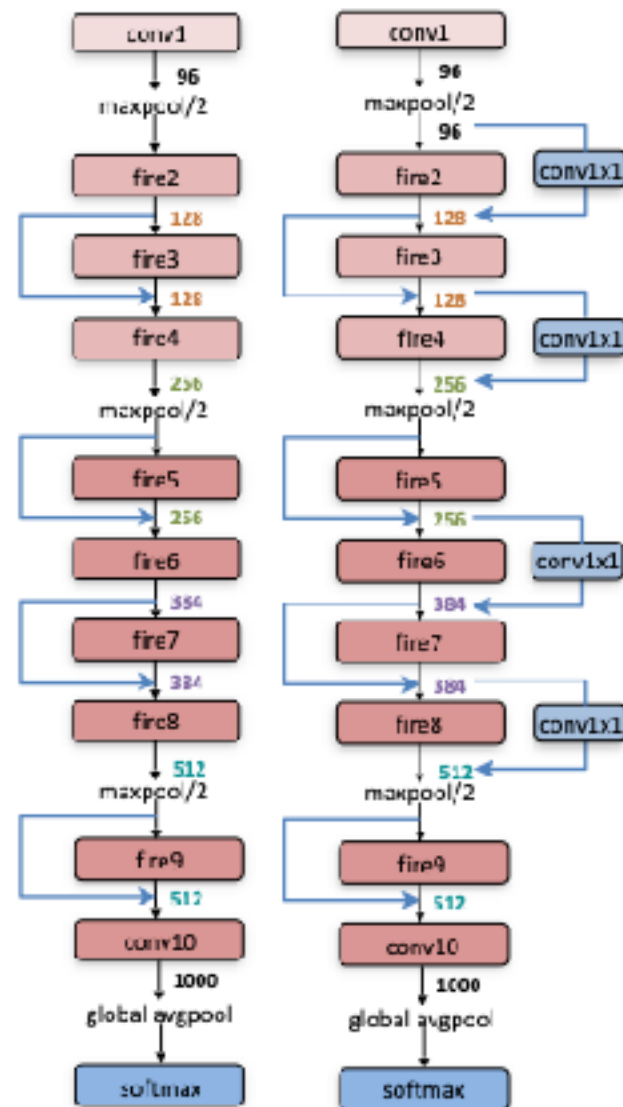
Forrest N. Iandola¹, Song Han², Matthew W. Moskewicz¹, Khalid Ashraf¹,

William J. Dally², Kurt Keutzer¹

¹DeepScale* & UC Berkeley ²Stanford University

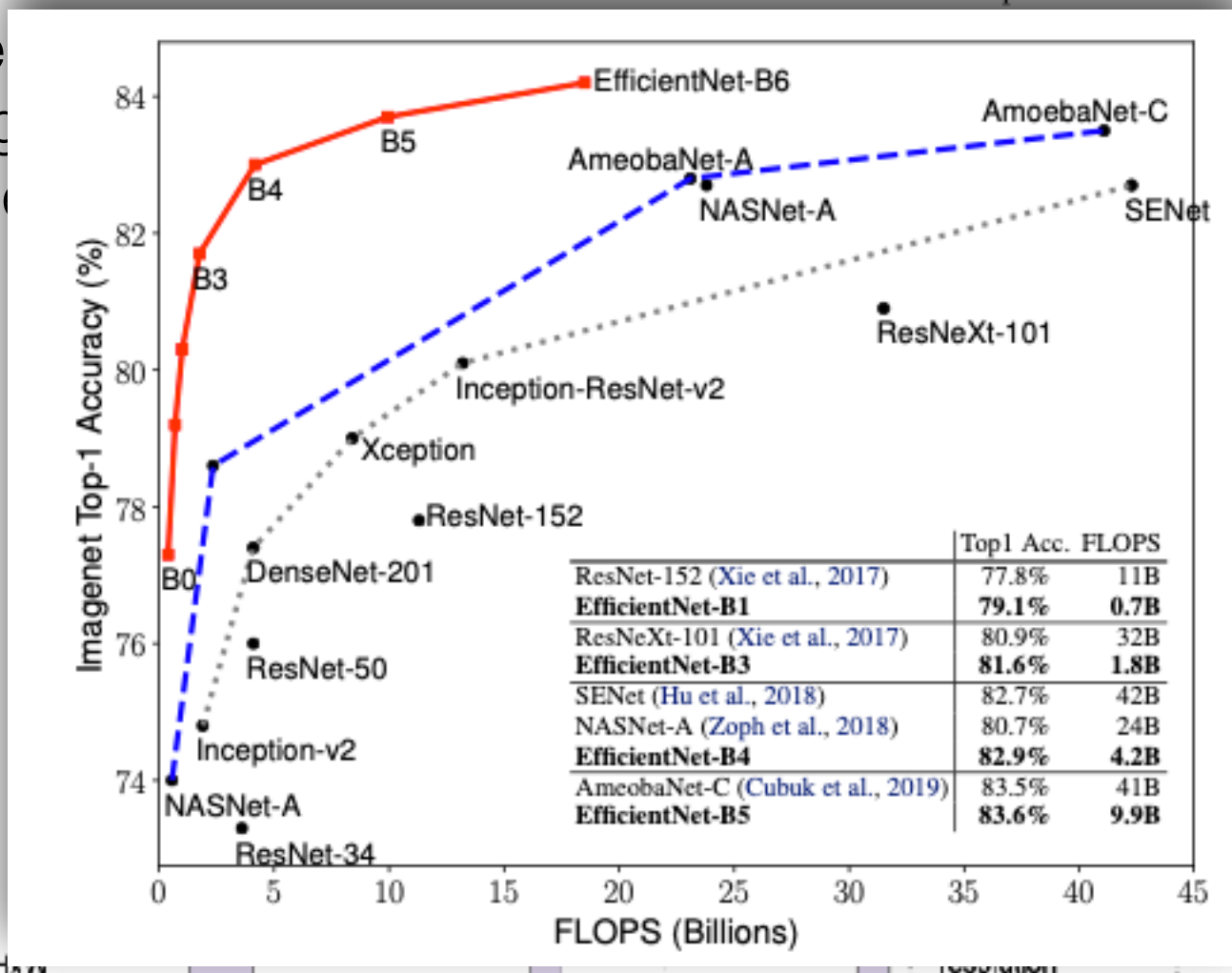
{forresti, moskewicz, kashraf, keutzer}@eecs.berkeley.edu

{songhan, dally}@stanford.edu



Efficient Net (2019)

- Define scaling between



2
 $\gamma \geq 1$
 ular, we find
 $\alpha = 1.2, \beta =$
 $2, \gamma^2 \approx 2.$

wider

higher resolution

(a) baseline

(b) width scaling
channels

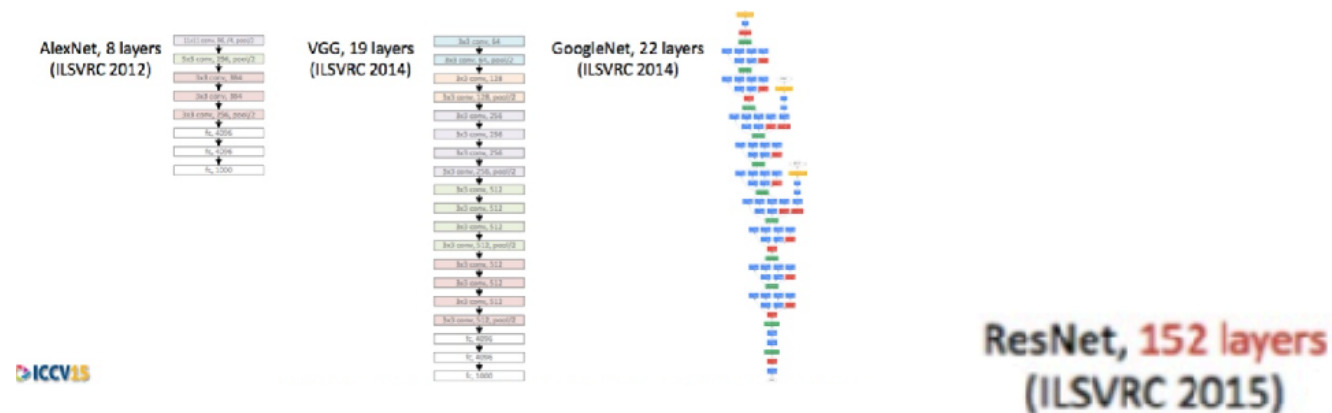
(c) depth scaling
layers

(d) resolution scaling
pooling

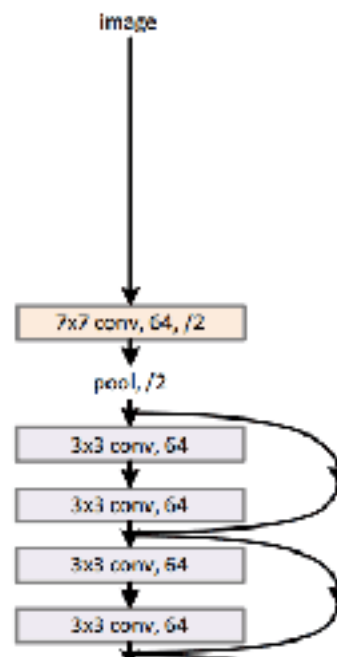
(e) compound scaling

How big are these networks?

How big are these networks?



34-layer residual



Self Test

- We have seen a lot of different networks.
- The most important concept to understand in using convolutional neural networks is:
 - A. Use proper initialization of layers
 - B. Have plenty of data or use expansion
 - C. Set aside time for training
 - D. Use batch/group normalization

Even more Convolutional
Neural Networks
...in TensorFlow
...with Keras

“if time” Demo



12. More Advanced CNN Techniques.ipynb

Next Time:

- CNN Lab Discussion (Town Hall, if needed)
- Intro to Recurrent Neural Network Architectures
 - RNNs, GRUs, LSTMs
 - Training for characters