Introduction to Deep Learning for Computer Vision

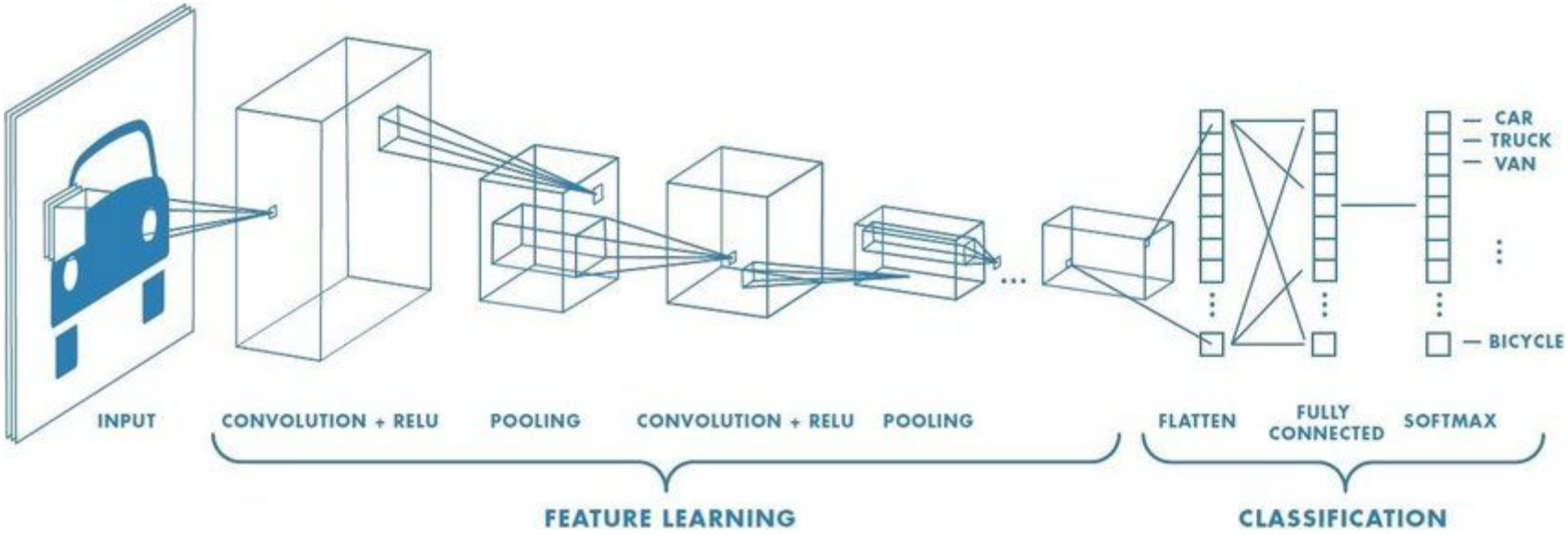
Adhyayan '23 - ACA Summer School
Department of Computer Science and Engineering
Indian Institute of Technology Kanpur

Lecture 4

Course Logistics

- Reminder: Quiz 1 will happen at the end of this week.
  - Quiz 1 will be done in a google form.
  - The form will be released at 12 PM on 18.06.2023 (Sunday)
  - The form will stop taking responses at 12 AM on 19.06.2023 (Monday) [12 hours after the release].
  - Questions will be MCQ.
  - Open-book test: Students are free to use any online resource they want to.
- Reading Materials regarding Week 1 contents are posted in the course ReadMe File.

CNN Recap


CAR
TRUCK
VAN
BICYCLE
INPUT
CONVOLUTION + RELU
POOLING
CONVOLUTION + RELU
POOLING
FLATTEN
FULLY CONNECTED
SOFTMAX
FEATURE LEARNING
CLASSIFICATION

CNN Recap: Important Terminologies

1. **Convolution**: The process of applying a filter/kernel to an input image to extract relevant features.

2. **Kernel**: A small matrix used in the convolution operation that is applied to the input image to extract features. The kernel contains learnable parameters that are adjusted during training to capture specific patterns or characteristics in the input data.

3. **Feature Map**: The output map generated by convolving the input with a set of learned filters.

4. **Pooling**: Downsampling the feature map by summarizing local information, such as taking the maximum value (*Max Pooling*) or average value (*Average Pooling*).

5. **Stride**: The number of pixels by which the convolutional filter/kernel is shifted during the convolution operation.

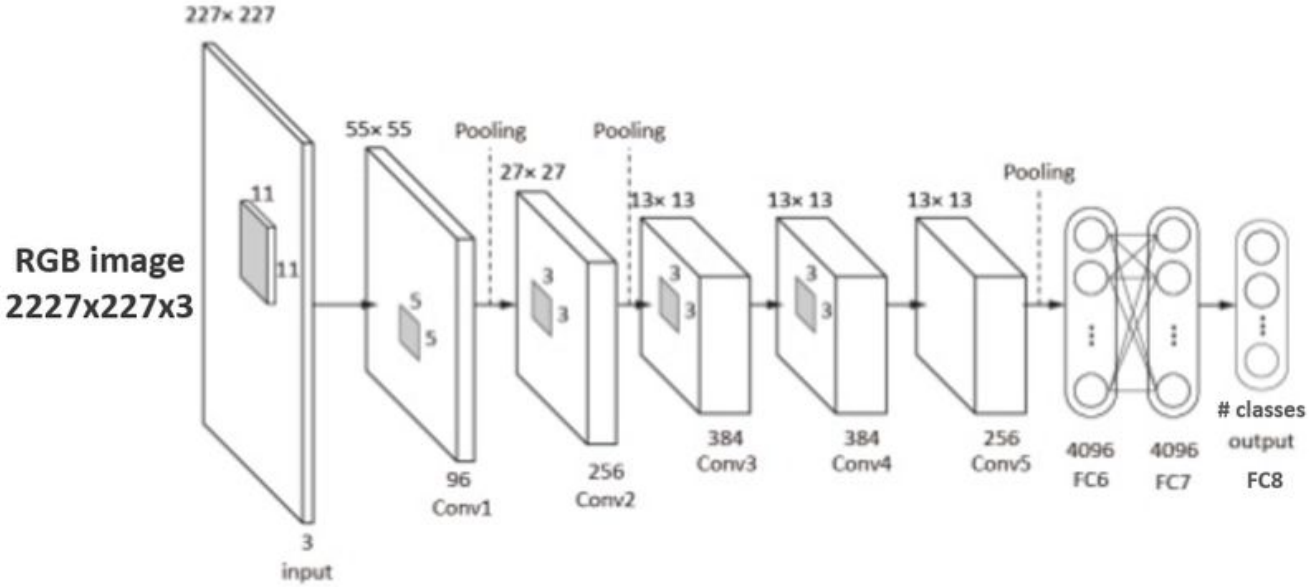
6. **Padding**: Adding additional pixels to the input image to preserve spatial dimensions and avoid information loss during convolution. Can be *valid* or *same*.


ImageNet Large Scale Visual Recognition Challenges

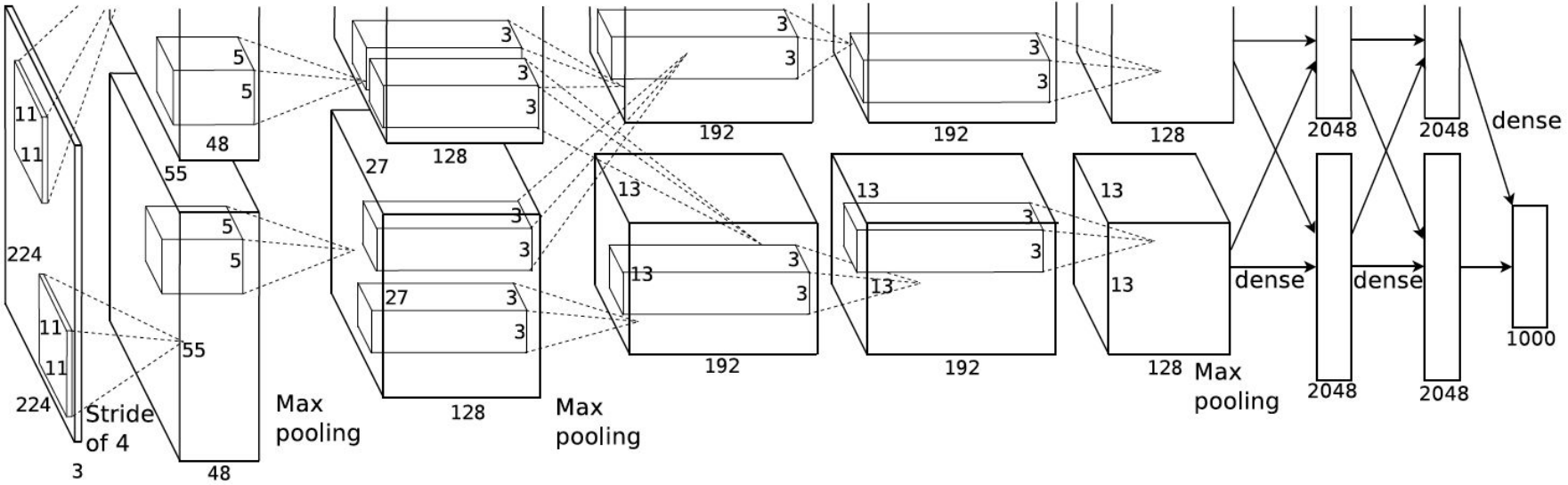

ILSVRC Challenge Winners (2010-2020)
Human Error Rate
Error Rate (%)
35
30
25
20
15
10
5
0
25.8
0
0
0
0
0
5.1%
Xerox (2011)

AlexNet (2012)

- Developed by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton.
- Multiple convolutional layers, ReLU activations, max pooling layers, and fully connected layers.



AlexNet (2012)



First CNN to be trained on GPUs parallelly.

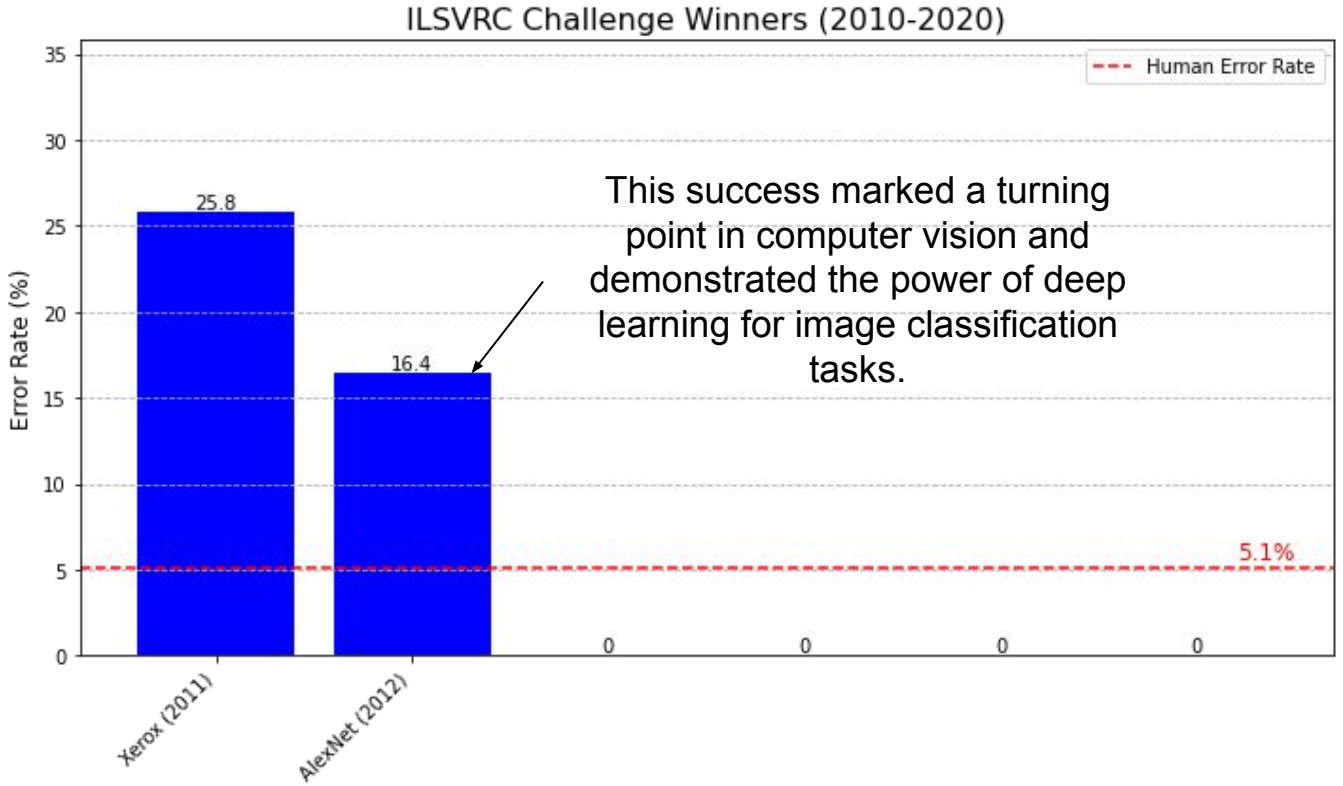
AlexNet (2012)

Hidden Layer		Design
Convolution	1	96 filters in size 11x11 with max pooling in size 3x3
	2	256 filters in size 5x5 with max pooling in size 3x3
	3	384 filters in size 3x3 without pooling in size 3x3
	4	384 filters in size 3x3 without pooling
	5	256 filters in size 3x3 with max pooling in size 3x3
Fully Connected	1	4096 nodes with LeakyRelu activation function
	2	4096 nodes with LeakyRelu activation function
	3	100 nodes with LeakyRelu activation function

61M Params!

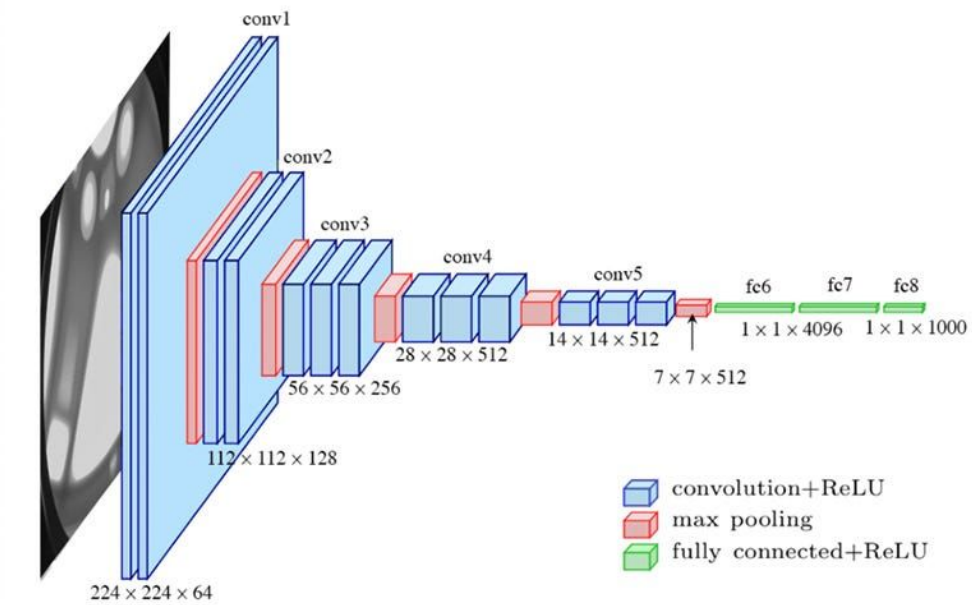
- In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012, AlexNet achieved a top-5 error rate of 16.4%, significantly outperforming other methods.
- AlexNet architecture served as a foundation for subsequent CNN designs like VGGNet, GoogLeNet, and ResNet.

AlexNet (2012)



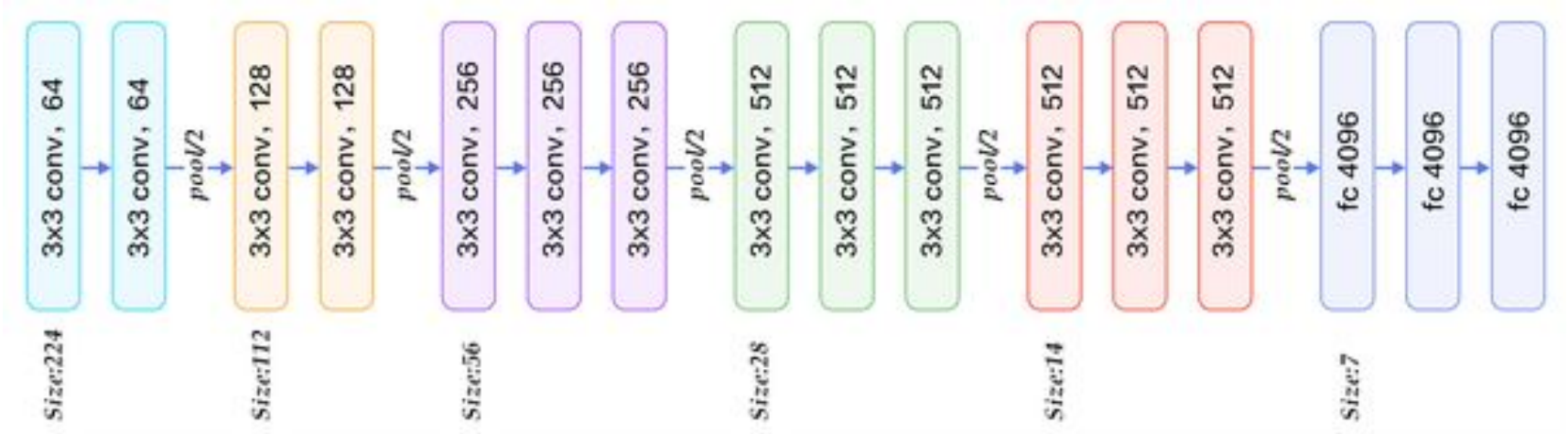
VGGNet (2014)

Why not go even deeper?



VGGNet (2014)

Variants: **VGG16** and **VGG19**, which have 16 and 19 weight layers, respectively.

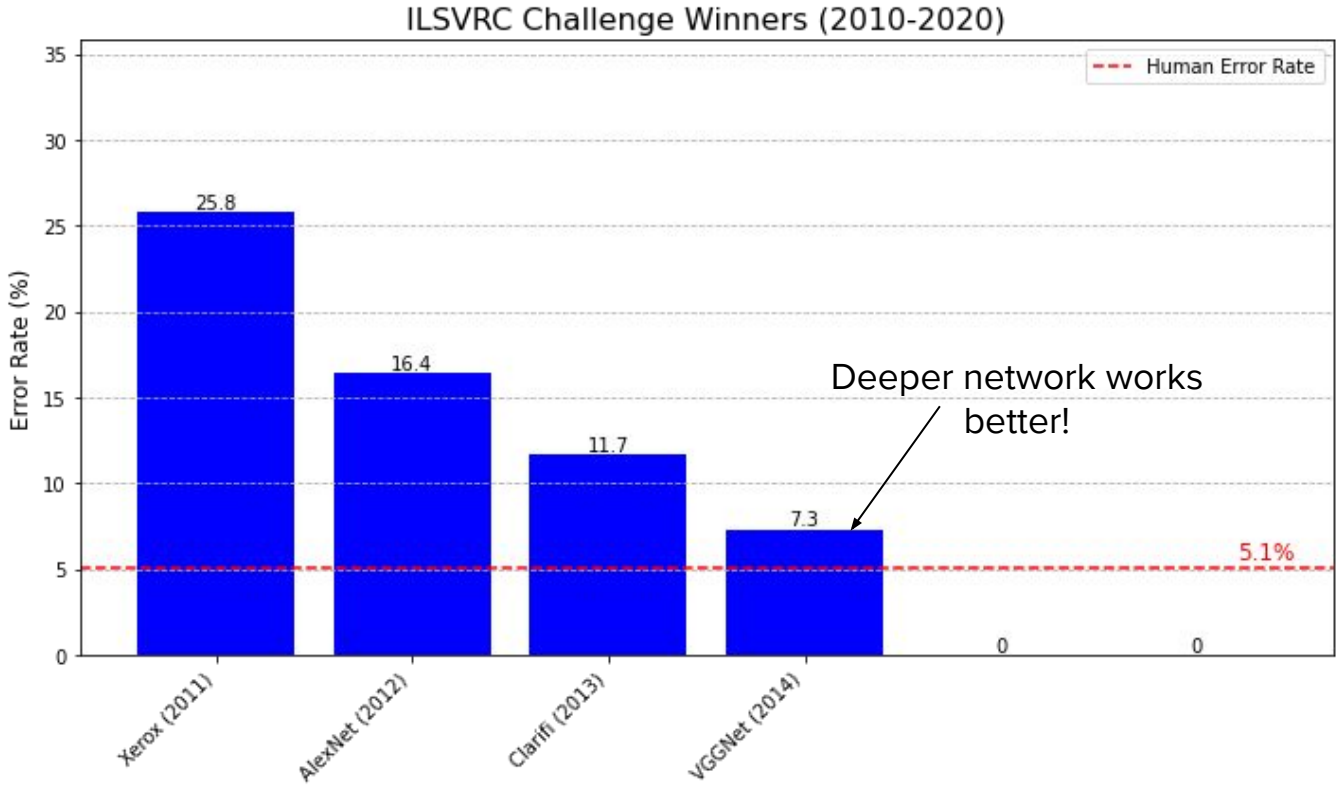


VGGNet (2014)

Variants: **VGG16** and **VGG19**, which have 16 and 19 weight layers, respectively.

144M Params!

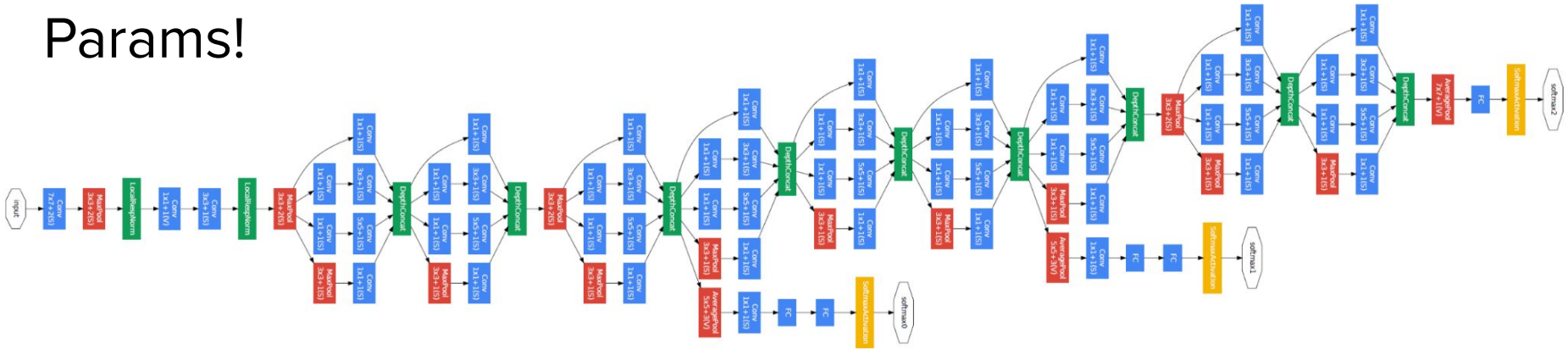
VGGNet (2014)


ILSVRC Challenge Winners (2010-2020)
Human Error Rate
Error Rate (%)
35
30
25
20
15
10
5
0
25.8
16.4
11.7
7.3
0
0
5.1%
Deeper network works better!
Xerox (2011)
AlexNet (2012)
Clarifi (2013)
VGGNet (2014)

GoogleNet (2014) - The Rise of 1x1 Convolution!

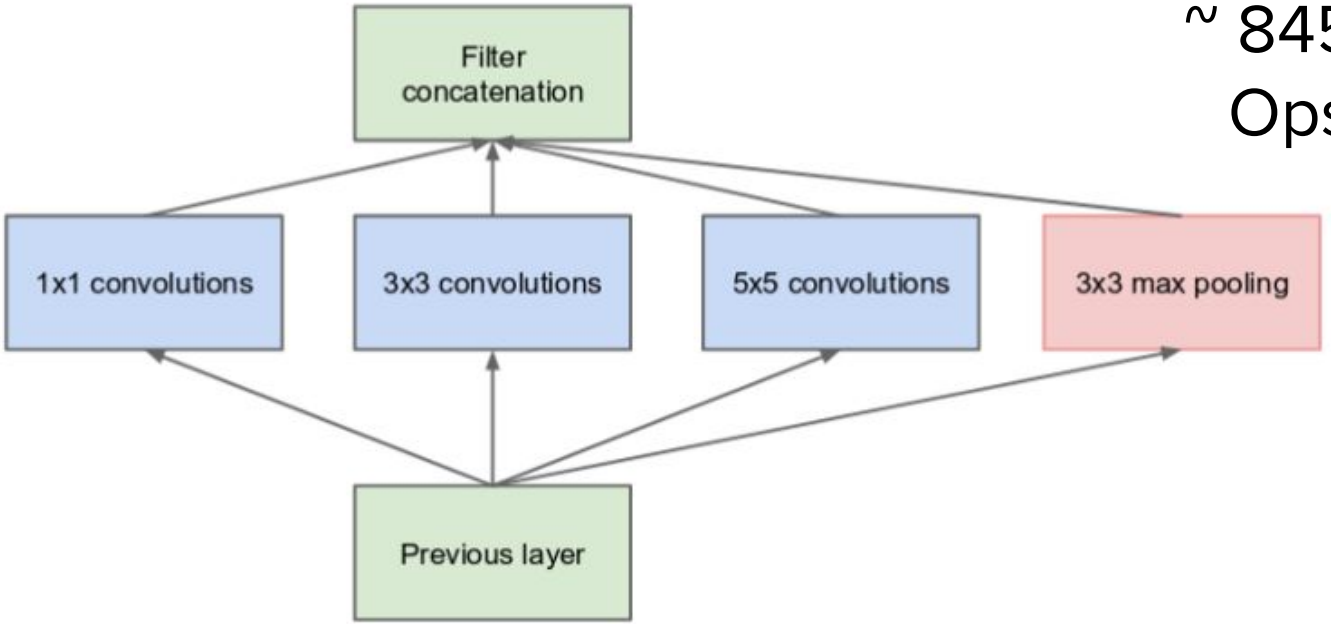
Going even deeper! But, there's a catch!

6M
Params!



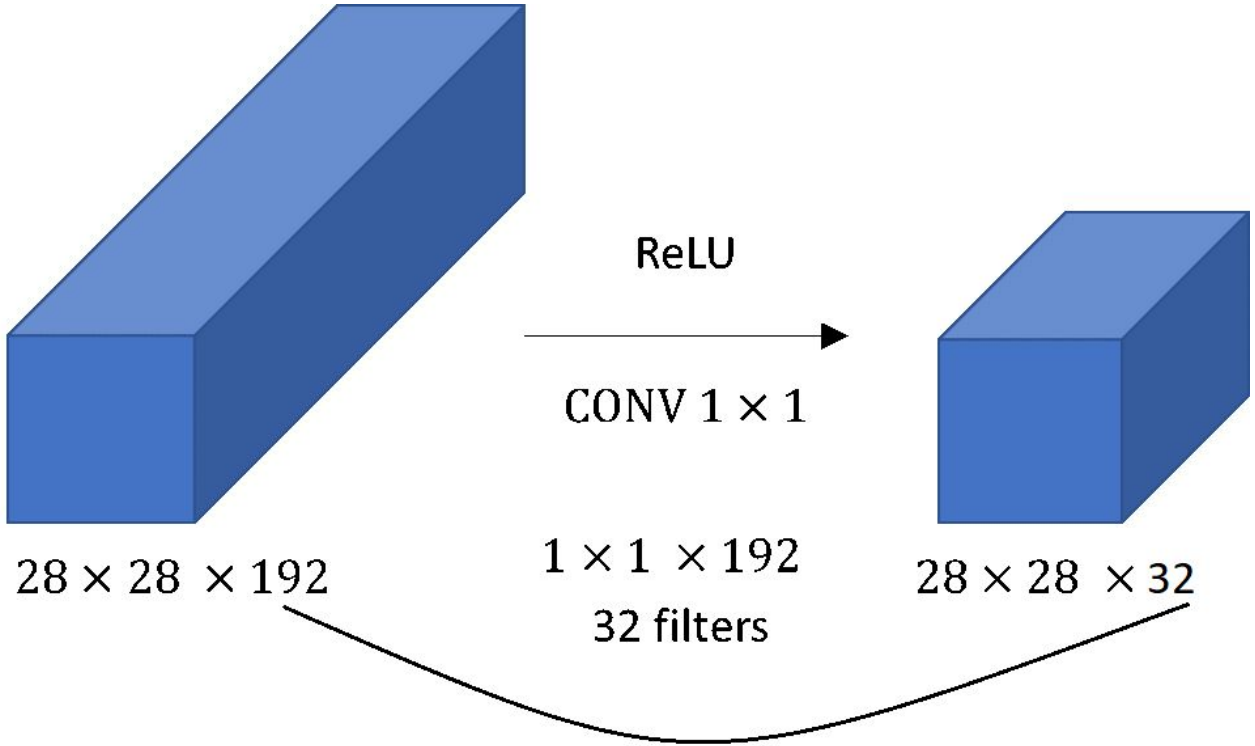
GoogleNet (2014) - Inception Module

~ 845M Ops!



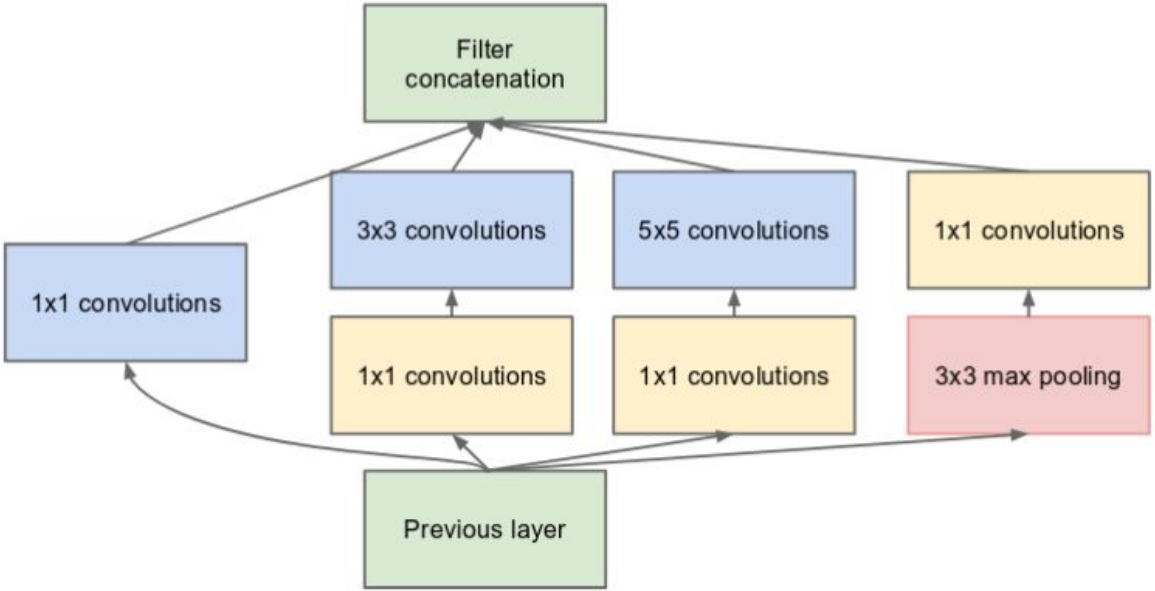
(a) Inception module, naïve version

GoogleNet (2014) - 1 x 1 Convolution



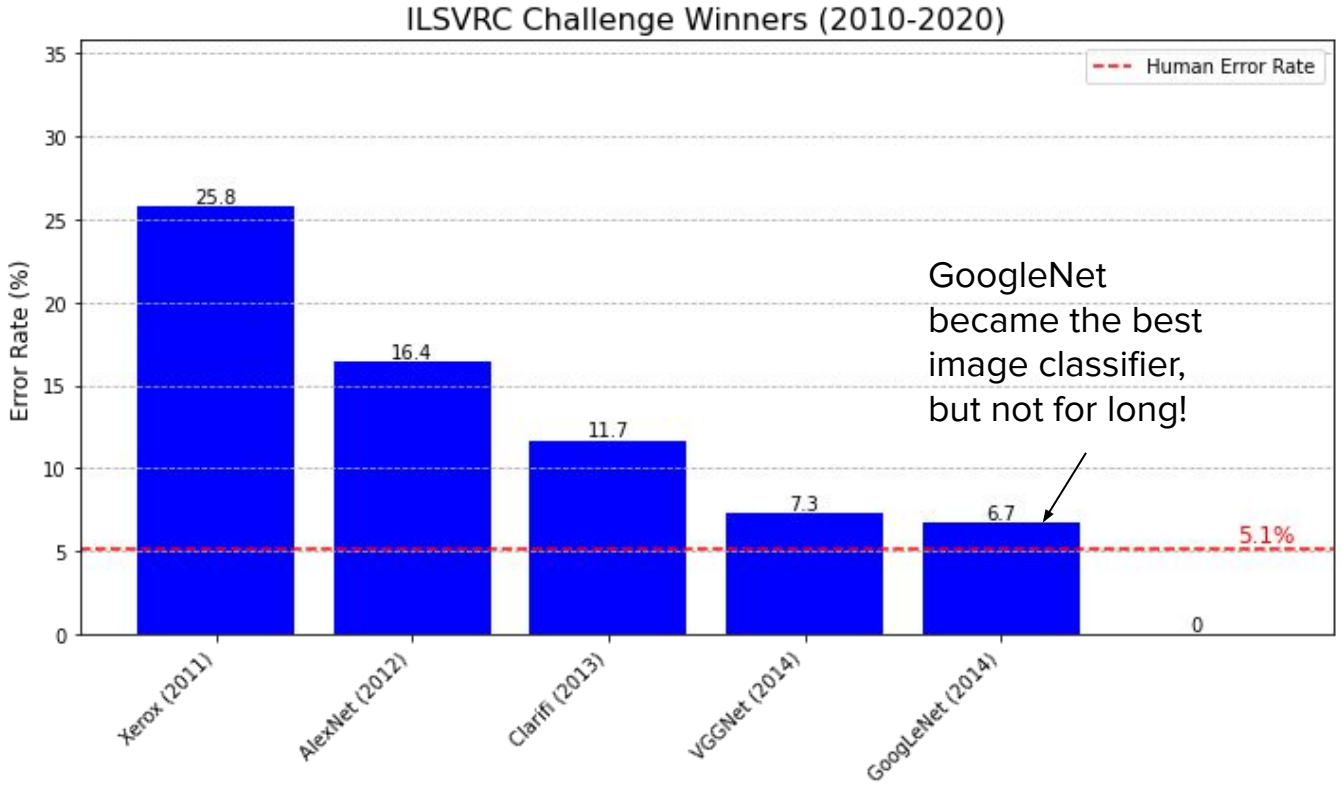
GoogleNet (2014) - Inception Module

~ 358M Ops!

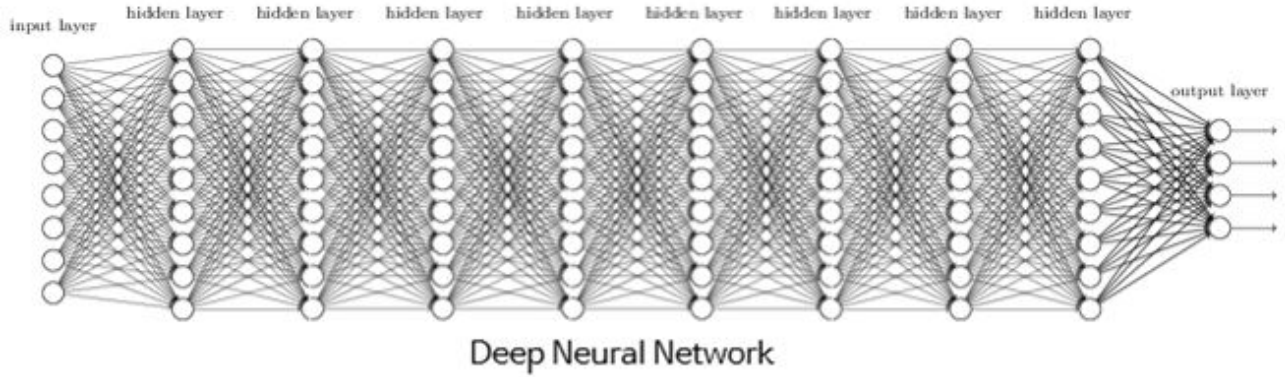


(b) Inception module with dimension reductions

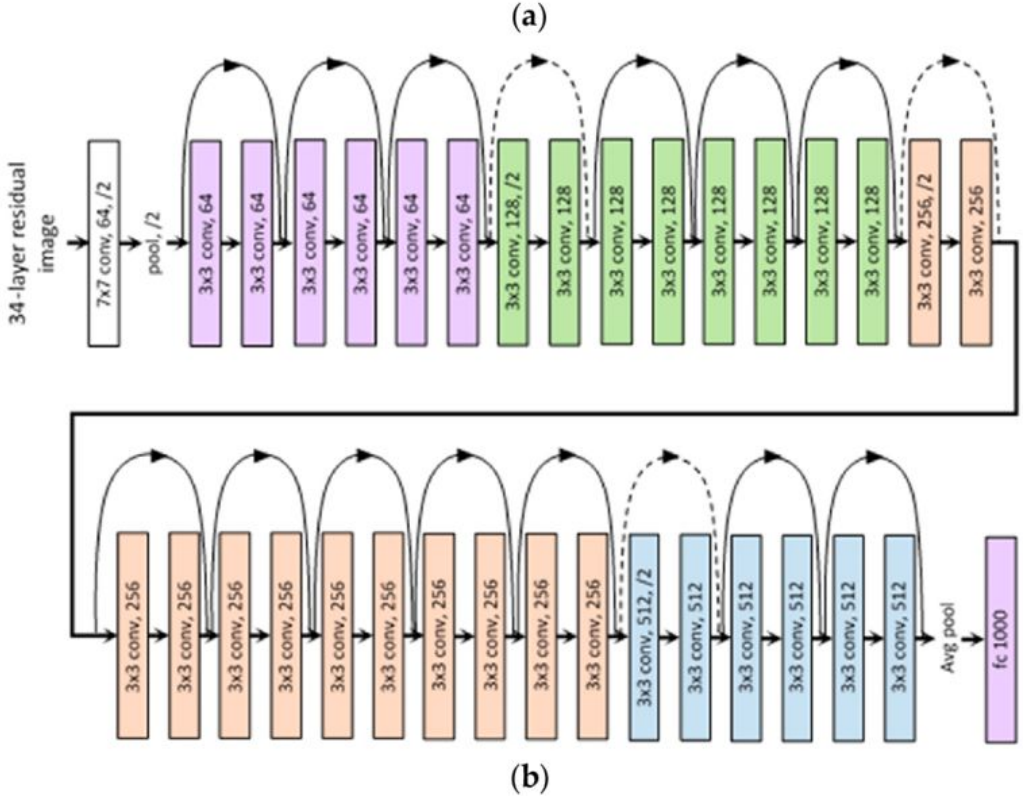
GoogleNet (2014)



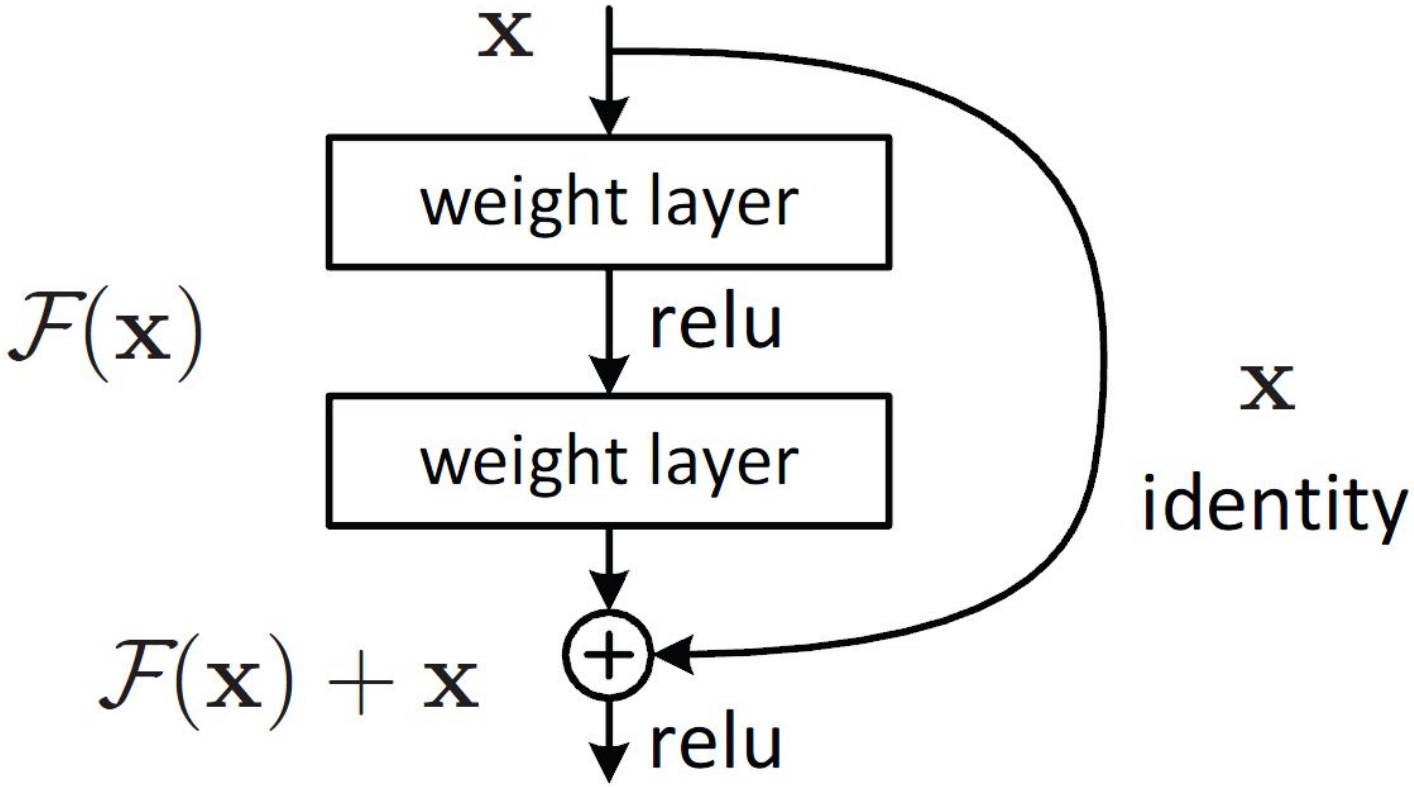
Vanishing Gradient Problem



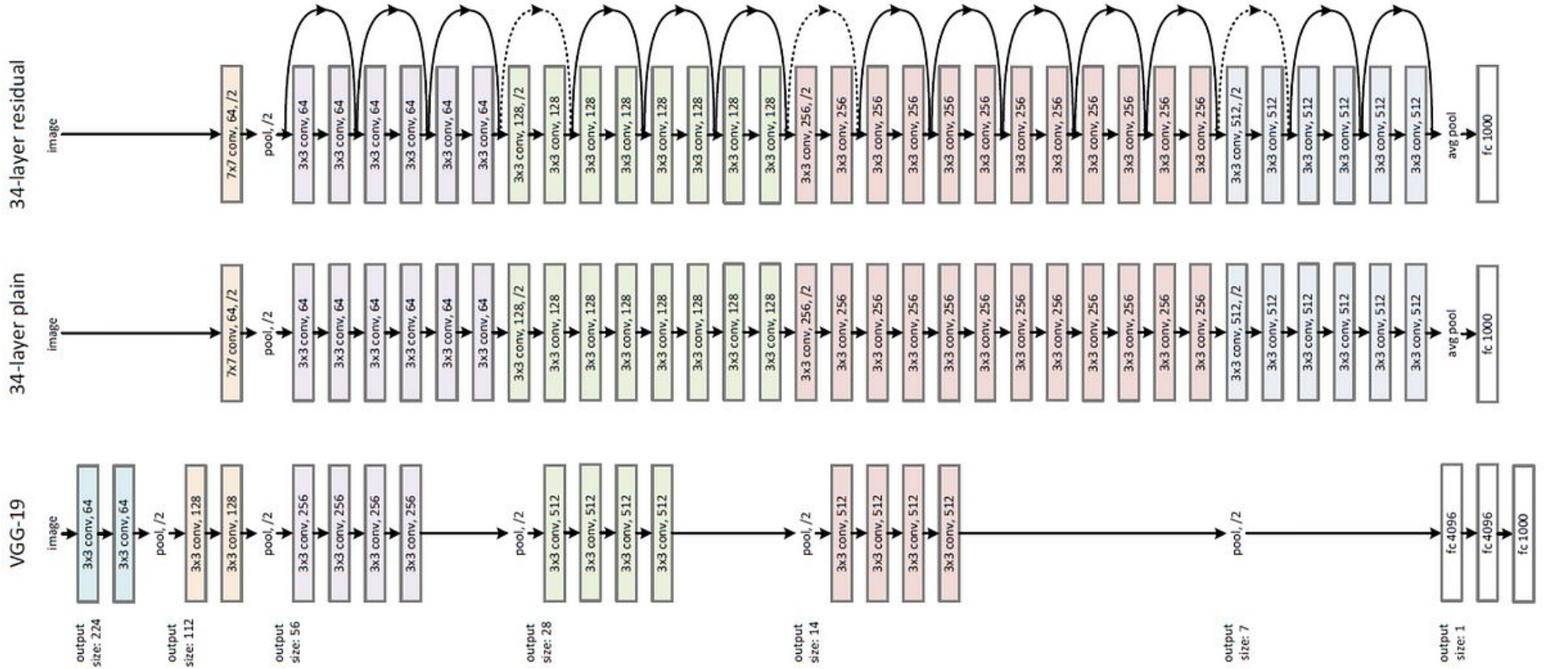
ResNet (2015)


(a)
34-layer residual
image
7x7 conv, 64, /2
pool, /2
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 256, /2
3x3 conv, 256
3x3 conv, 512, /2
3x3 conv, 512
Avg pool
fc 1000
(b)

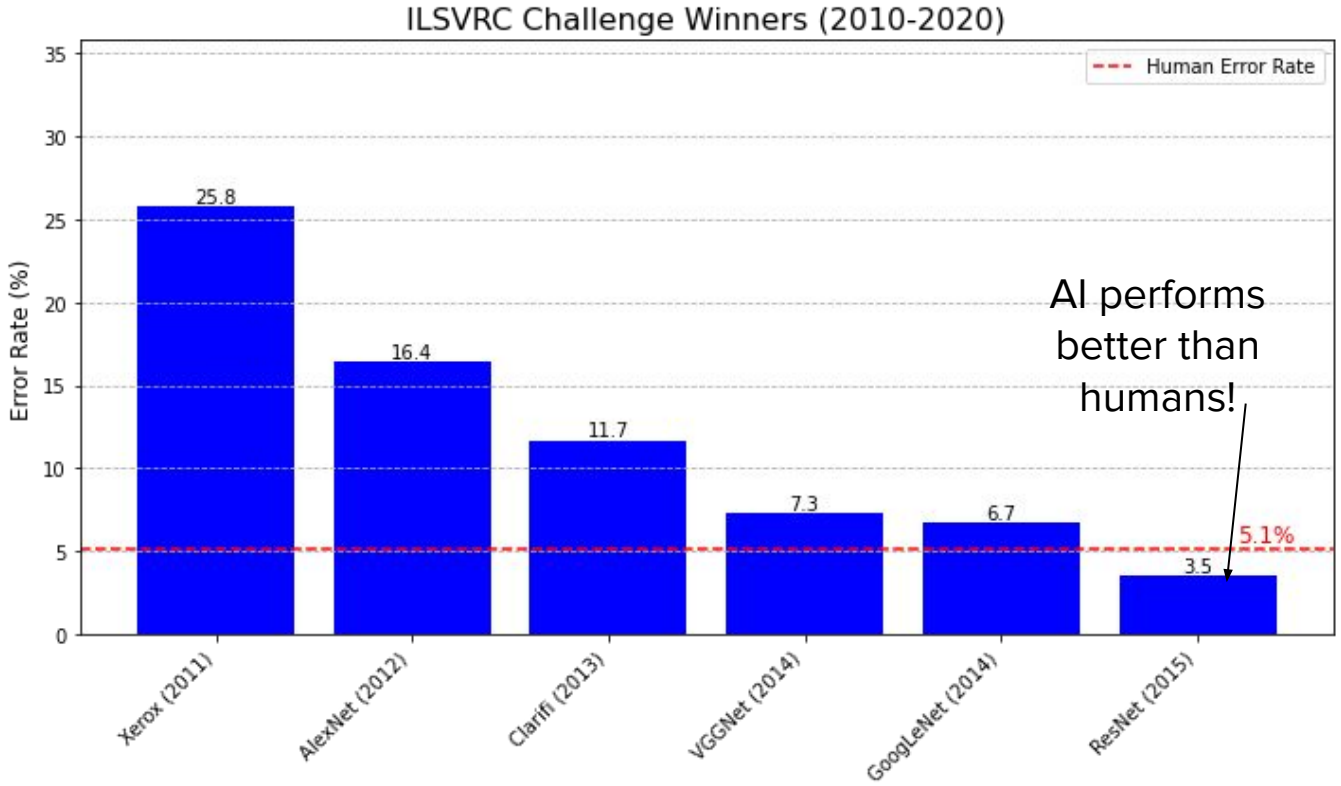
ResNet (2015) - Residual Connection



ResNet (2015) - Deep isn't a problem!



ResNet (2015)


ILSVRC Challenge Winners (2010-2020)
Human Error Rate
Error Rate (%)
35
30
25
20
15
10
5
0
25.8
16.4
11.7
7.3
6.7
3.5
5.1%
AI performs better than humans!
Xerox (2011)
AlexNet (2012)
Clarifi (2013)
VGGNet (2014)
GoogLeNet (2014)
ResNet (2015)

Next Lecture: Training Neural Networks better!