DATA 442: Neural Networks & Deep Learning

Dan Runfola – danr@wm.edu

icss.wm.edu/data442/

What is a “Good” Learning rate?

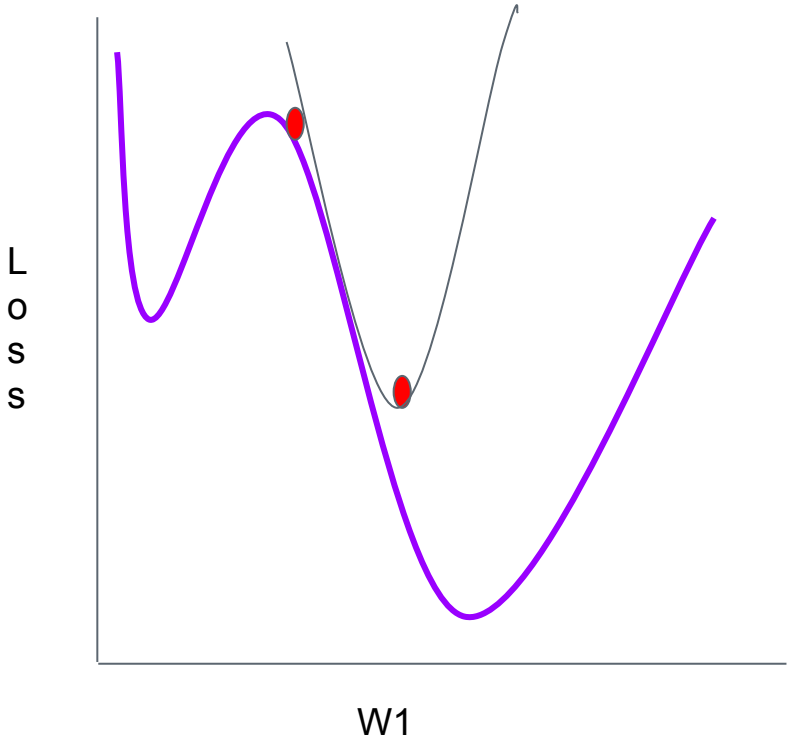
Step Decay - Every *k* iterations, the learning rate is cut by half.

Exponential Decay:

$$\alpha_{i+1} = \alpha_i e^{-ki}$$

Inverse Decay:

$$\alpha_{i+1} = \alpha_i / (1 + ki)$$



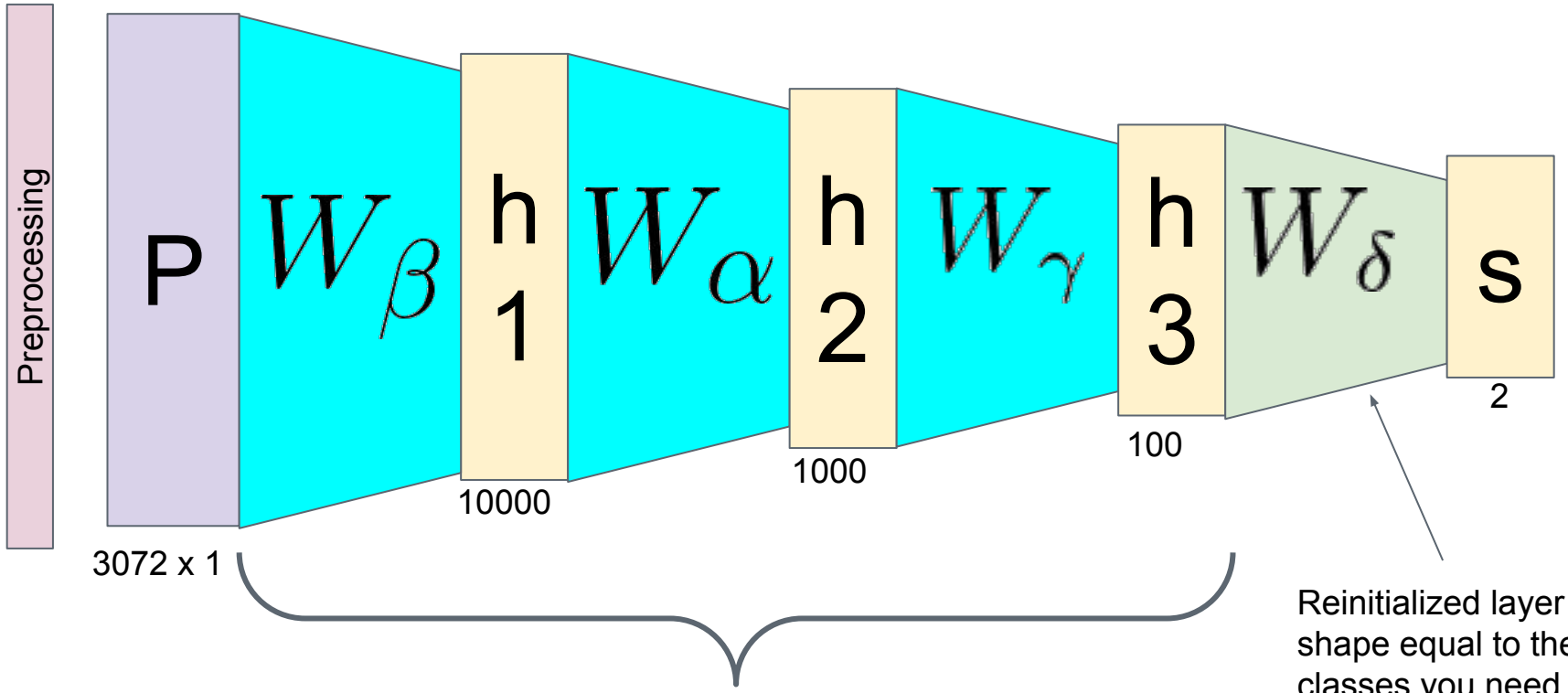
BFGS - Broyden Fletcher Goldfarb Shanno algorithm.
Approximates Hessian with low-rank updates.

ORIGINAL

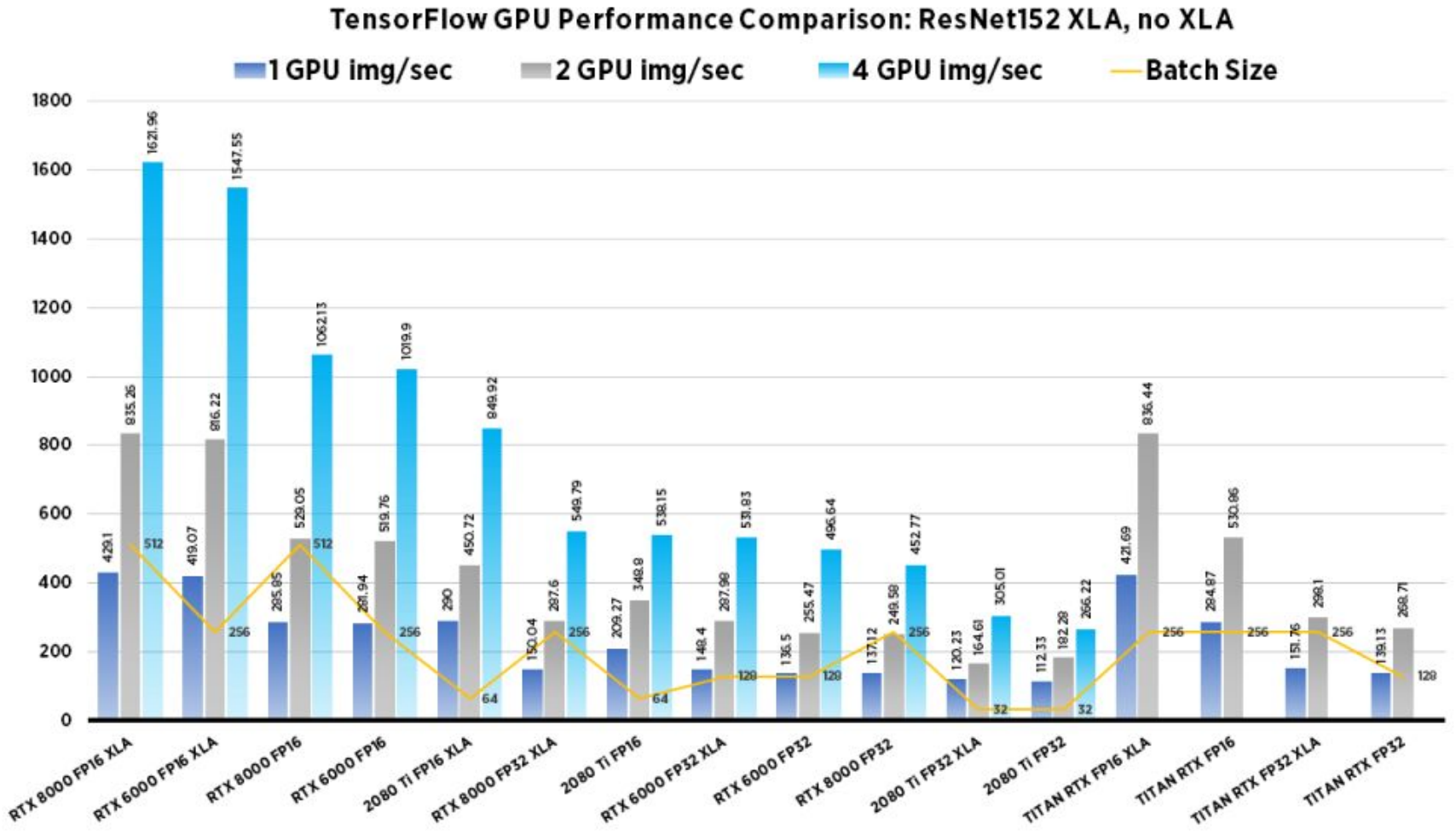
FLIP

Contrast / Brightness

Cropping



```
Epoch 11: Iteration 4: Estimated Time Remaining ==============>>>>>>>>>>>>>>>>>>>> 345 hours, 32 minutes, 12 seconds
```



CPU

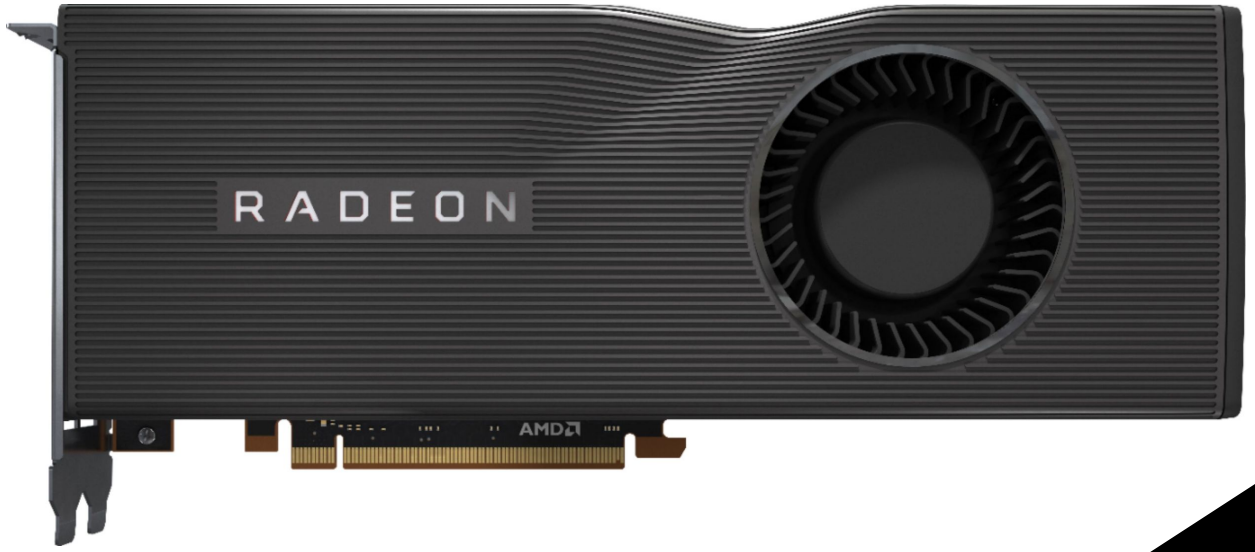

GTX 1070

GPU


PSU
CD-ROM
RAM
CPU
VIDEO CARD
HARD DRIVE
FAN
MOTHERBOARD

Graphics Processing Unit

Graphics Processing Unit



Graphics Processing Unit

CPU

Only a Few Cores
(Counted in Hundreds, at most, and normally in ten or less).

Very fast clock speeds (4GHz +)

Uses Physical Memory located Elsewhere on Motherboard

GPU

CPU

Only a Few Cores
(Counted in Hundreds, at most, and normally in ten or less).

Very fast clock speeds (4GHz +)

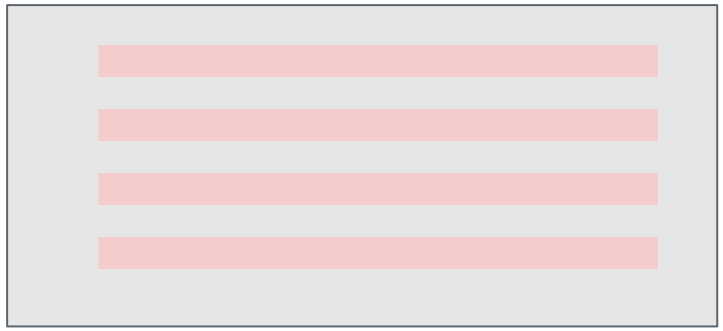
Uses Physical Memory located Elsewhere on Motherboard

GPU

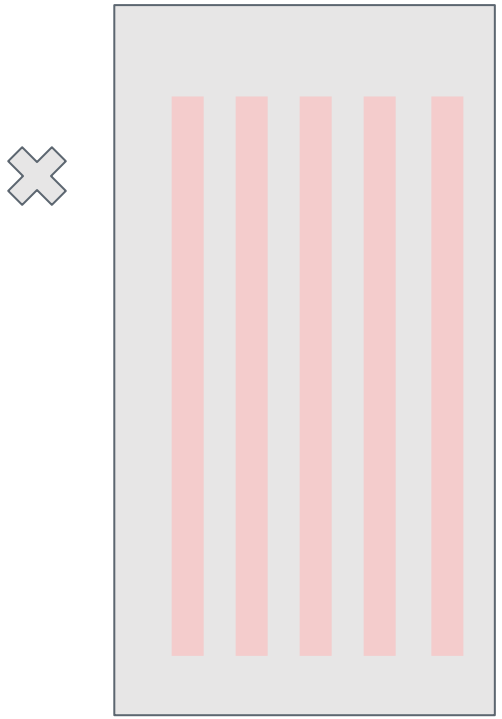
LOTS of cores
Thousands - i.e., a RTX 2080 TI has 4,352 cores.

Slower clock speeds (1-2 GHz)

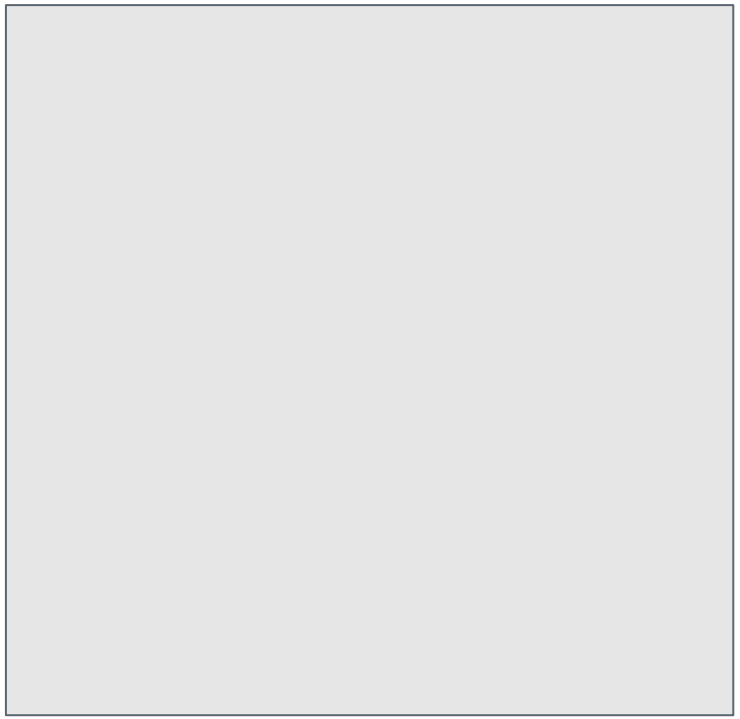
Physical memory is integrated with the card.



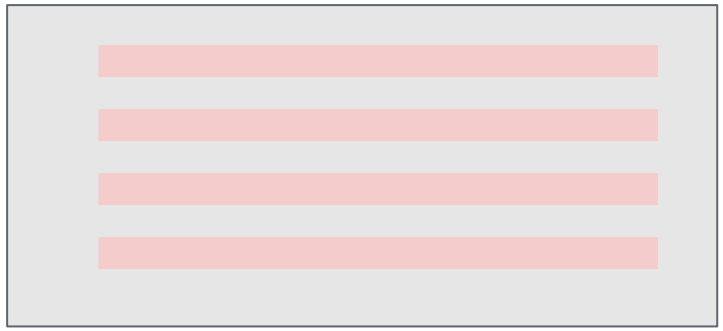
4X5 Matrix



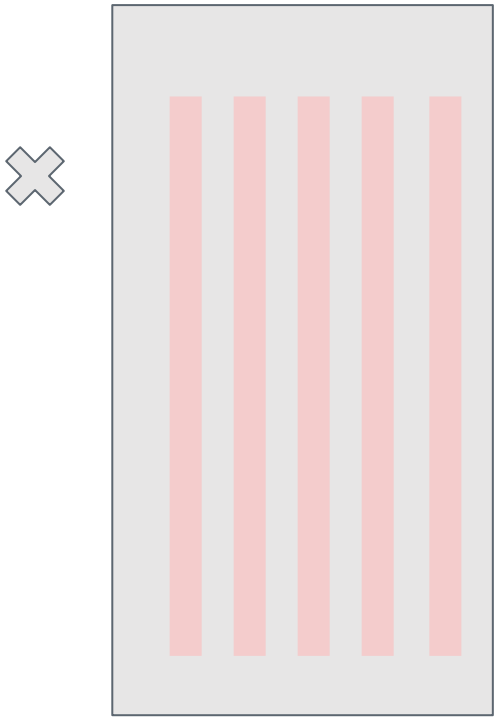
5x5 Matrix



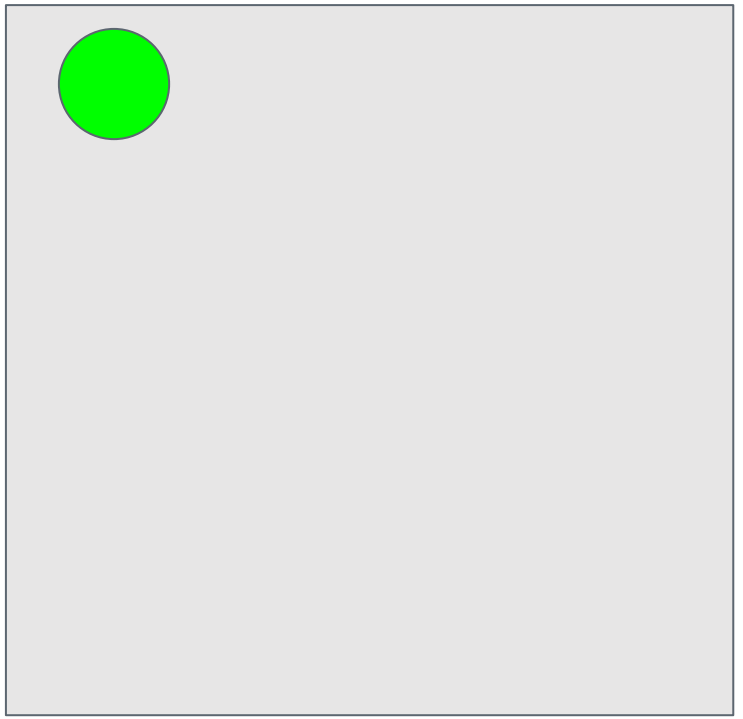
4x5 Matrix



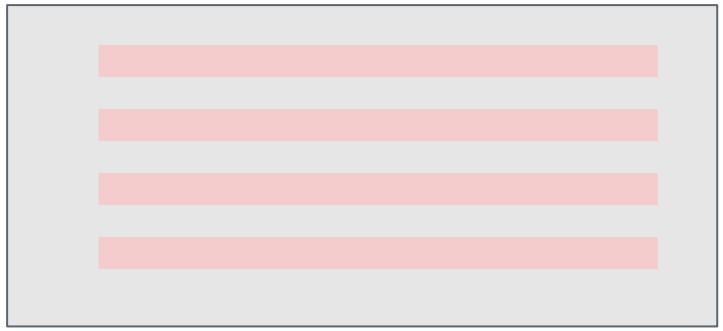
4X5 Matrix



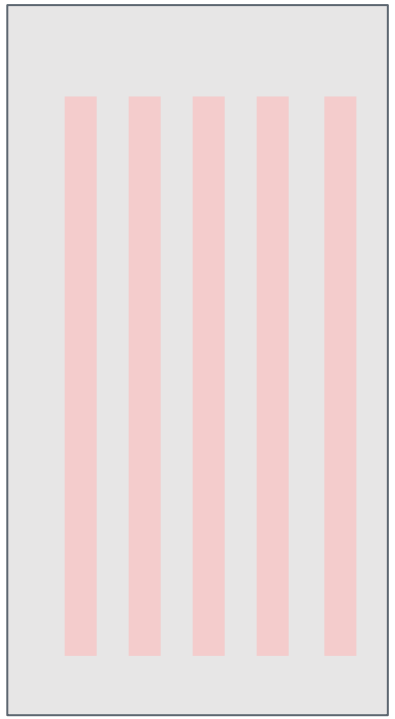
5x5 Matrix



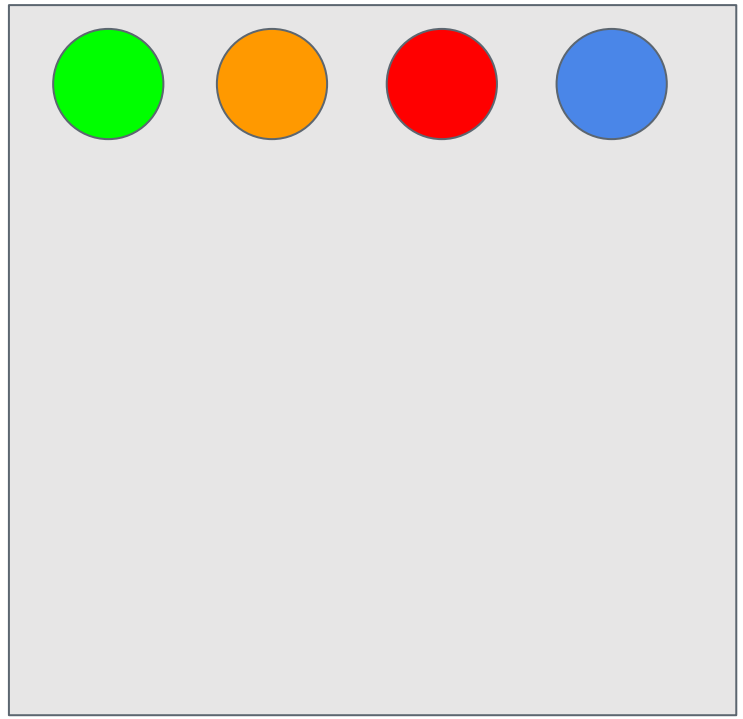
4x5 Matrix



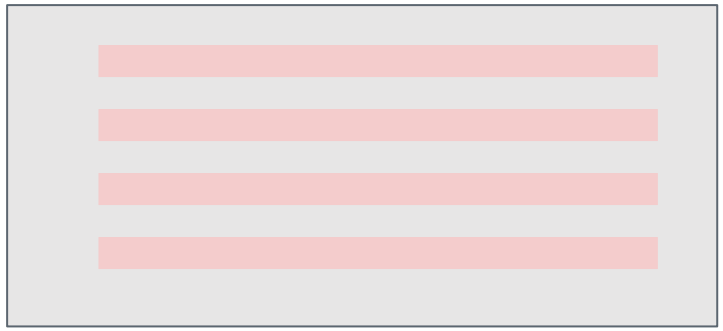
4X5 Matrix



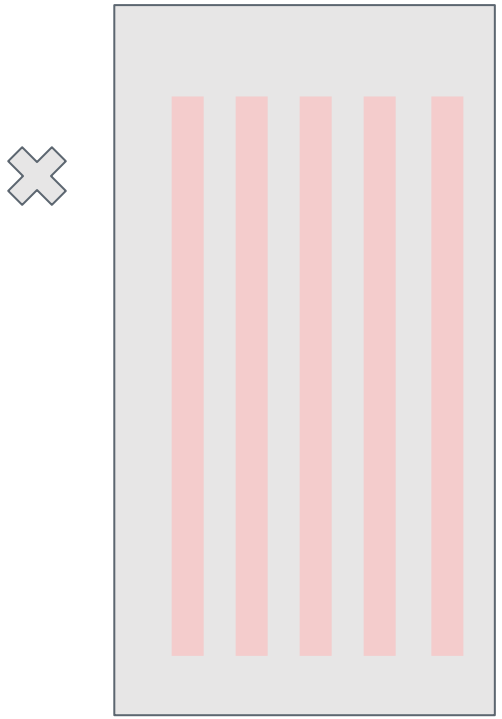
5x5 Matrix



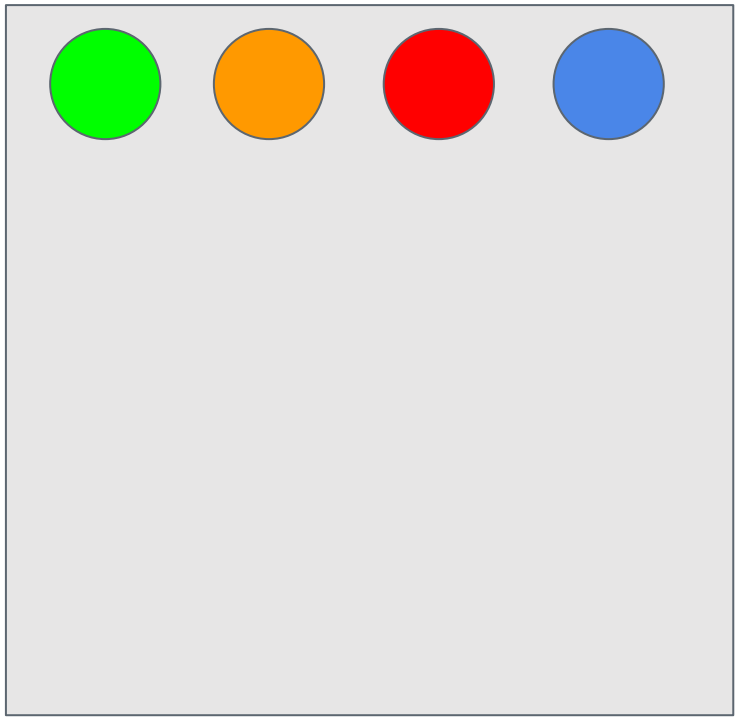
4x5 Matrix



4X5 Matrix



5x5 Matrix



4x5 Matrix

LOTS of cores

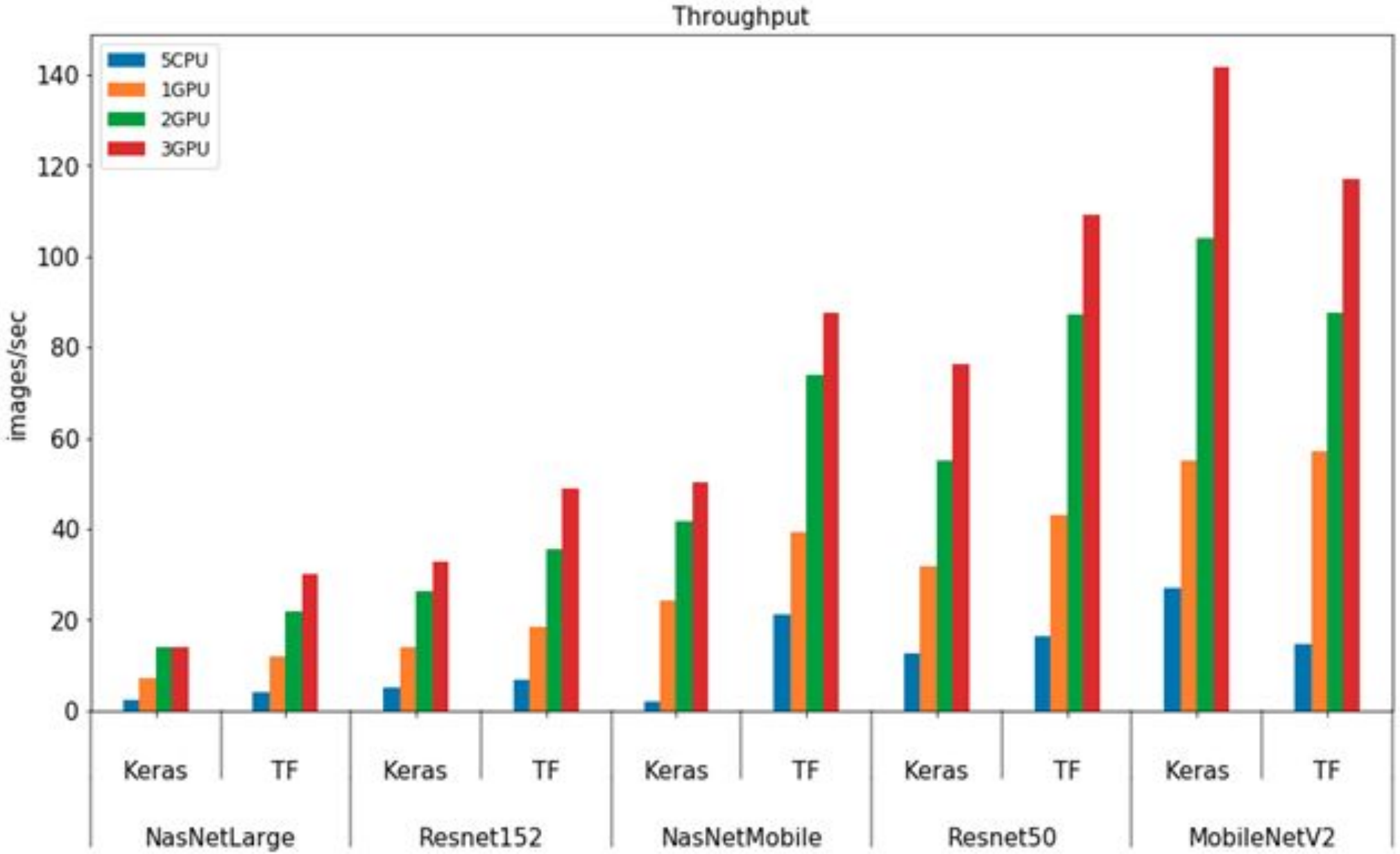
Thousands - i.e., a RTX 2080 TI has 4,352 cores.

Slower clock speeds (1-2 GHz)

Physical memory is integrated with the card.

Low-level GPU Programming

Comparison:
https://arxiv.org/vc/arxiv/papers/1005/1005.2581v1.pdf



https://azure.microsoft.com/en-us/blog/gpus-vs-cpus-for-deployment-of-deep-learning-models/


PSU
CD-ROM
RAM
CPU
VIDEO CARD
HARD DRIVE
FAN
MOTHERBOARD

Deep Learning Frameworks

Majors Players:

Torch / PyTorch (Facebook)

TensorFlow / Keras (Google)

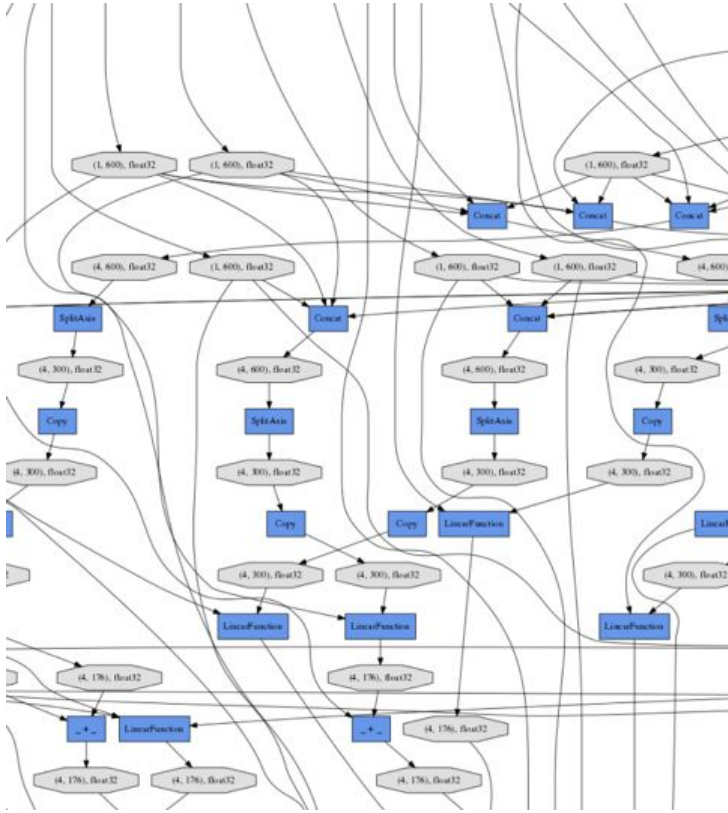
Old / Less Used / Integrated / Non-English:

Caffe (UC Berkeley); Theano (U Montreal); Caffe 2 (Facebook); PaddlePaddle (Baidu); CNTK (MSFT); MXNET (Amazon, MIT, CMU)

https://www.paddlepaddle.org.cn/

What do these things do?

- Simplify Building our Computational Graphs
  - Including computing gradients (autoGrad!)
- Integrate with GPUs and other dedicated cards.



What framework to use?

Torch / PyTorch

(~) TorchHub

TensorFlow / Keras

(+) TensorHub

What framework to use?

Torch / PyTorch

(~) TorchHub

(+) Dynamic Graphs

TensorFlow / Keras

(+) TensorHub

(~) Static Graphs

What framework to use?

Torch / PyTorch

(~) TorchHub

(+) Dynamic Graphs

(-) Expert Readable

TensorFlow / Keras

(+) TensorHub

(~) Static Graphs

(+) Human Readable (Keras!)

What framework to use?

Torch / PyTorch

(~) TorchHub

(+) Dynamic Graphs

(-) Expert Readable

(~) Fast

TensorFlow / Keras

(+) TensorHub

(~) Static Graphs

(+) Human Readable (Keras!)

(~) TF: Fast

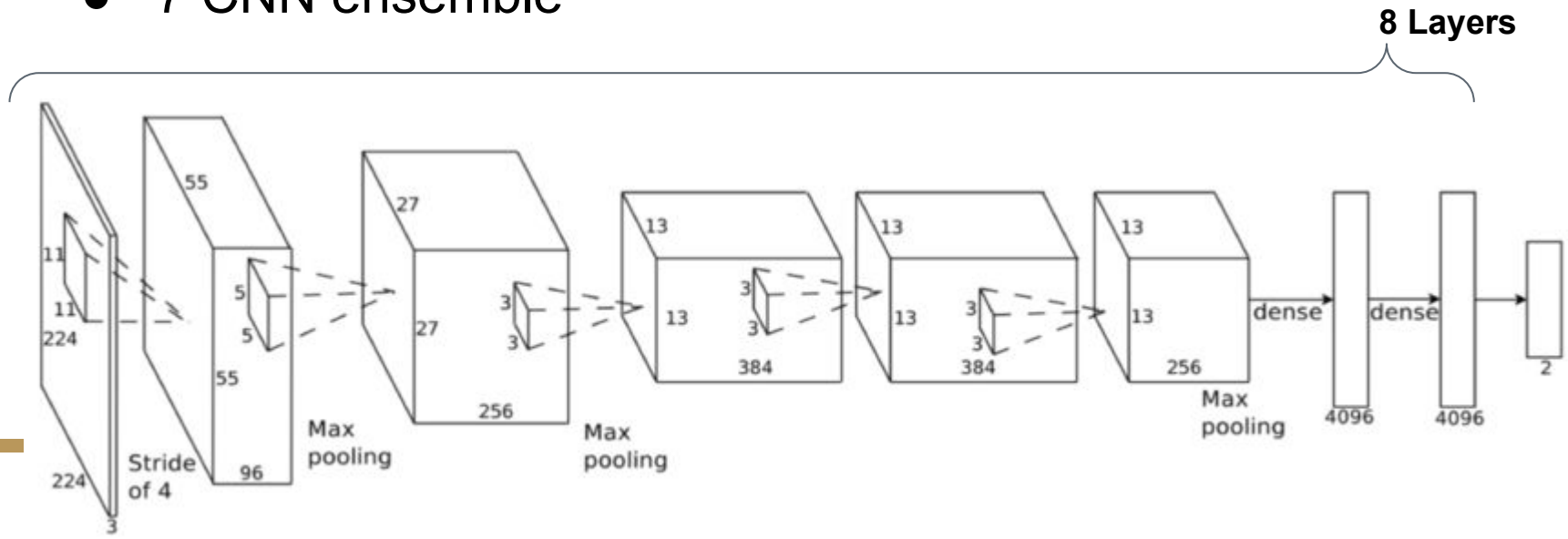
(-) Keras: Slower

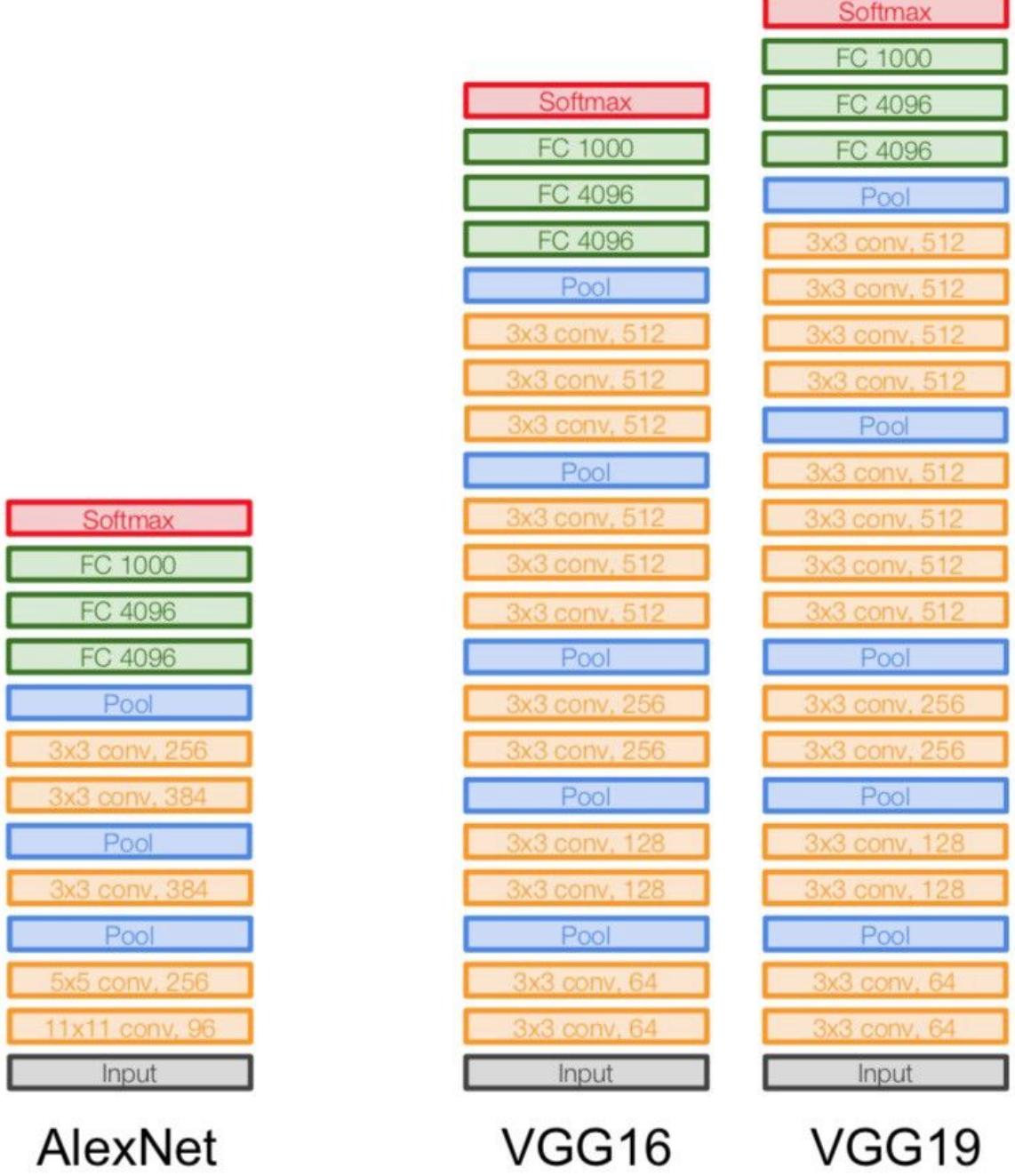
Next Up: Model Architecture

- What layer(s) should I use for my problem?
- How should I design my convolutions?
- **What have successful models done?**

AlexNet

- First CNN winner of the ImageNet annual competition (2012)! Classification Error: 16.4%
- First integration of ReLU with CNN architecture in the competition
- Manually Adapted Learning Rate (1e-2, reduced when accuracy no longer improving)
- 7 CNN ensemble



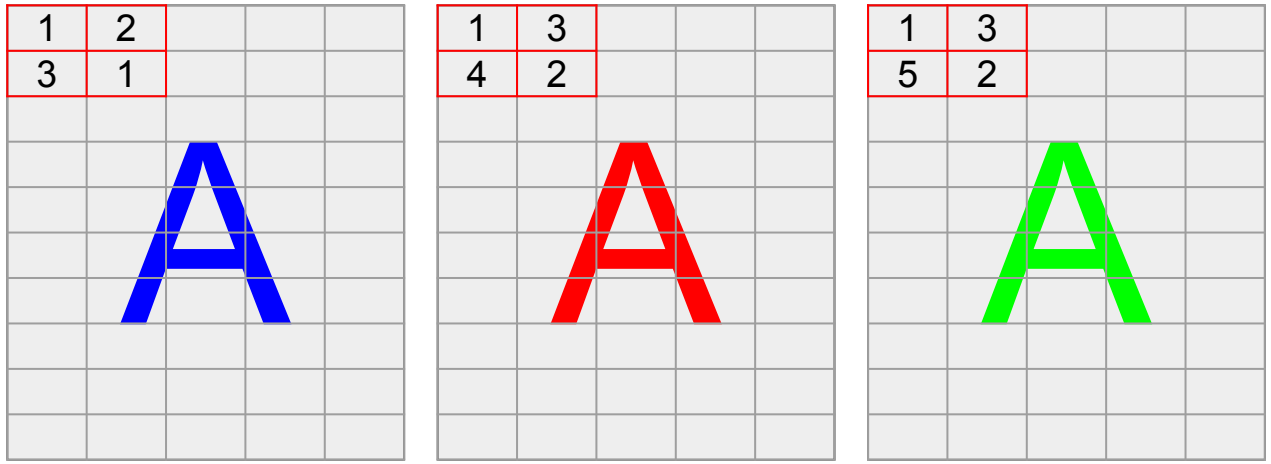


VGGNet

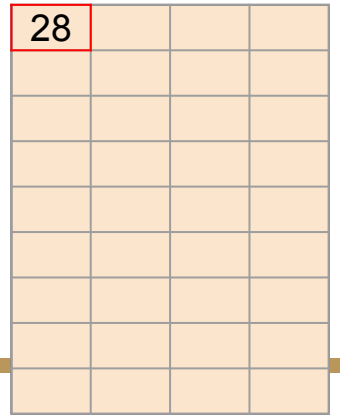
- 16/19 Layers
- All filters 3x3
- Generally similar to AlexNet (Pooling interspersed)

https://medium.com/deep-learning-g/cnn-architectures-vggnet-e09d7fe79c45

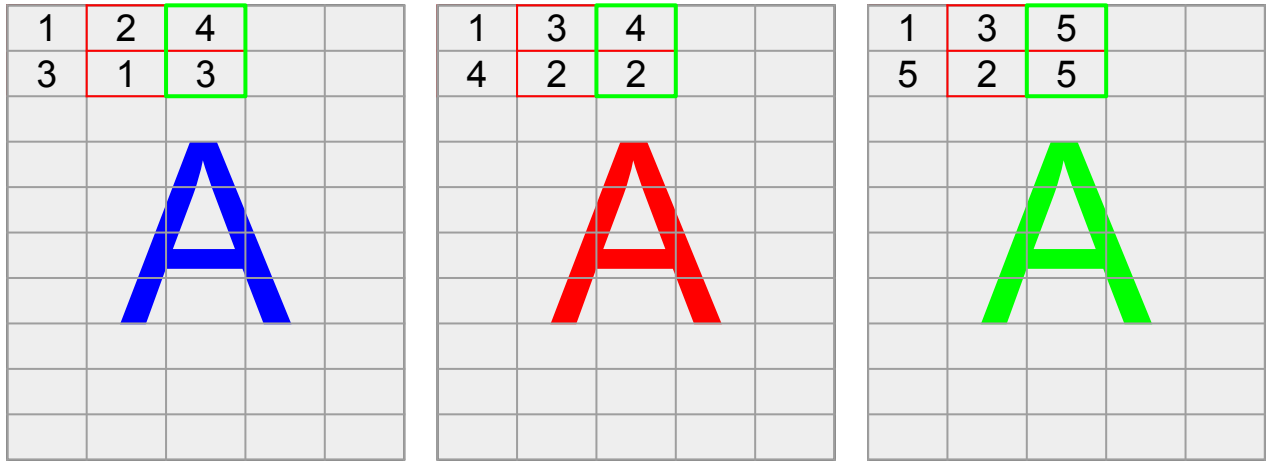
Why is a filter of 3x3 + more depth > 11x11 with less depth?



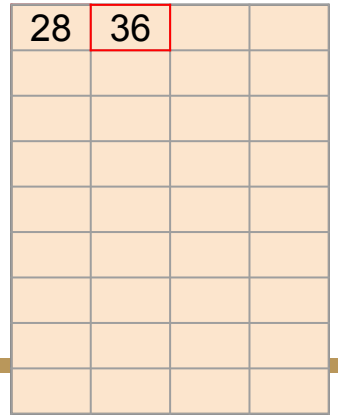
Total Unique Pixels:
12



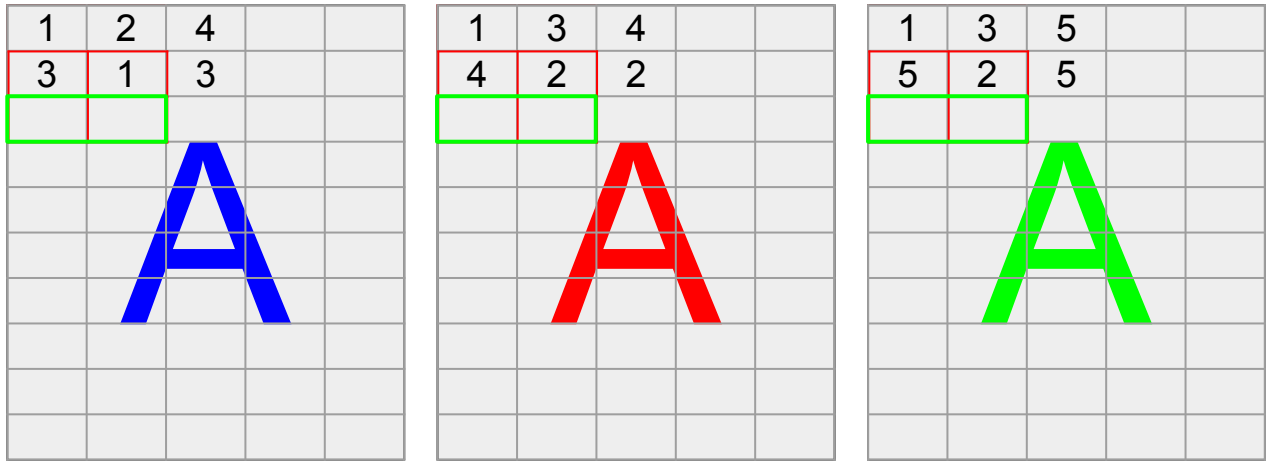
Why is a filter of 3x3 + more depth > 11x11 with less depth?



Total Unique Pixels:
18

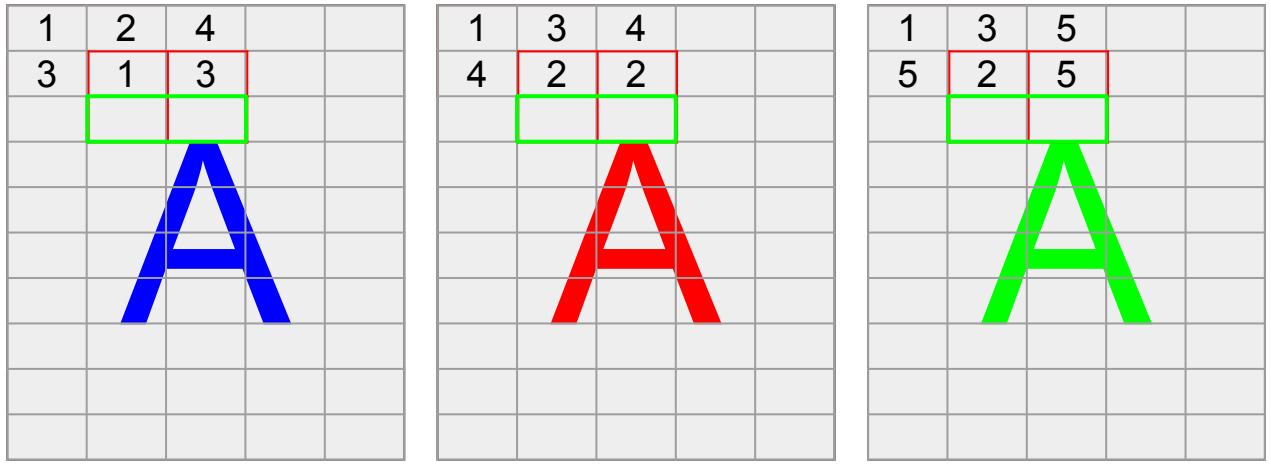


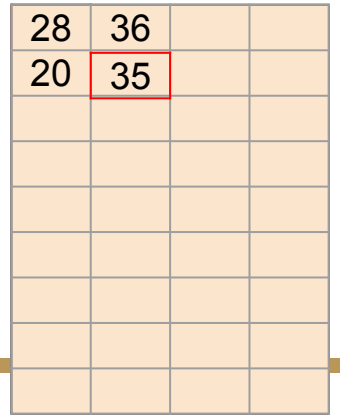
Why is a filter of 3x3 + more depth > 11x11 with less depth?



Total Unique Pixels:
24

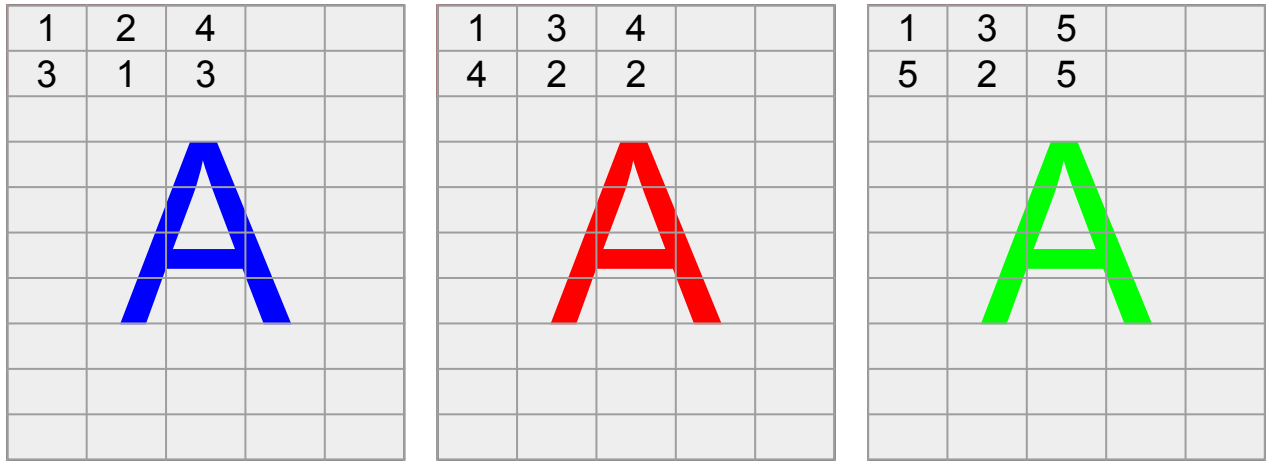
Why is a filter of 3x3 + more depth > 11x11 with less depth?

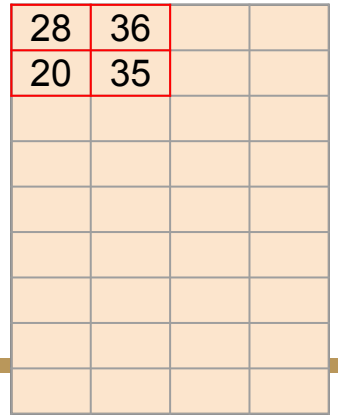




Total Unique Pixels:
27

Why is a filter of 3x3 + more depth > 11x11 with less depth?



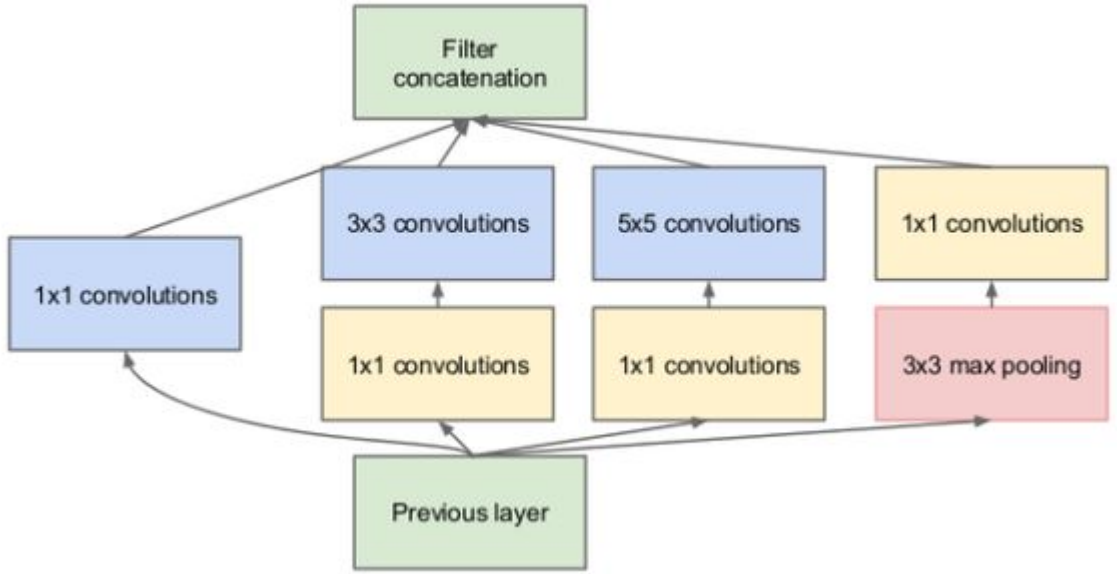


Total Unique Pixels:
27

GoogleNet (aka Inception v1)

Focused on Computational Efficiency; similar accuracy (better in some cases) to VGG.

5 million total parameters (vs >60 million for AlexNet; 138 million for VGG16).

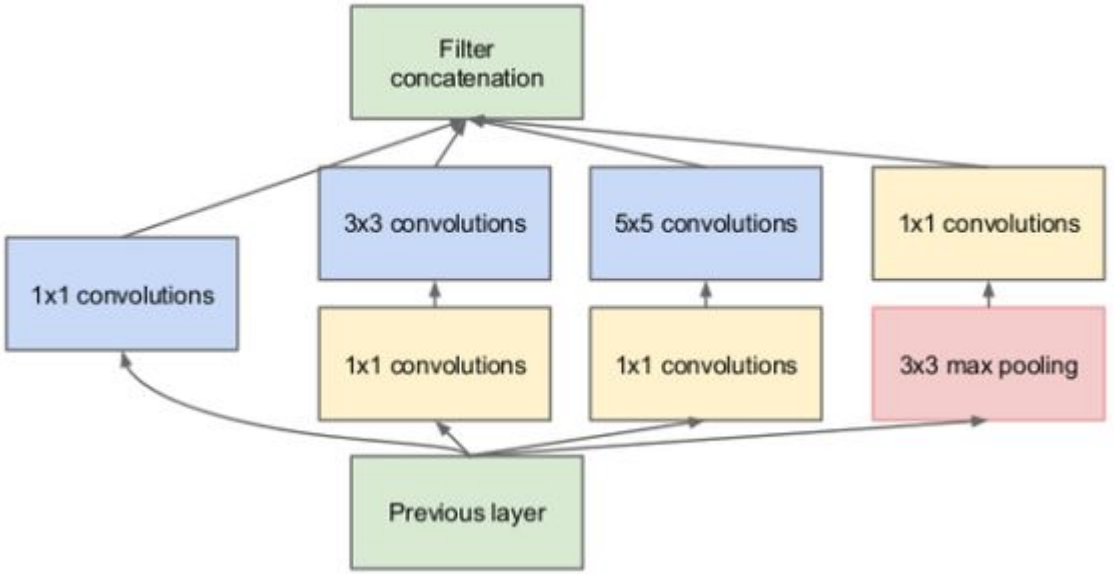


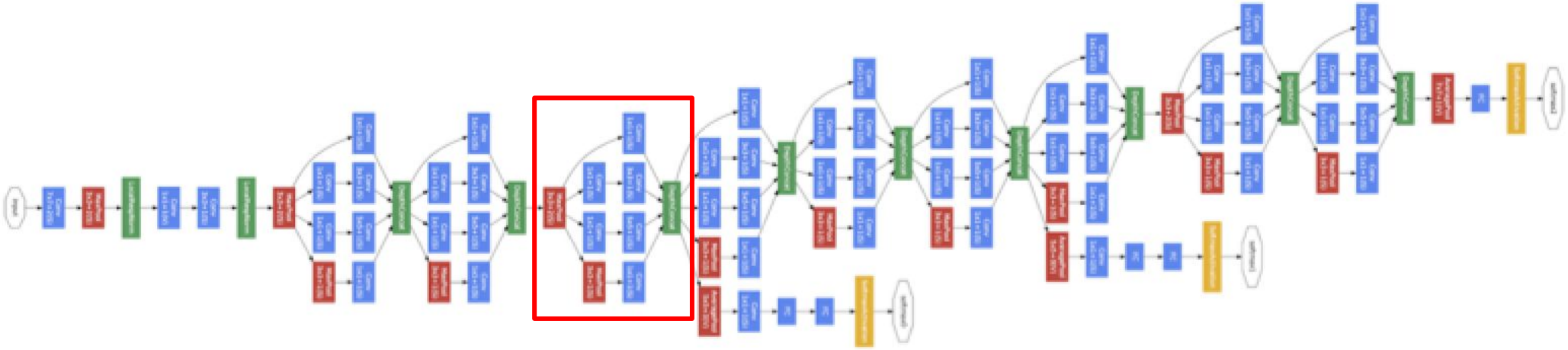
Inception Modules.

GoogleNet/Inception v1

Basic Goal: Build a very performant “micro network” (Inception Layer), and then build a giant stack of them.

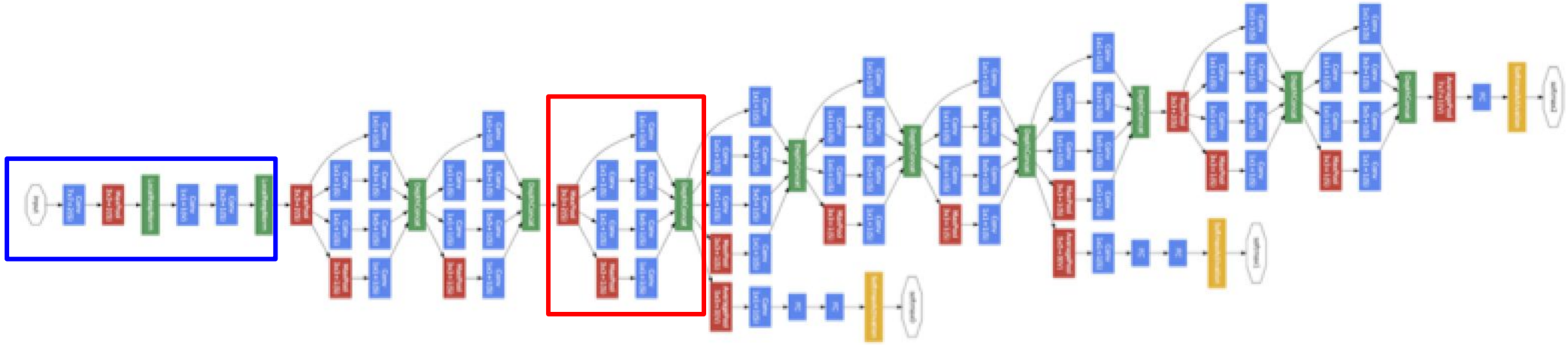
All outputs are concatenated together, and passed into the next inception layer.





Inception Module in GoogleNet Architecture

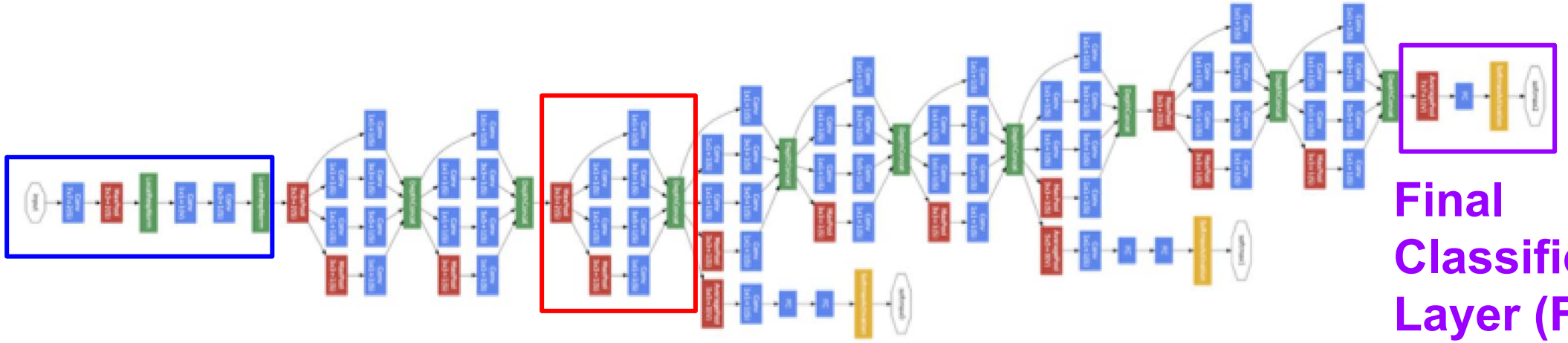
Convolution
Pooling
Softmax
Other



Inception Module in GoogleNet Architecture

At the start of the network is a Stem, which contains a traditional set of Convolutional and Pooling layers.

Convolution
Pooling
Softmax
Other

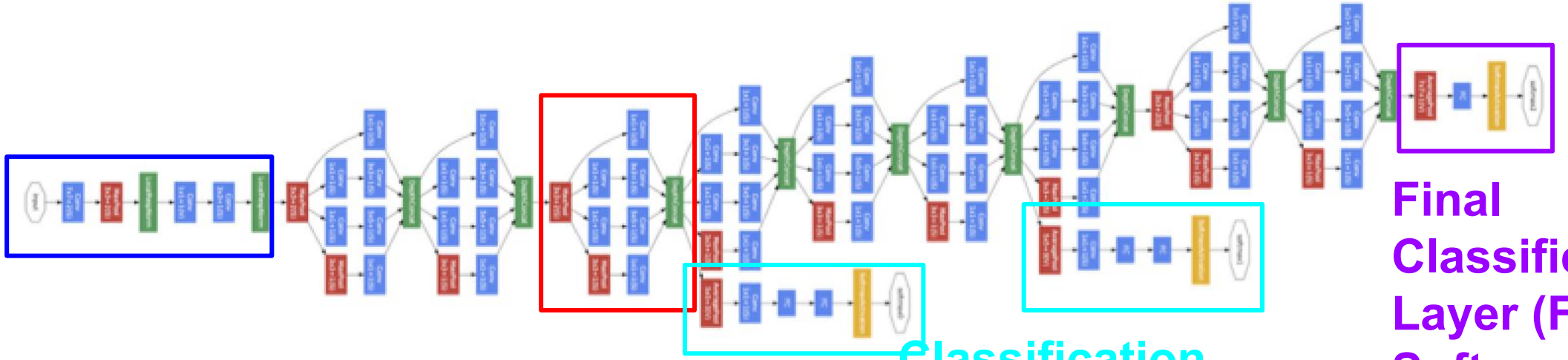


Final Classification Layer (FC -> Softmax)

Convolution
Pooling
Softmax
Other

Inception Module in GoogleNet Architecture

At the start of the network is a Stem, which contains a traditional set of Convolutional and Pooling layers.



Final Classification Layer (FC -> Softmax)

Classification Outputs which add gradient to earlier layers.

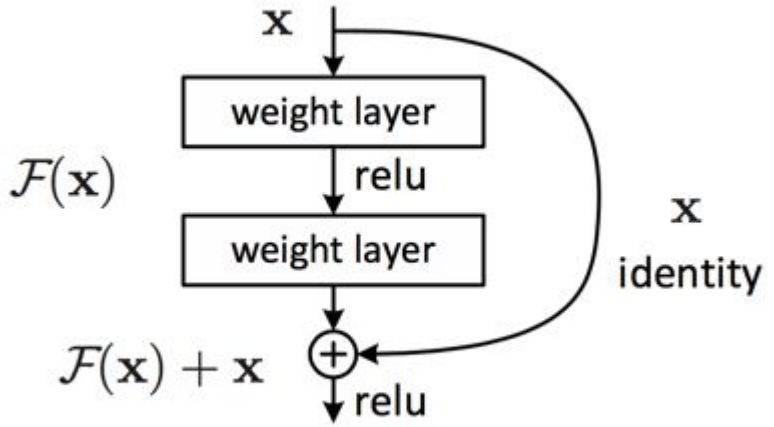
Convolution
Pooling
Softmax
Other

Inception Module in GoogleNet Architecture

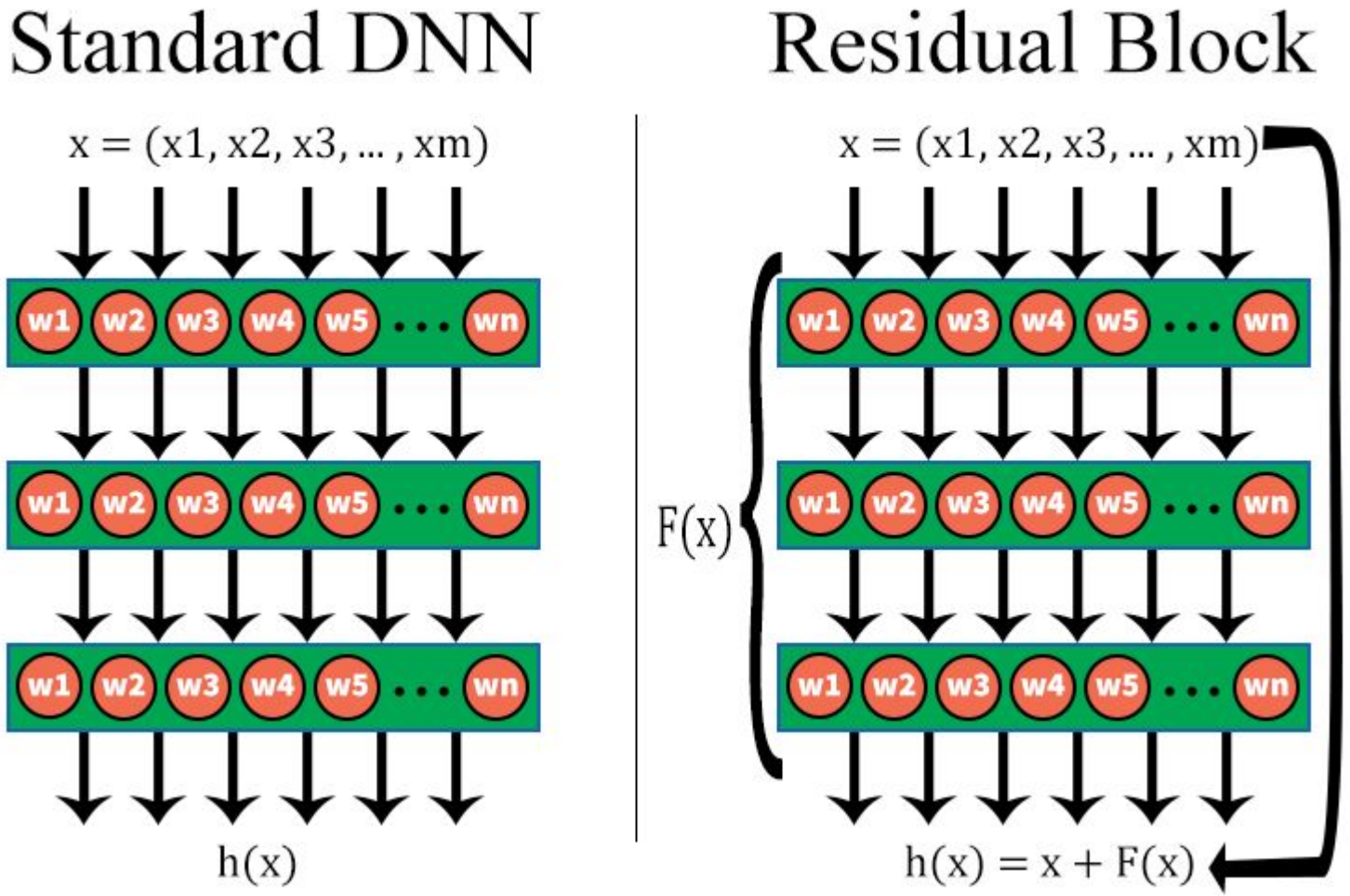
At the start of the network is a Stem, which contains a traditional set of Convolutional and Pooling layers.

The Elephant in the Room: ResNet (2015)

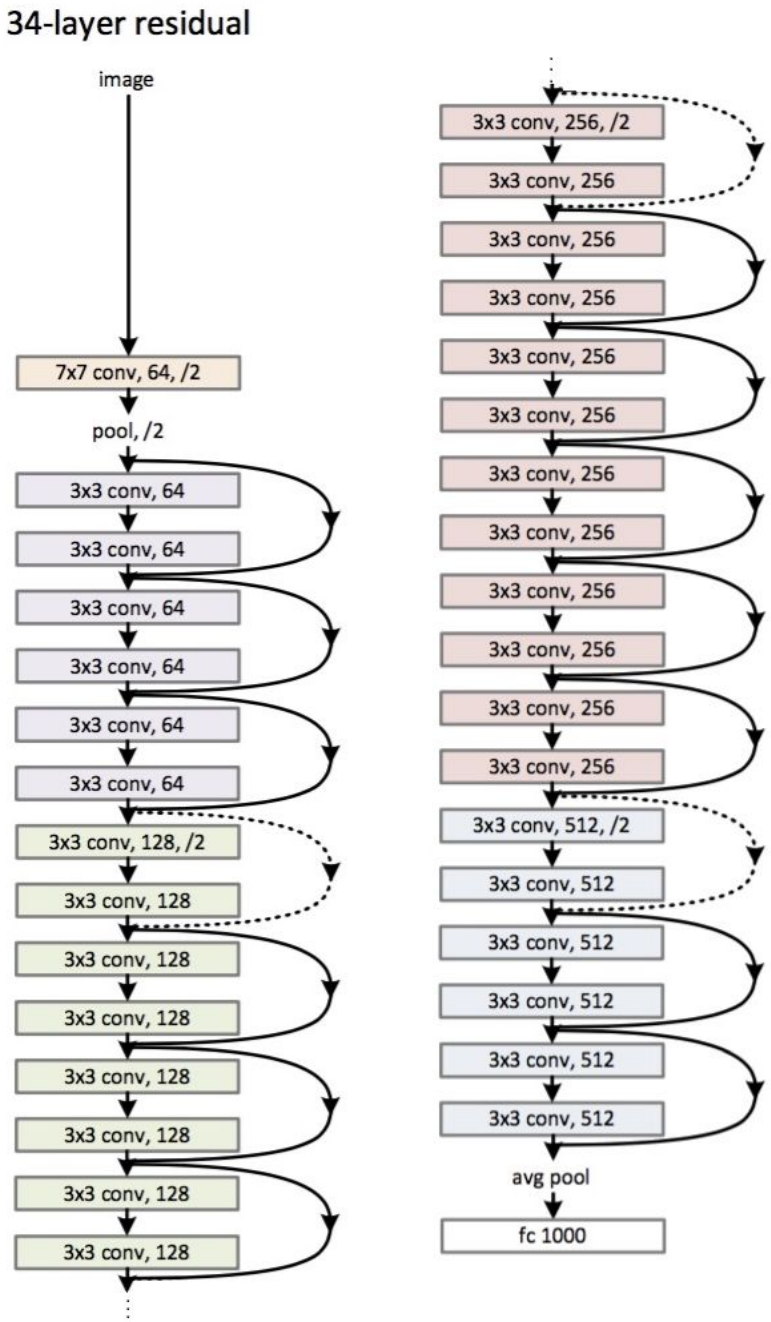
- Huge increase in accuracy (down to 3.57% error on ImageNet)
- Huge increase in depth - 152 layers!



ResNet



https://wiki.math.uwaterloo.ca/statwiki/index.php?title=File:Residual-Block-vs-DNN.png



- FC1000 is to translate the final convolutional layer out to the number of classes you're targeting.

Summary

- Broad overview of hardware (CPU vs. GPU)
- Discussion of current frameworks for deep learning, pros/cons
- Discussion of popular architectures for image recognition
  - VGG
  - ResNet
  - GoogleNet / Inception