CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019

Outline for April 12

- Proposal and Project highlights
- Finish Convolutional Neural Networks
  - Weights on CONV layers
  - Parameter analysis
  - Strides and pooling
- Other NN architectures
- Next week: unsupervised learning

- Lab 7 due MONDAY
- Final project details posted
- Office hours TODAY 1-3pm

Outline for April 12

- Proposal and Project highlights
- Finish Convolutional Neural Networks
  - Weights on CONV layers
  - Parameter analysis
  - Strides and pooling
- Other NN architectures
- Next week: unsupervised learning

Project Proposal

1) Dataset
2) Methods/Algorithms
3) Scientific Question
4) Evaluation and Interpretation of Results
5) References

Datasets

- Kaggle: Wide variety of datasets (may need to create an account).
- UCI Machine Learning Repository: Also contains a wide variety of datasets (options on the left allow you to search by task, attribute type, etc which can be very useful).
- ImageNET: Large database of images (larger than CIFAR-10, which is also an option).
- **Climate data**: The faculty at Swarthmore have been encouraged to incorporate climate analysis into our curriculum. If you are interested, I would encourage you to choose a climate-oriented project. There are a number of government and climate research sites with data. Here are a few:
  - Climate.gov
  - GlobalChange.gov
  - NOAA
  - EPA
- Wikipedia Data List: Up-to-date list of datasets organized by category (may or may not be freely available):
  - Image data
  - Text data
  - Sound data
  - Signal data
  - Physical data
  - Biological data
  - Anomaly data
  - Question answering data
  - Multivariate data
- 1000 genomes (human DNA data): If you're interested in exploring a biological project, let me know. This specific dataset contains DNA data from humans around the world, but there are many other datasets from other species.
- The 50 Best Free Datasets for Machine Learning: There are a number of these type of lists floating around - this one looks decent but make sure that you can actually download the data.

Project Lab Notebook

- As you as you receive your git repo, start creating a “lab notebook” in your README
- This should say who was working, what date, how long, and briefly what you did

Sara: 03-07-18 (2hrs)

- now averaging the Markov chain, fixed all the results
- combined ancestral 1000 genomes still running (need to start similar for SGDP)
- started new runs with filtering to only have selected alleles in the “selected pop” and only have ancestral alleles in the “reference panel”

Project Deliverables

- Main deliverable: presentation

  - Group of 1: 5 min
  - Group of 2: 9 min
  - Group of 3: 12 min

- On git:
  - Lab Notebook
  - Project Code
  - Presentation Slides

Outline for April 12

- Proposal and Project highlights
- Finish Convolutional Neural Networks
  - Weights on CONV layers
  - Parameter analysis
  - Strides and pooling
- Other NN architectures
- Next week: unsupervised learning

Visualization of an entire network

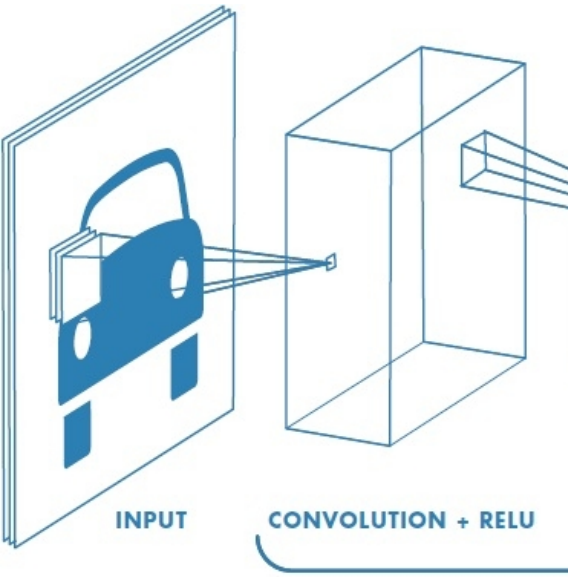


Image from MathWorks: https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html

Visualization of an entire network

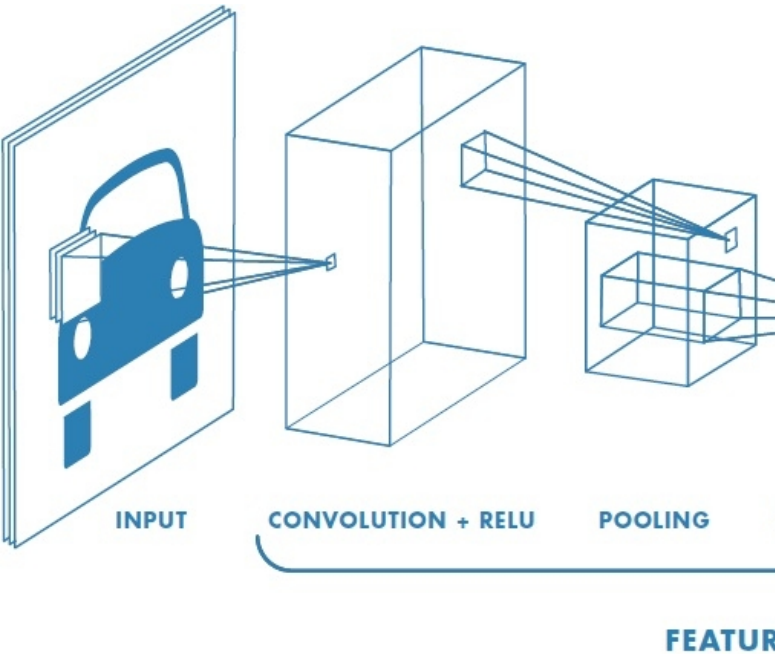


Image from MathWorks: https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html

Visualization of an entire network

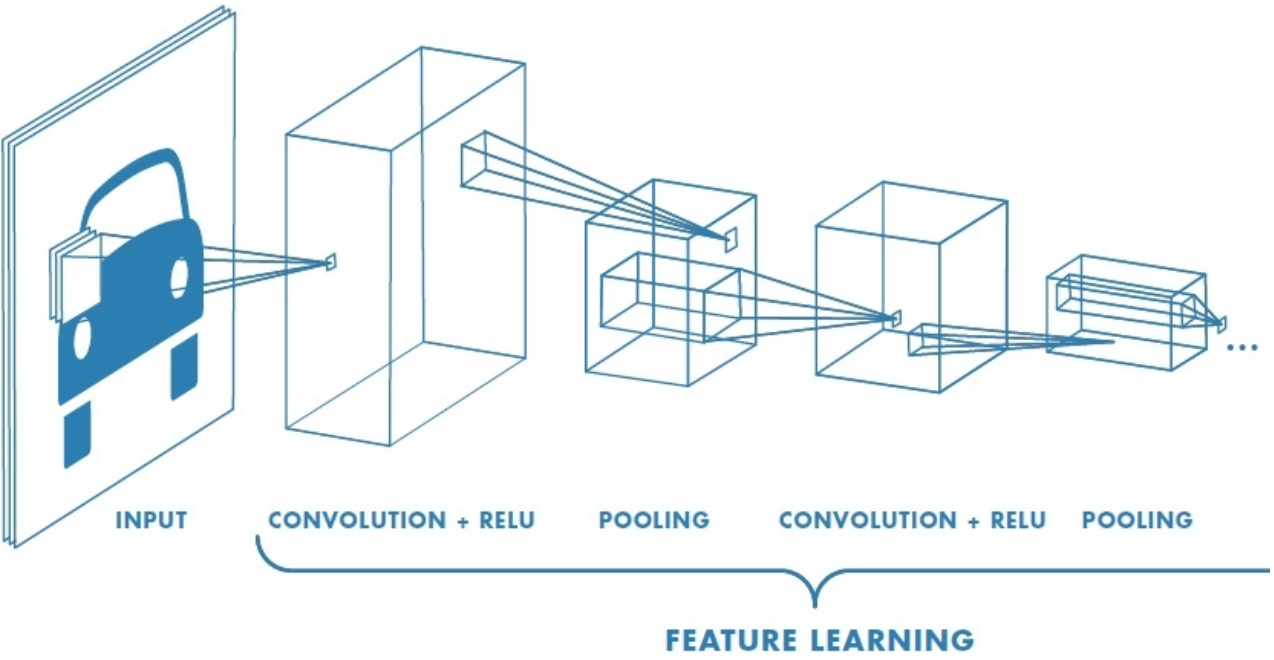


Image from MathWorks: https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html

Visualization of an entire network

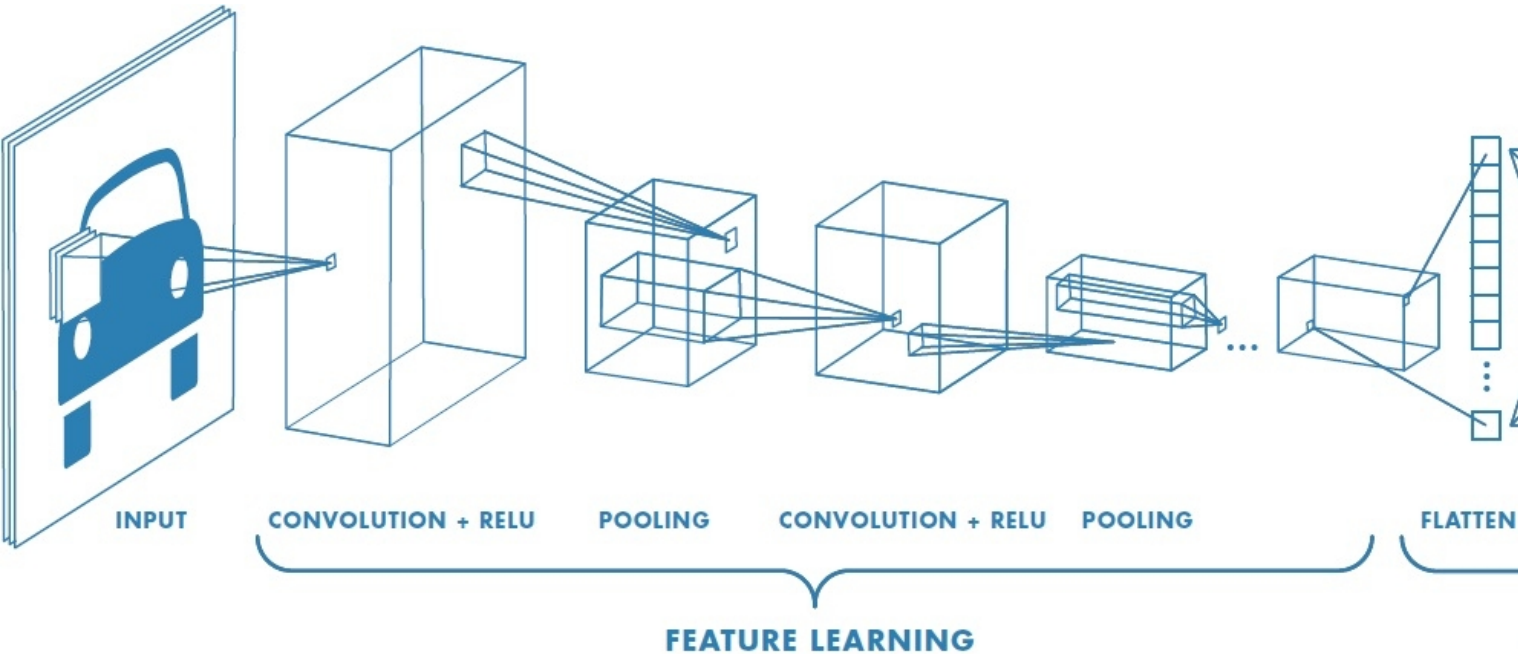


Image from MathWorks: https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html

Visualization of an entire network

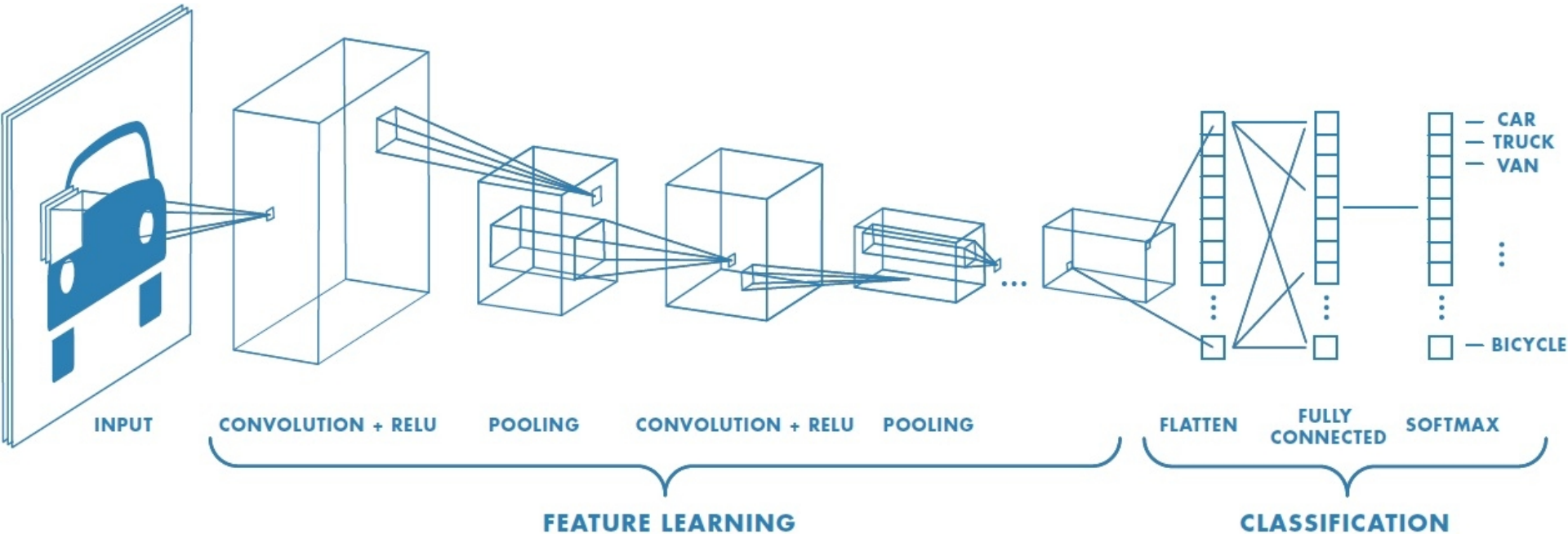
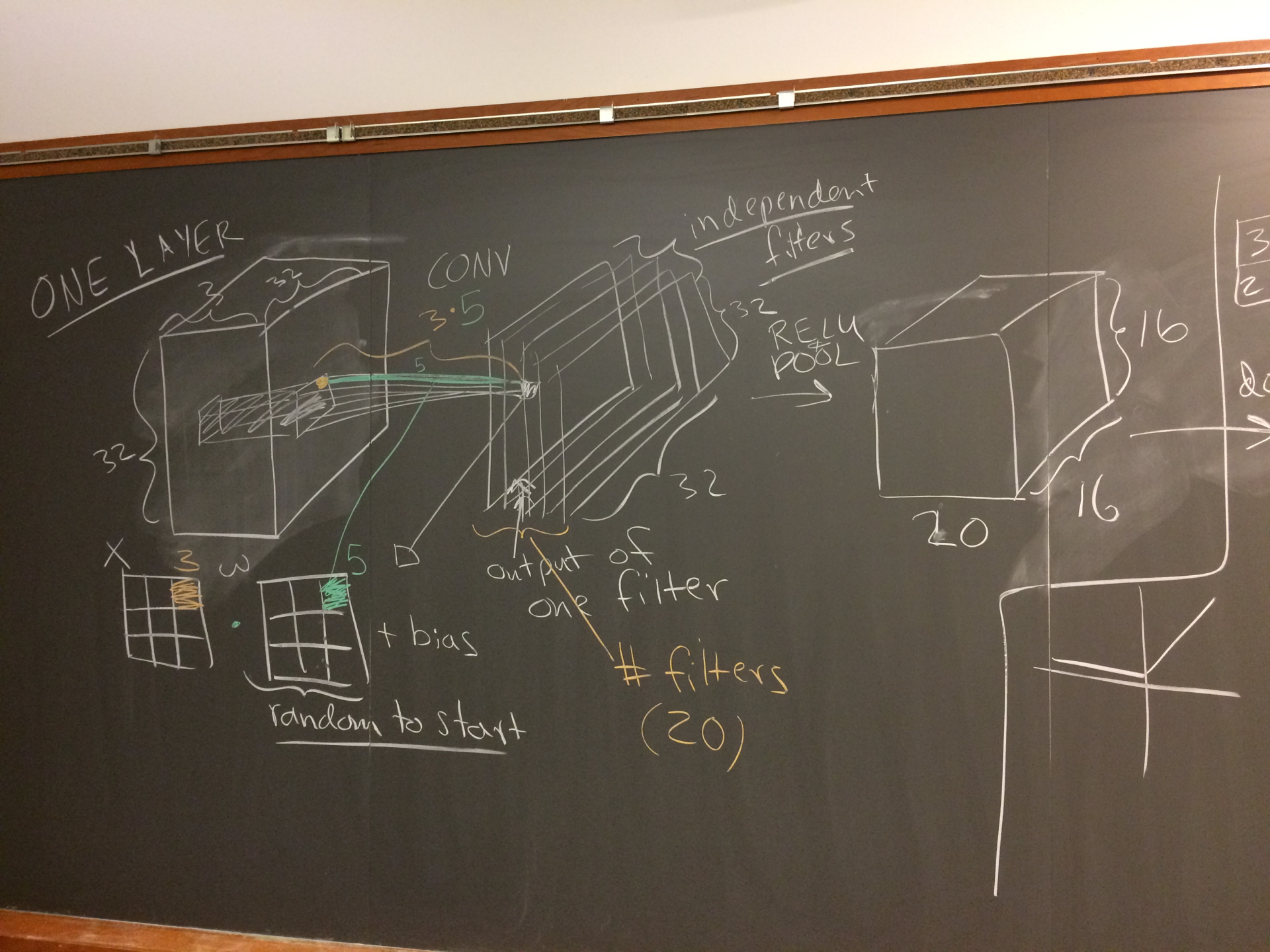
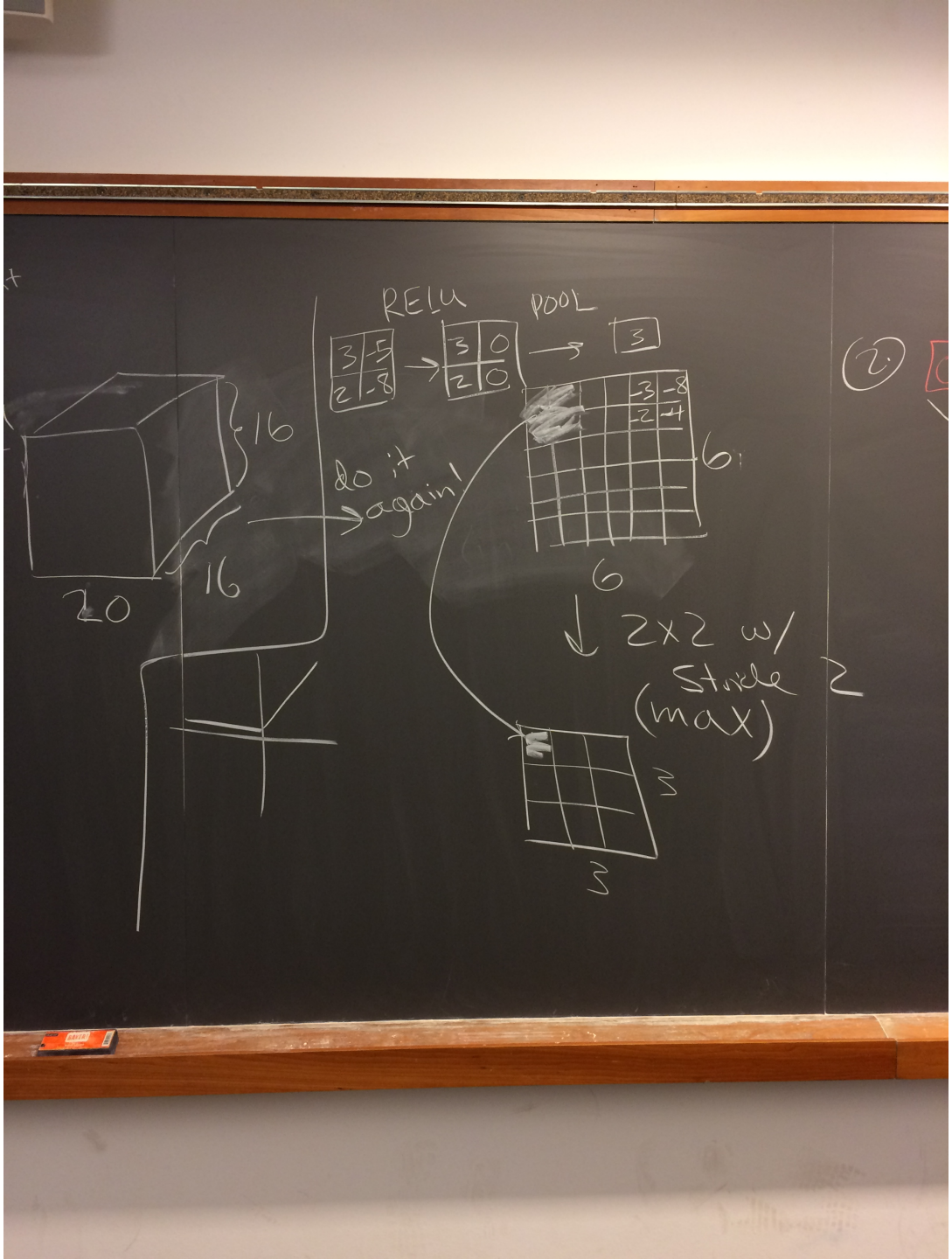


Image from MathWorks: https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html


ONE LAYER
3
32
32
CONV
3·5
5
independent filters
32
RELU & POOL
16
16
20
32
X
3
3
5
+ bias
output of one filter
filters (20)
random to start


RELU
POOL
3
16
16
20
do it again!
6
6
2x2 w/ Stride 2 (max)
3
3

Handout 16

(a) params:

CONV, FC

no params:

RELU, POOL, FLATTEN

(b)

5x

weights for one filter

(c)

3x

Note that some of these operations require hyper-parameters!

- CONV: filter size, number of filters, stride, padding
- POOL: filter size, stride
- FC: if using a hidden layer, number of units in this layer

(b) $5 \times 5 \times 3 \times 20 + 20$

from input

weights for one filter

one 3D filter

filters

new depth!

1520

(c) $3 \times 3 \times 20 \times 10 + 10$

1810

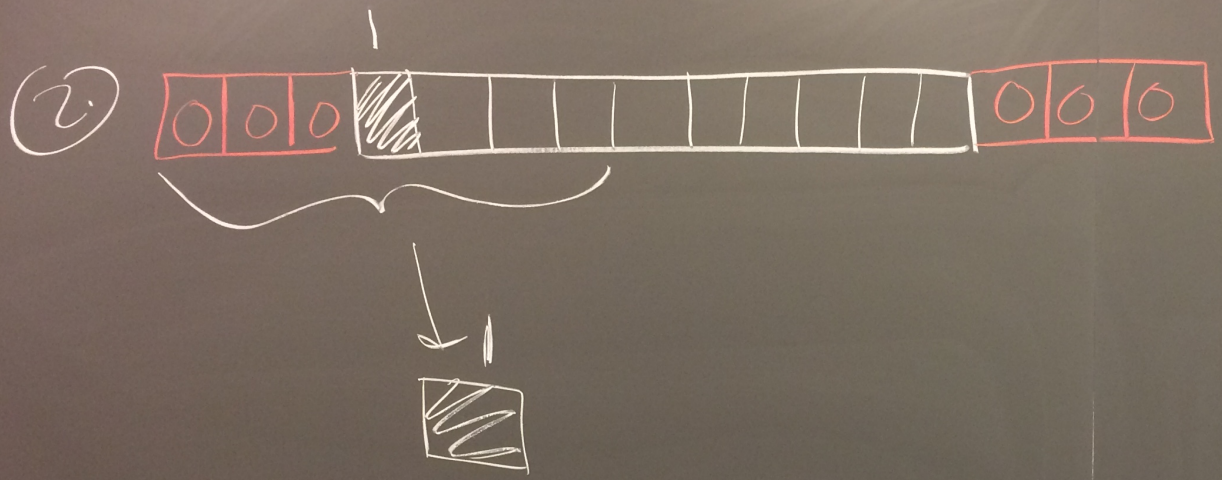
(d) $8 \cdot 8 \cdot 10 \times 10 + 10$

6410

(e) 9740

(f)

☆ much better!



Outline for April 12

- Proposal and Project highlights
- Finish Convolutional Neural Networks
  - Weights on CONV layers
  - Parameter analysis
  - Strides and pooling
- Other NN architectures NEXT TIME!
- Next week: unsupervised learning