Machine Learning & Image Classification

ROB 102: Introduction to AI & Programming

Lecture 10

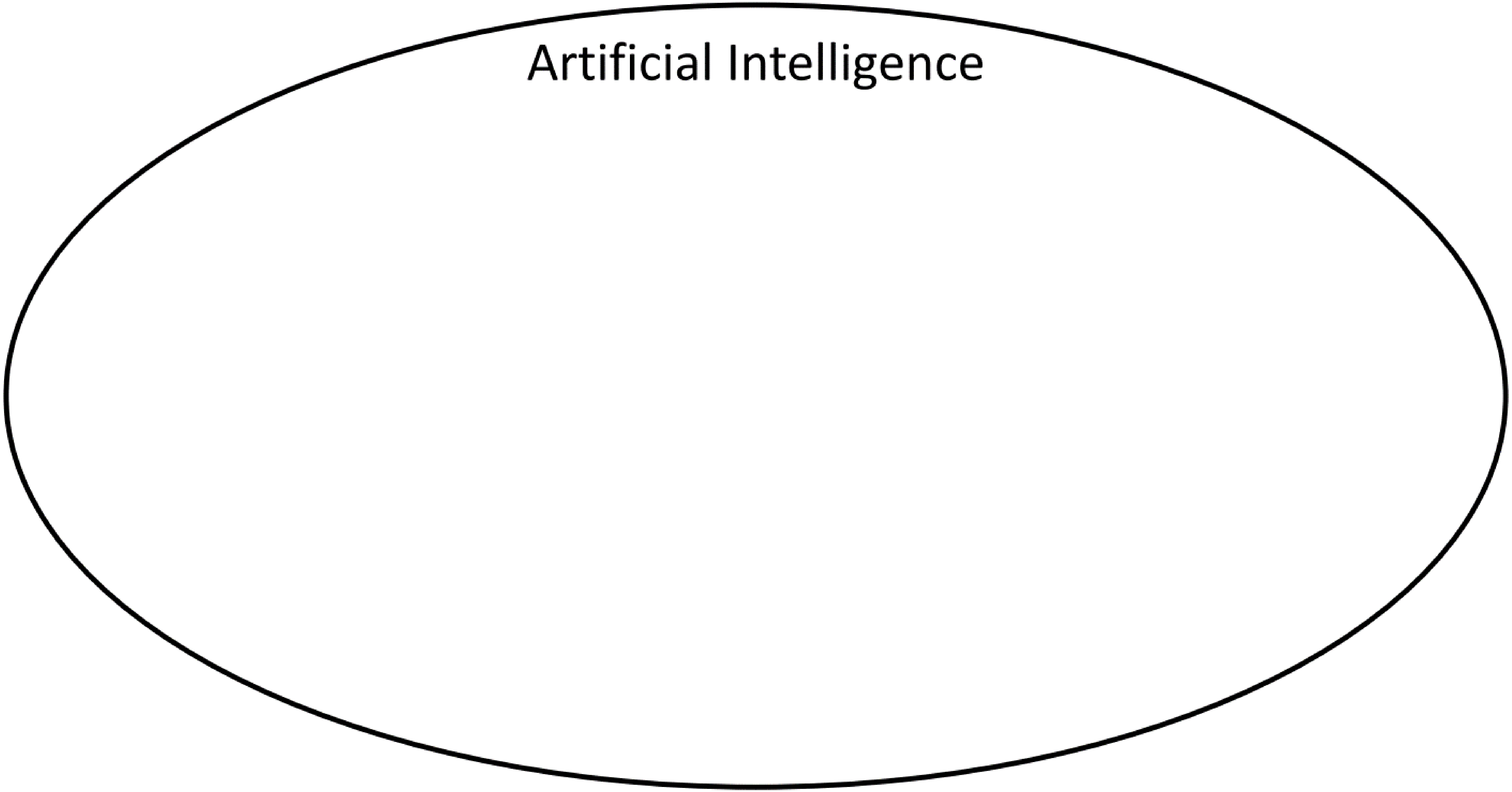
2021/11/17

How can we make a robot see?

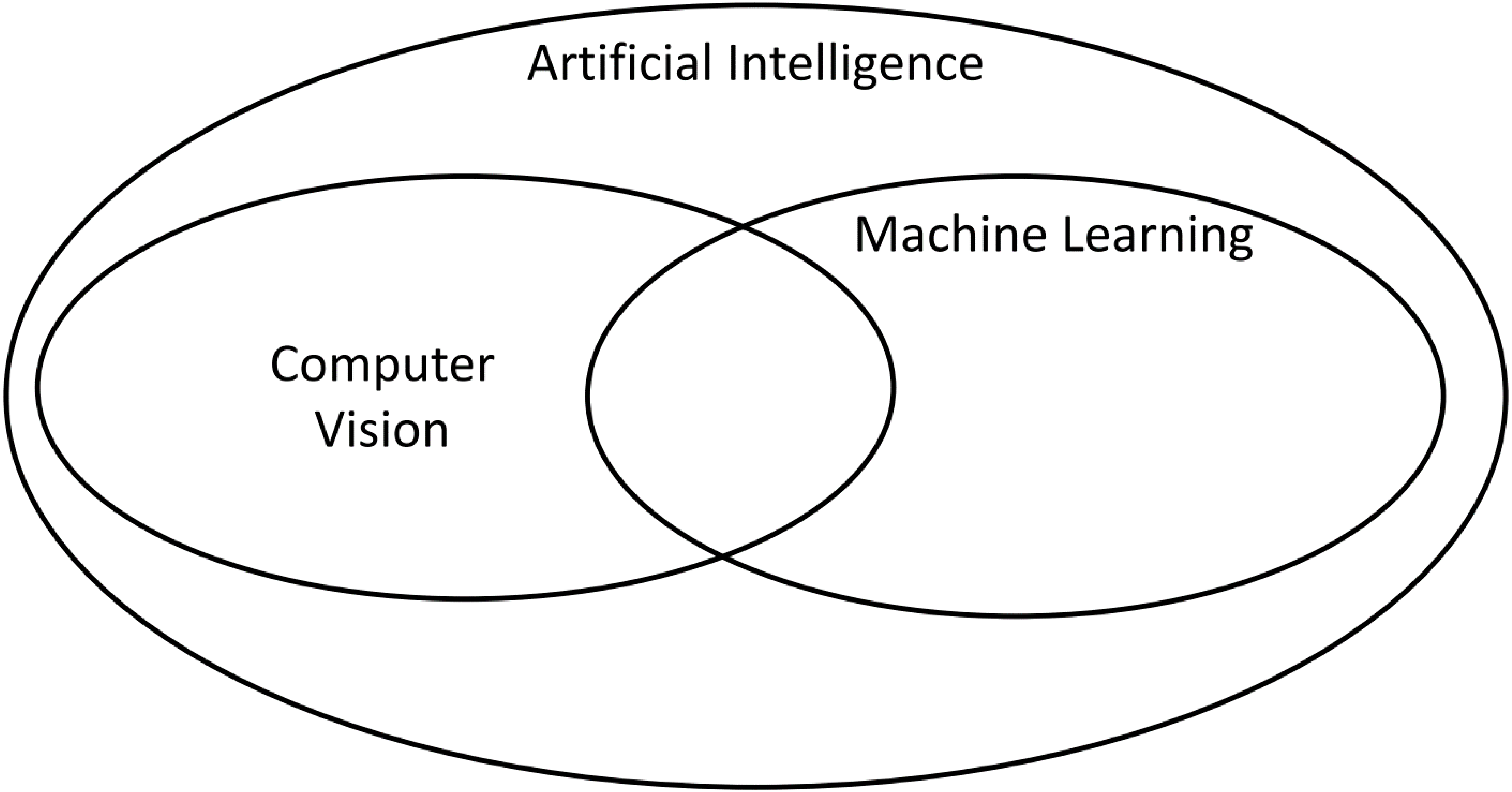
How can we make a robot see?

camera!

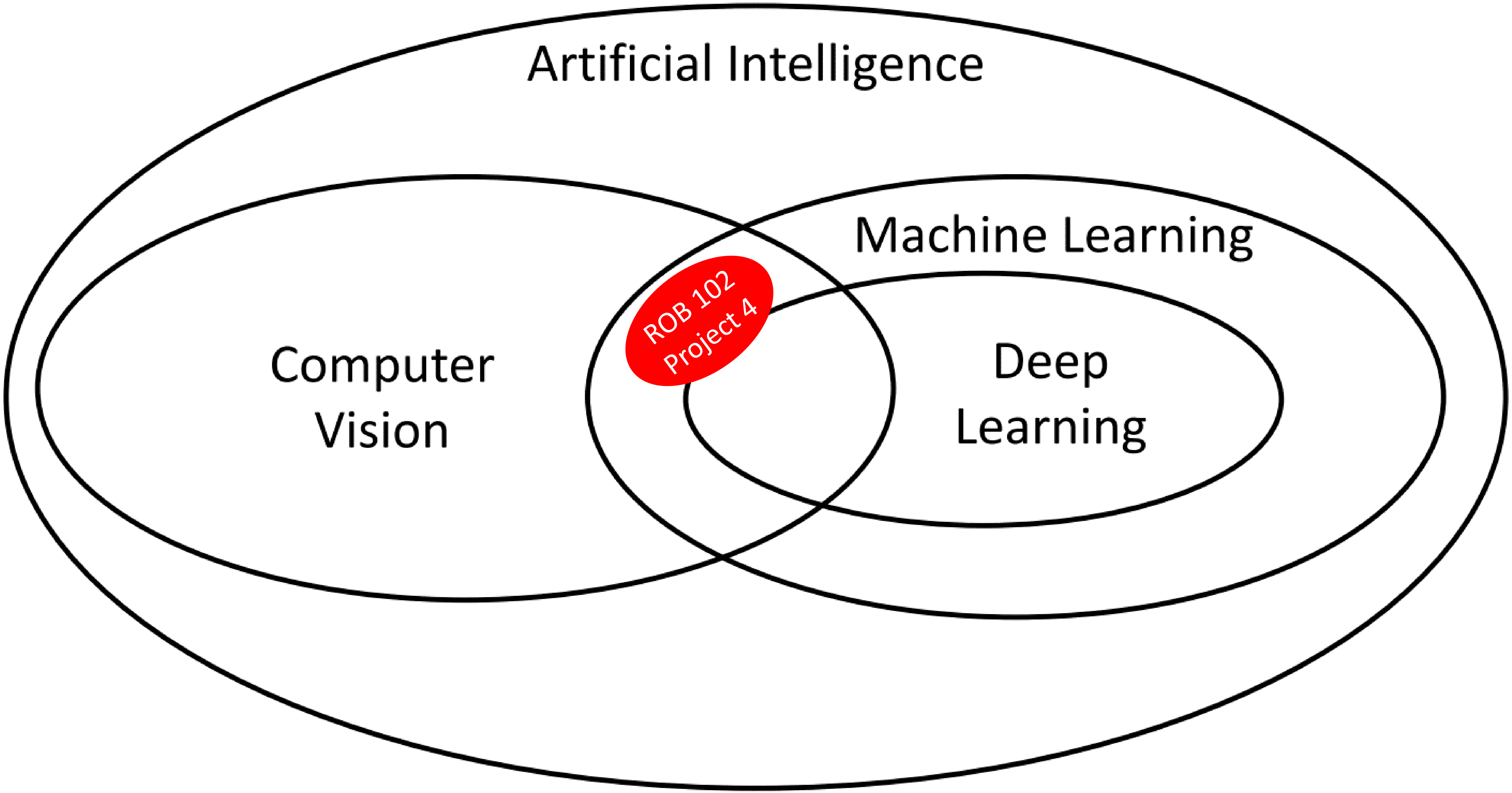
ROB 102 (Project 4): We will perform a **computer vision** task using **machine learning** algorithms.



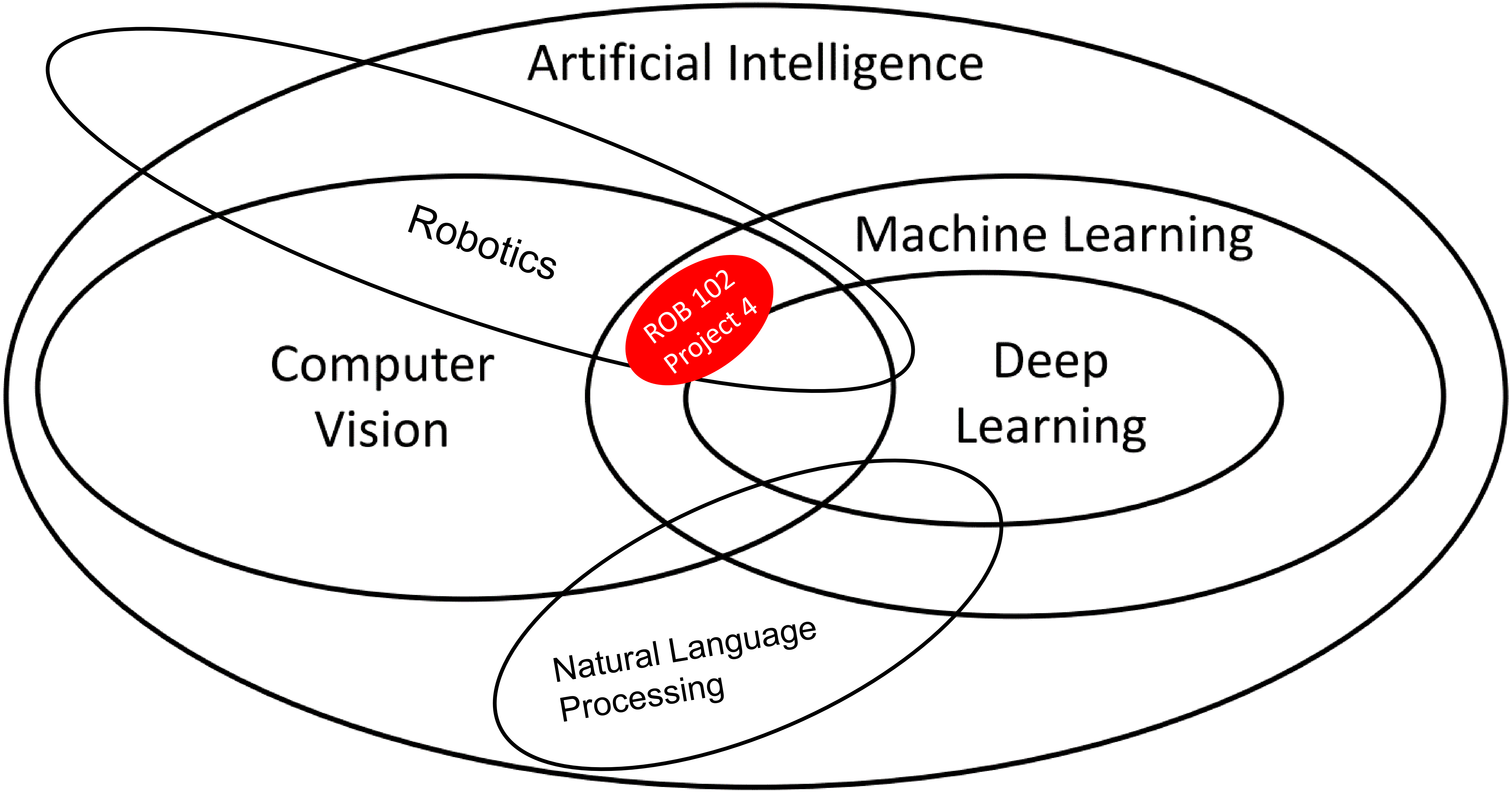
Credit: Justin Johnson, EECS 498 F20 (link)



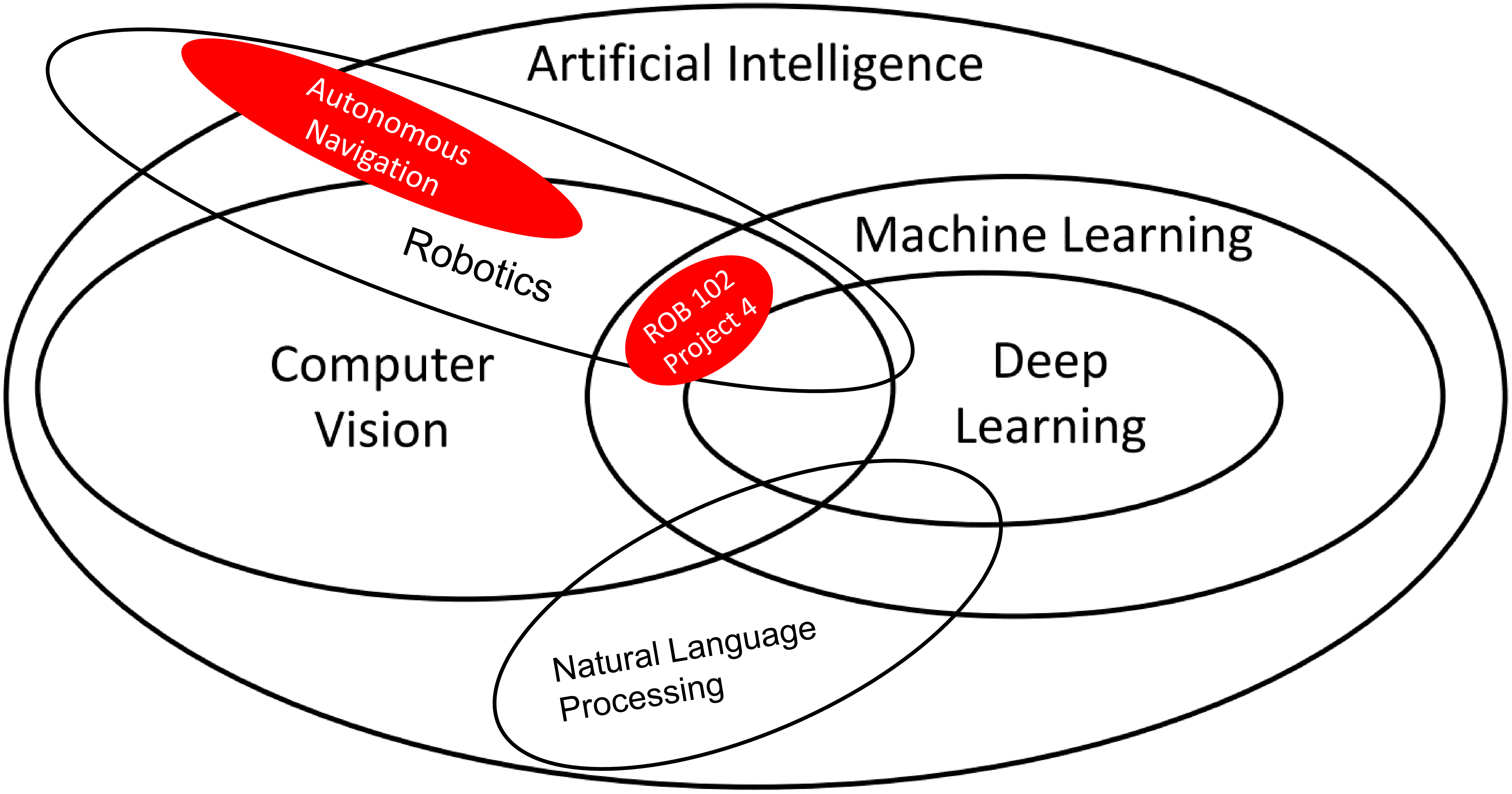
Credit: Justin Johnson, EECS 498 F20 (link)



Credit: Justin Johnson, EECS 498 F20 (link)



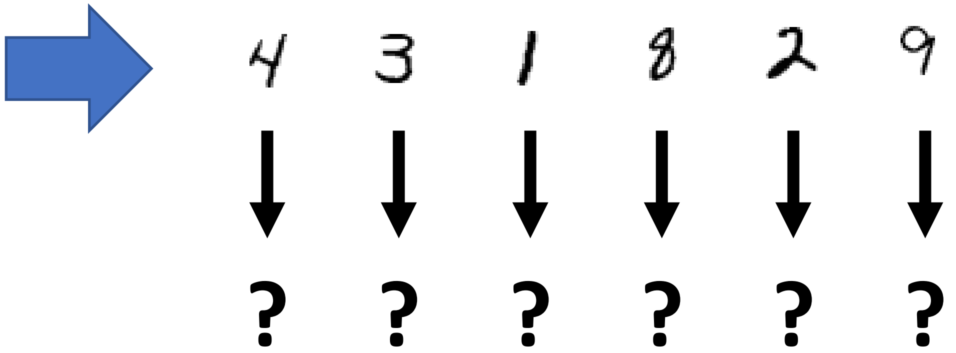
Credit: Justin Johnson, EECS 498 F20 (link)



Credit: Justin Johnson, EECS 498 F20 (link)

Goal for ROB 102 Project 4: Perform **image classification** with machine learning algorithms.

Image: CC4.0 (link)



Machine Learning Terminology

Supervised Learning

Unsupervised Learning

Reinforcement Learning

Supervised Learning

Given input data X and labels for that data y, learn a function to perform prediction on new data:

$$f(X) = y.$$

Supervised Learning

IMAGENET Large Scale Visual Recognition Challenge

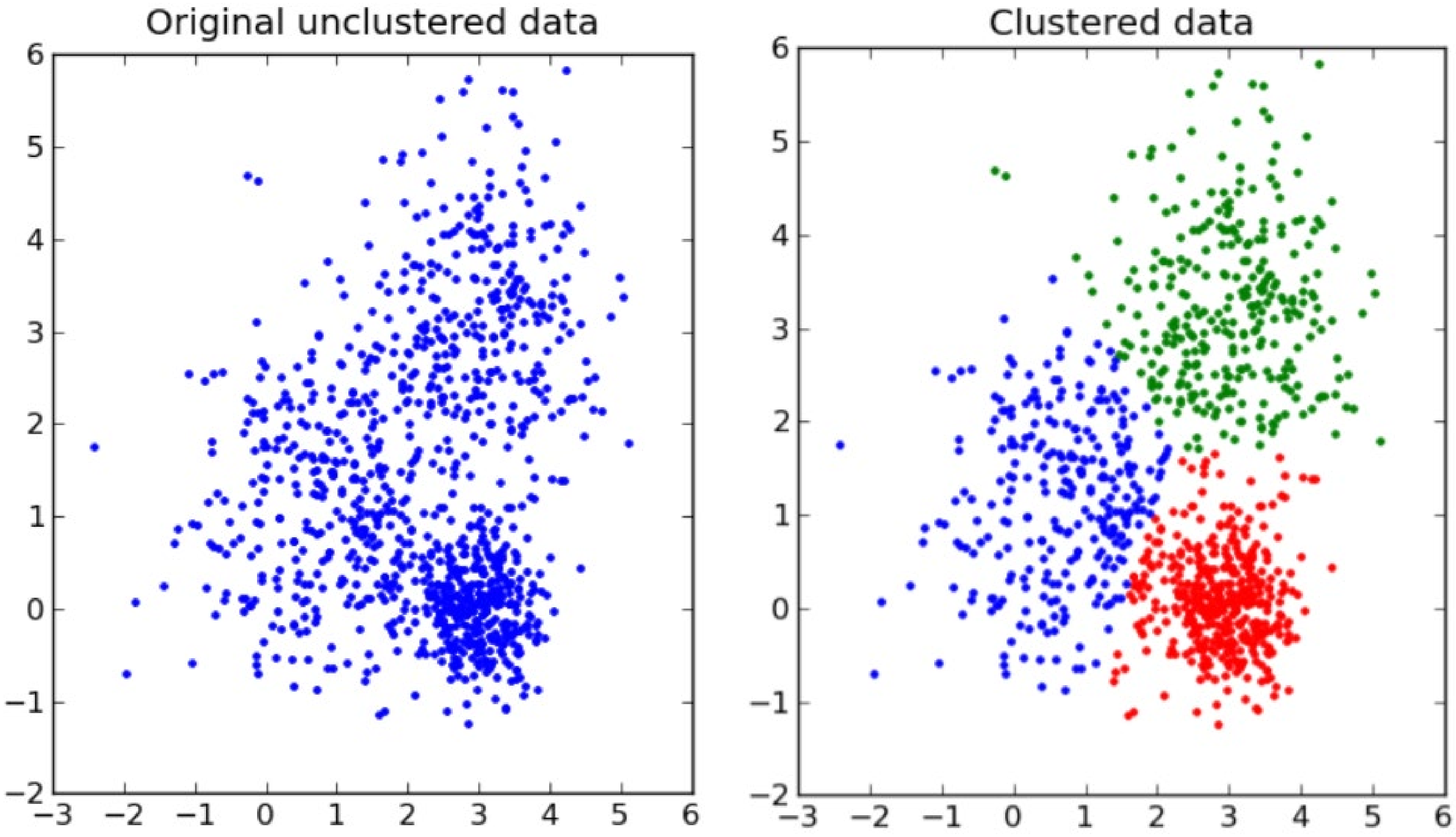
The Image Classification Challenge:
1,000 object classes
1,431,167 images

Output:
Scale
T-shirt
Steel drum
Drumstick
Mud turtle

Deng et al, 2009
Russakovsky et al. IJCV 2015

Unsupervised Learning

Given unlabeled data, learn a function over the input.



Source: Pineau, 2017 (link)

Reinforcement Learning

An agent learns a policy that maximizes a reward.

DeepMind (link)

Beat the world Go champion in 2017.

Reinforcement Learning

Outperforming humans at Atari games

Agent57: Outperforming the human Atari benchmark. [Puigdomènech, 2020]

Machine Learning Terminology

Supervised Learning

Image Classification

Unsupervised Learning

Reinforcement Learning

Machine Learning Terminology

Supervised Learning

Image Classification

ROB 102

Unsupervised Learning

Reinforcement Learning

What is in this picture?

Image: Alastair Pollock Photography/Getty Images

"shark"

What is in this picture?

Image: Alastair Pollock Photography/Getty Images

“shark”

Image: Benjamin Lowy/The New York Times

“shark”

What is in this picture?


Image: Alastair Pollock Photography/Getty Images

“shark”


Image: Benjamin Lowy/The New York Times

“shark”


Image: WeRateDogs

What is in this picture?

Image: Alastair Pollock Photography/Getty Images

“shark”

Image: Benjamin Lowy/The New York Times

“shark”

Image: WeRateDogs

~~“shark”~~

“dog”

Solve the following equation:

$$10230823.5849 \times 3729.2201 + 19420186 = \ ?$$

Humans are very good at reasoning about images:

“shark”

“shark”

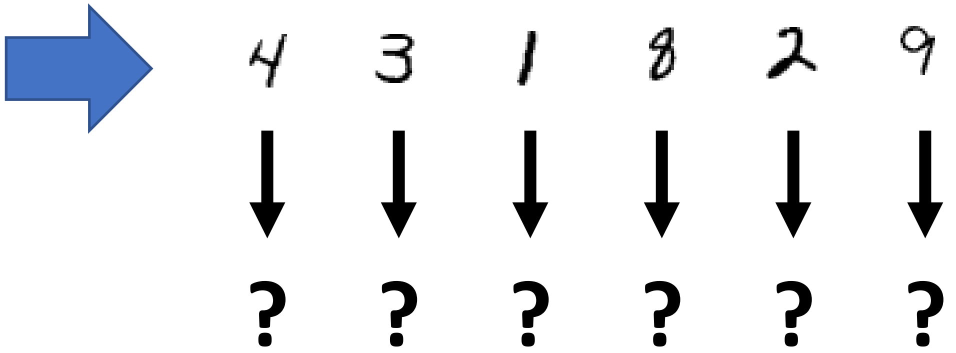
~~“shark”~~
“dog”

Computers are very good at doing arithmetic:

$$10230823.5849 \times 3729.2201 + 19420186 = \ ?$$

Goal for ROB 102 Project 4: Perform **<u>image classification</u>** with machine learning algorithms.

Image: CC4.0 (link)



Goal for ROB 102 Project 4: Perform <u>**image classification**</u> with machine learning algorithms.

Image: CC4.0 (link)

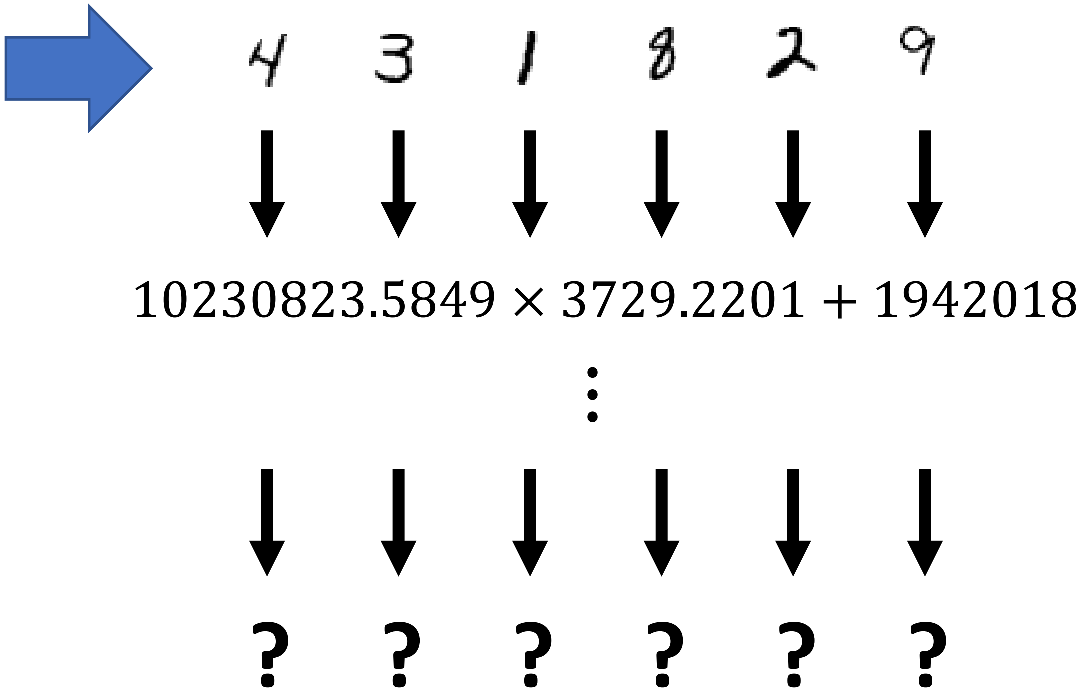
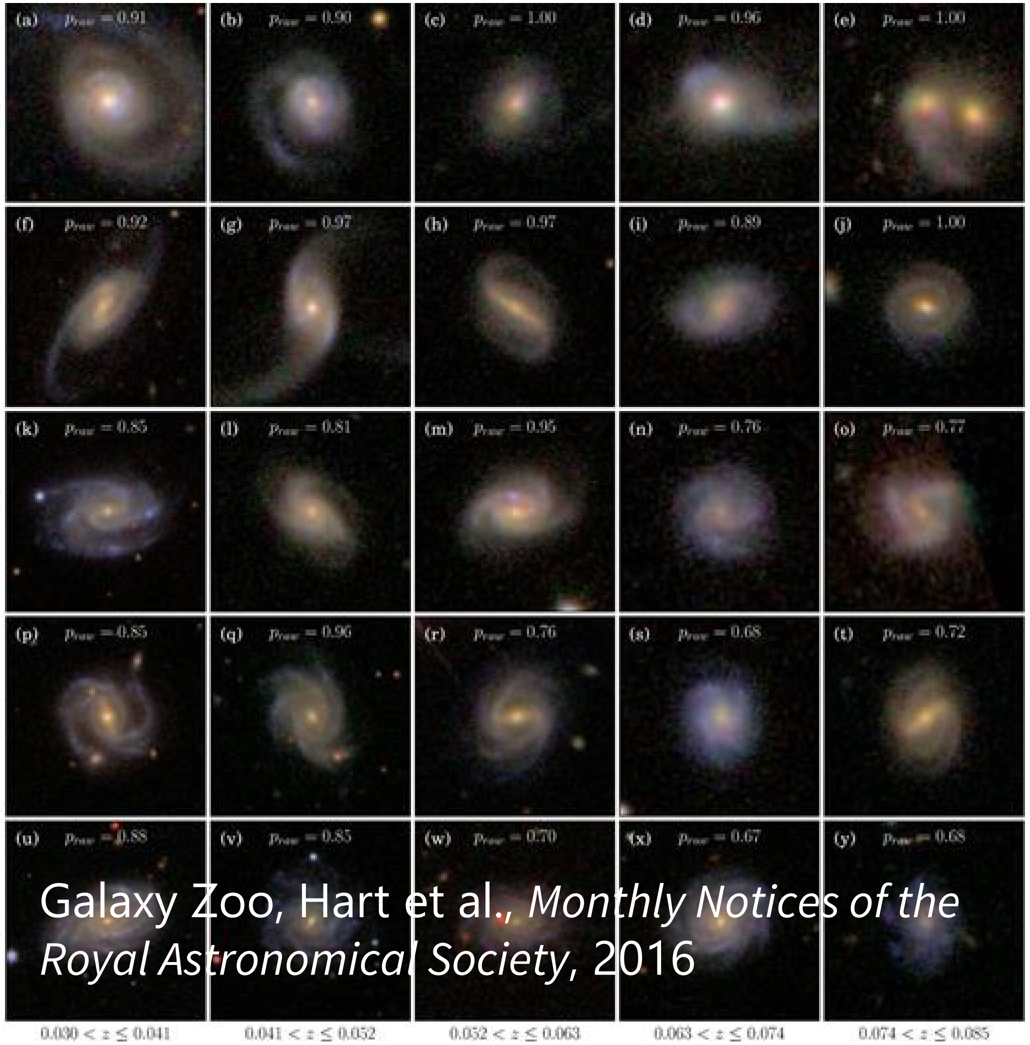


Image Classification



Galaxy Zoo, Hart et al., *Monthly Notices of the Royal Astronomical Society*, 2016

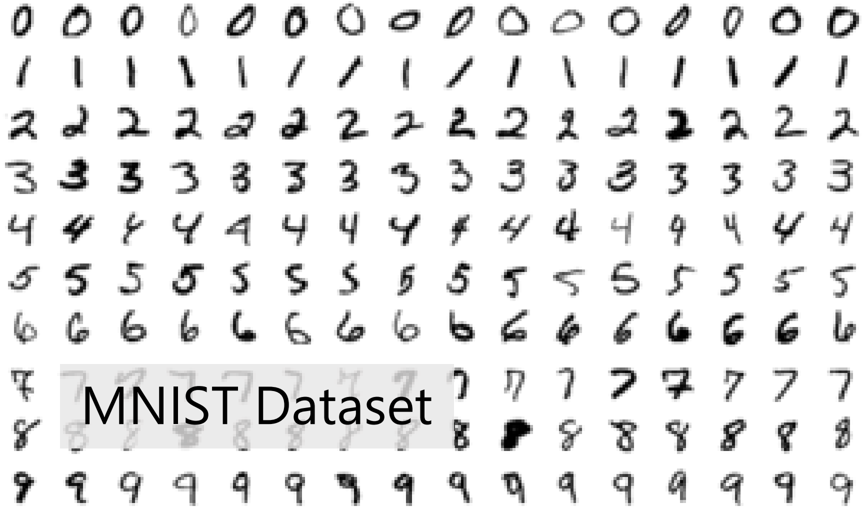
Galaxy Classification

Skin Cancer Detection

Esteva et al., Nature, 2017

Image: Matt Young (link)

Handwriting classification



MNIST Dataset

Image: CC4.0 (link)

Image Classification: A Building Block

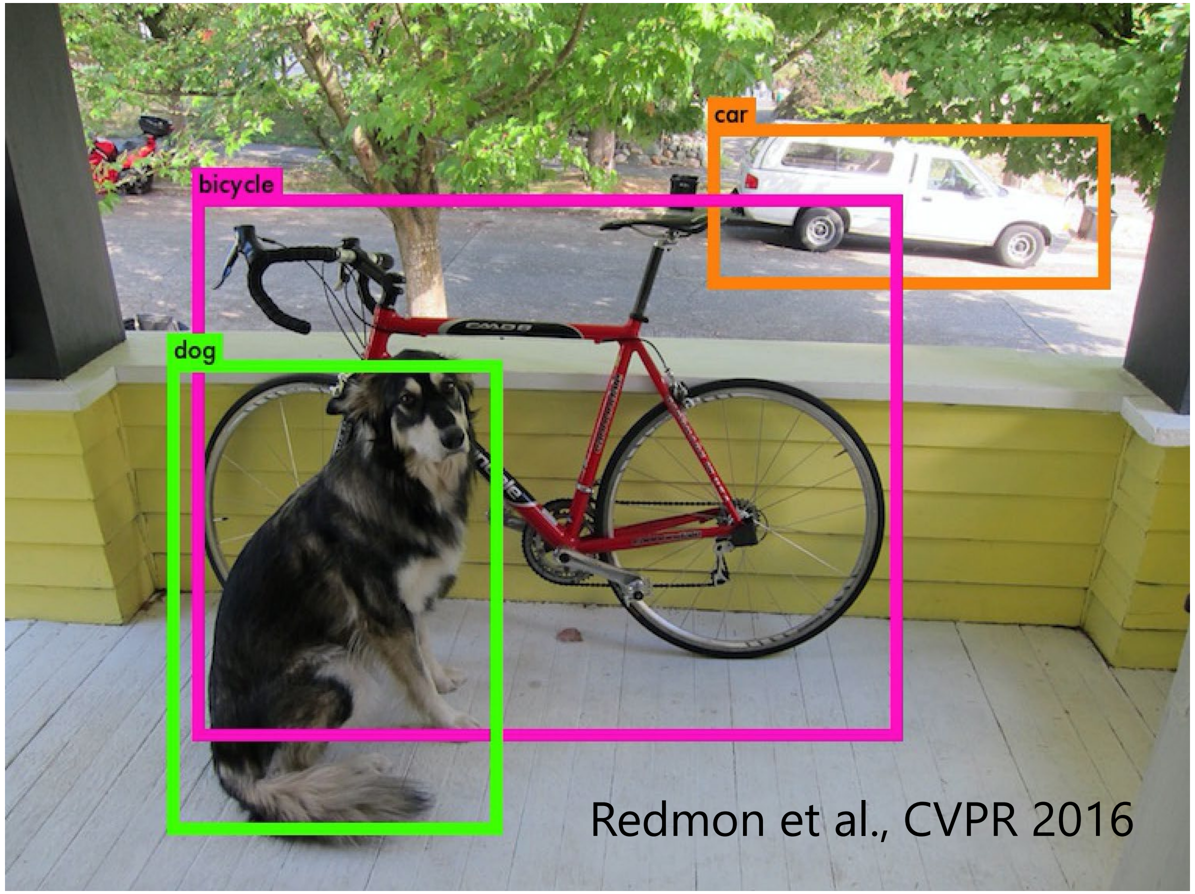


Image *Recognition*

Image *Segmentation*

How does a computer process an image?

097	097	097	097	097	097	097	097	096	097	097	096	096	096
100	100	100	100	100	100	101	101	102	101	100	100	100	099
105	105	105	105	105	105	105	103	102	102	101	103	104	105
109	109	109	109	109	110	107	118	145	132	120	112	106	103
113	113	113	112	112	113	110	129	160	160	164	162	157	151
118	117	118	123	119	118	112	125	142	134	135	139	139	175
123	121	125	162	166	157	149	153	160	151	150	146	137	168
127	127	125	168	147	117	139	135	126	147	147	149	156	160
133	130	150	179	145	132	160	134	150	150	111	145	126	121
138	134	179	185	141	090	166	117	120	153	111	153	114	126
144	151	188	178	159	154	172	147	159	170	147	185	105	122
152	157	184	183	142	127	141	133	137	141	131	147	144	147
130	147	185	180	139	131	154	121	140	147	107	147	120	128
035	102	194	175	149	140	179	128	146	168	096	163	101	125

An image is a grid of pixels. Each pixel has a numerical value.

For a greyscale image, the value corresponds to the intensity of that pixel.

How does a computer process an image?

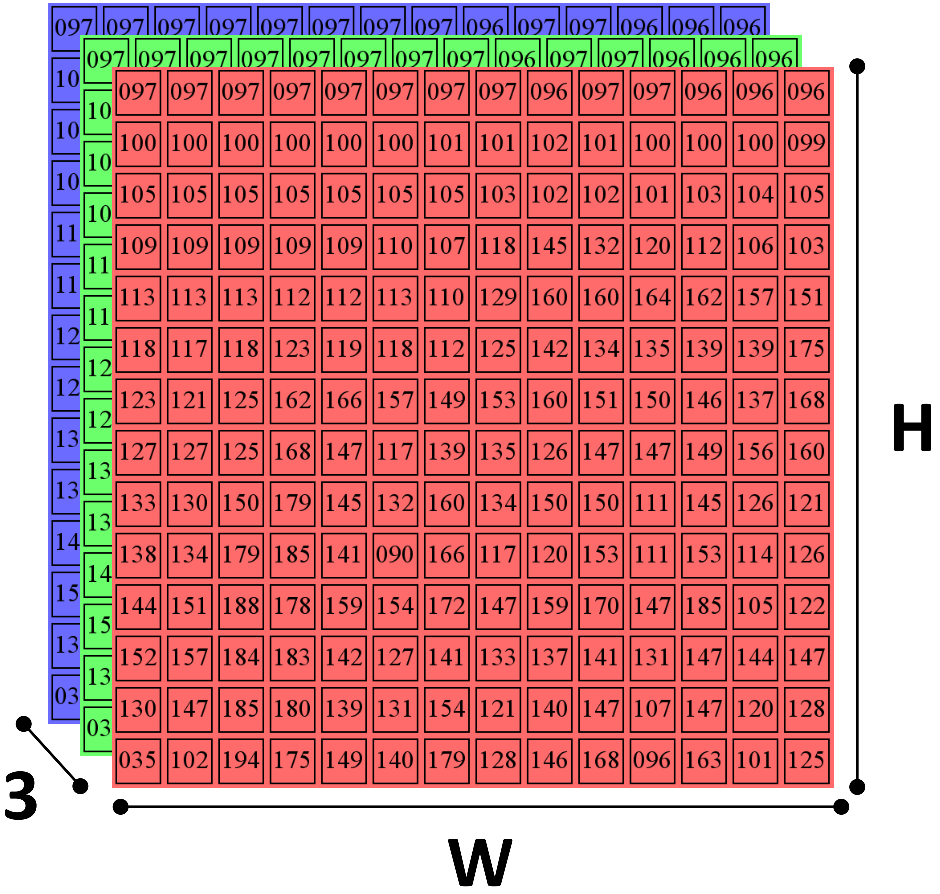
097	097	097	097	097	097	097	097	096	097	097	096	096	096
100	100	100	100	100	100	101	101	102	101	100	100	100	099
105	105	105	105	105	105	105	103	102	102	101	103	104	105
109	109	109	109	109	110	107	118	145	132	120	112	106	103
113	113	113	112	112	113	110	129	160	160	164	162	157	151
118	117	118	123	119	118	112	125	142	134	135	139	139	175
123	121	125	162	166	157	149	153	160	151	150	146	137	168
127	127	125	168	147	117	139	135	126	147	147	149	156	160
133	130	150	179	145	132	160	134	150	150	111	145	126	121
138	134	179	185	141	090	166	117	120	153	111	153	114	126
144	151	188	178	159	154	172	147	159	170	147	185	105	122
152	157	184	183	142	127	141	133	137	141	131	147	144	147
130	147	185	180	139	131	154	121	140	147	107	147	120	128
035	102	194	175	149	140	179	128	146	168	096	163	101	125

H

W

The image has size HxW.

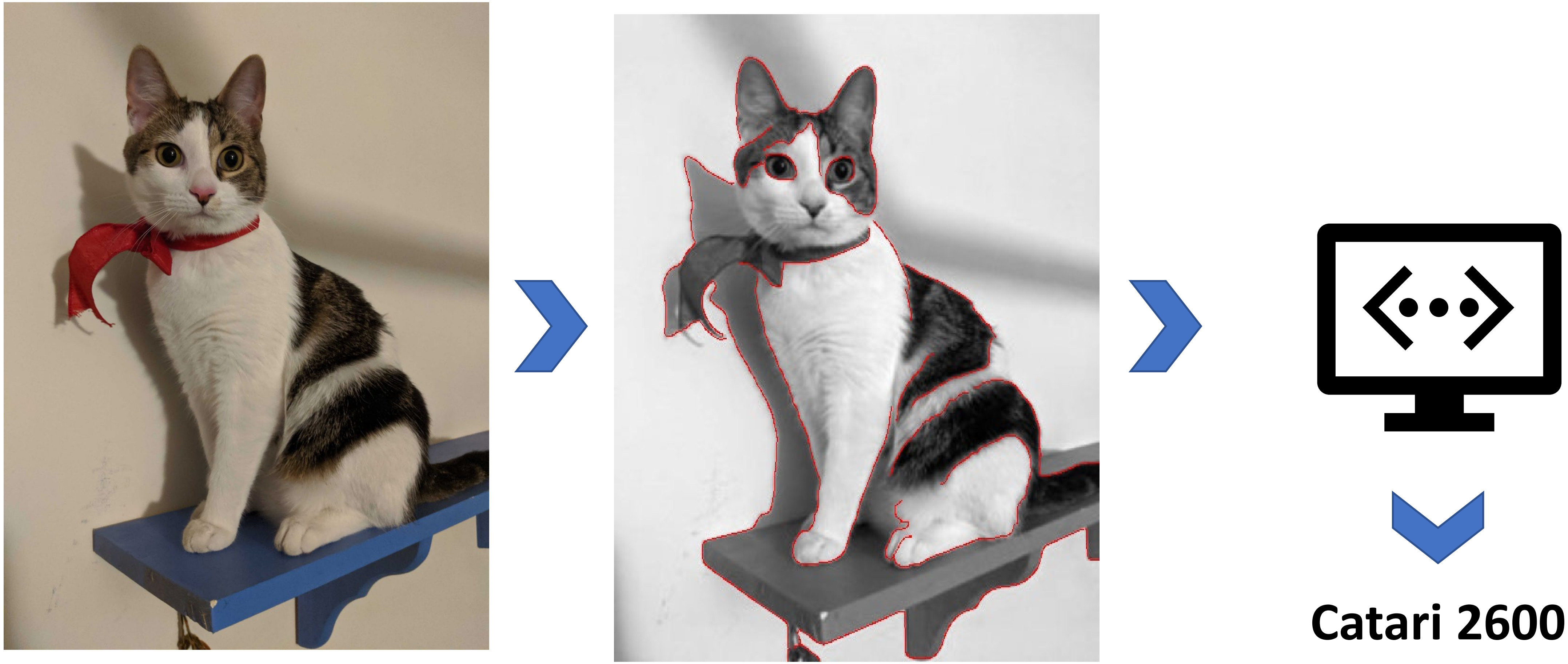
How does a computer process an image?



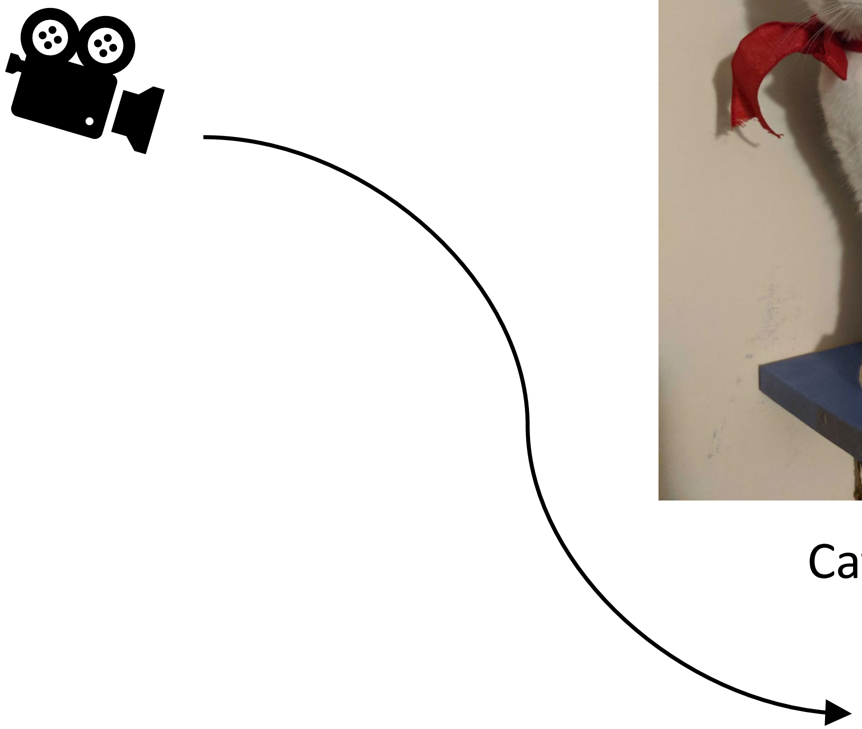
A color can be represented by 3 values at each pixel: red, green, blue (RGB)

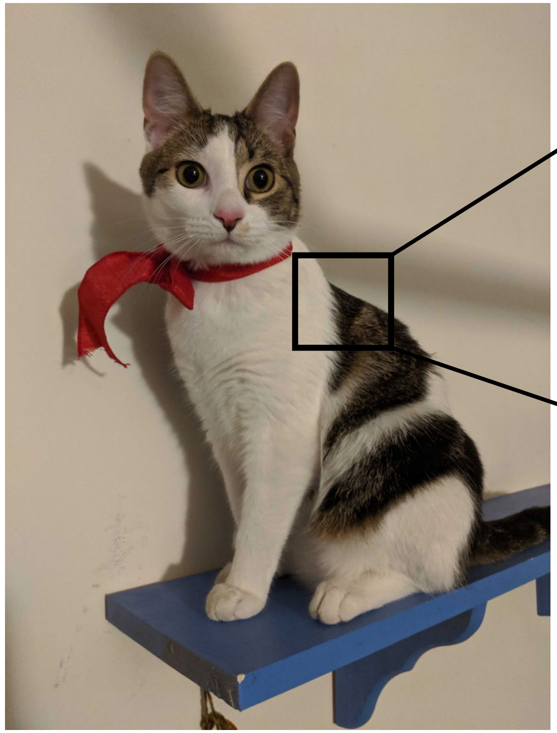
A full-color image can be represented by HxWx3 numbers.

We could try to code up a classifier…



Challenges: Viewpoint Variation



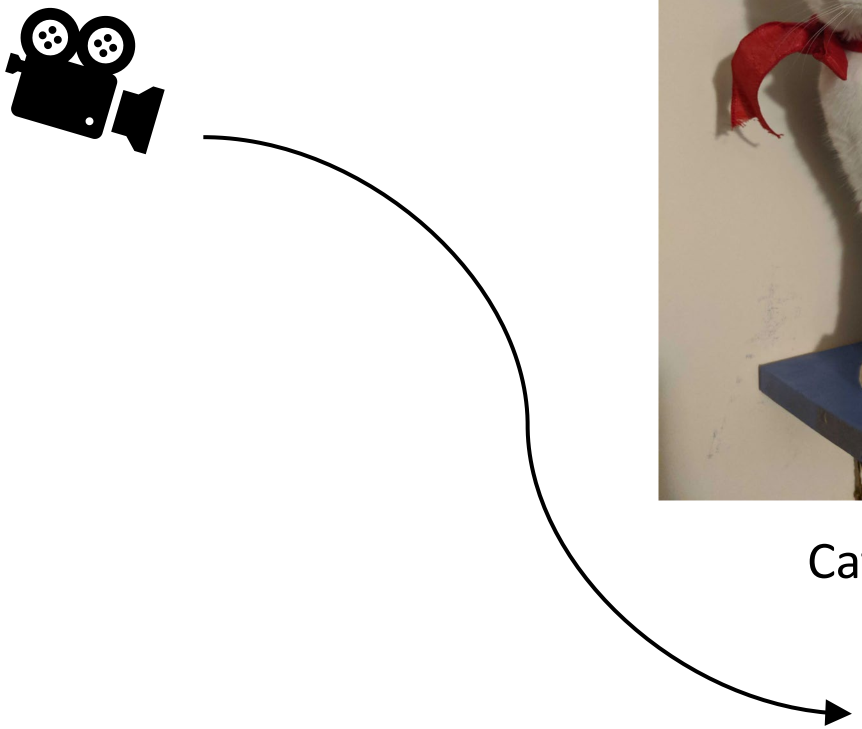


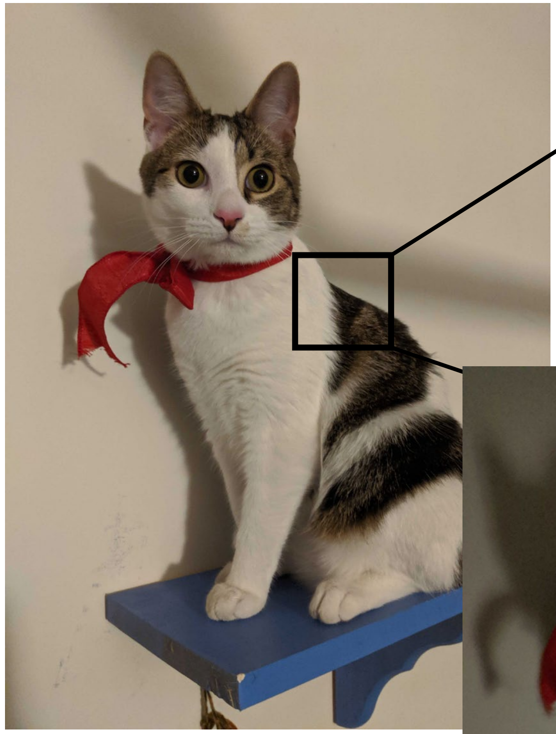
Catari 2600

097	097	097	097	097	097	097	097	096	097	097	096	096
100	100	100	100	100	100	101	101	102	101	100	100	100
105	105	105	105	105	105	105	103	102	102	101	103	104
109	109	109	109	109	110	107	118	145	132	120	112	106
113	113	113	112	112	113	110	129	160	160	164	162	157
118	117	118	123	119	118	112	125	142	134	135	139	139
123	121	125	162	166	157	149	153	160	151	150	146	137
127	127	125	168	147	117	139	135	126	147	147	149	156
133	130	150	179	145	132	160	134	150	150	111	145	126
138	134	179	185	141	090	166	117	120	153	111	153	114
144	151	188	178	159	154	172	147	159	170	147	185	105

Pixels change values when the camera moves

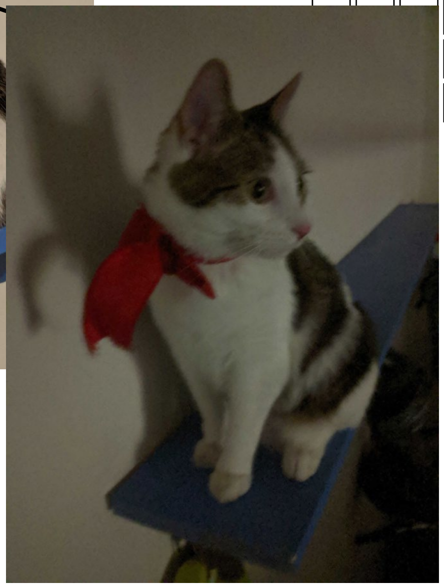
Challenges: Viewpoint Variation





Catari 2600

097	097	097	097	097	097	097	097	096	097	097	096	096
100	100	100	100	100	100	101	101	102	101	100	100	100
105	105	105	105	105	105	105	103	102	102	101	103	104
109	109	109	109	109	110	107	118	145	132	120	112	106
113	113	113	112	112	113	110	129	160	160	164	162	157
118	117	118	123	119	118	112	125	142	134	135	139	139
123	121	125	162	166	157	149	153	160	151	150	146	137
127	127	125	168	147	117	139	135	126	147	147	149	156
			179	145	132	160	134	150	150	111	145	126
			185	141	090	166	117	120	153	111	153	114
			178	159	154	172	147	159	170	147	185	105



Pixels change values when the camera moves

Challenges: Variation within Classes

This image is CC0 1.0 public domain

Credit: Justin Johnson, EECS 498 F20 (link)

Challenges: Background Clutter

This image is CC0 1.0 public domain

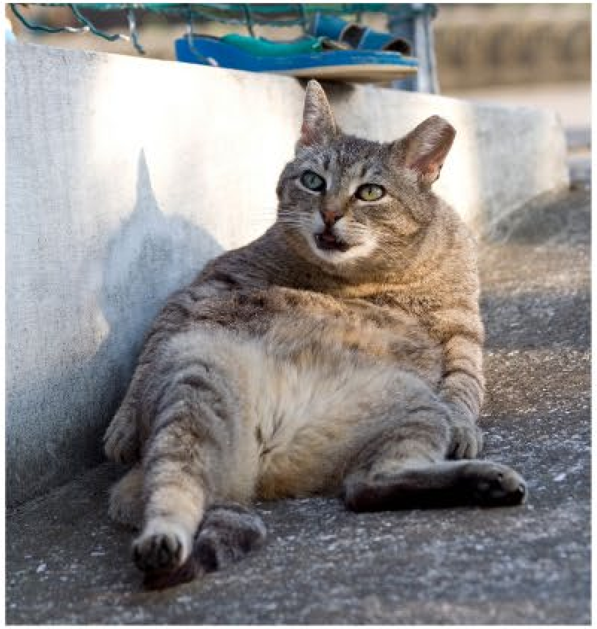
This image is CC0 1.0 public domain

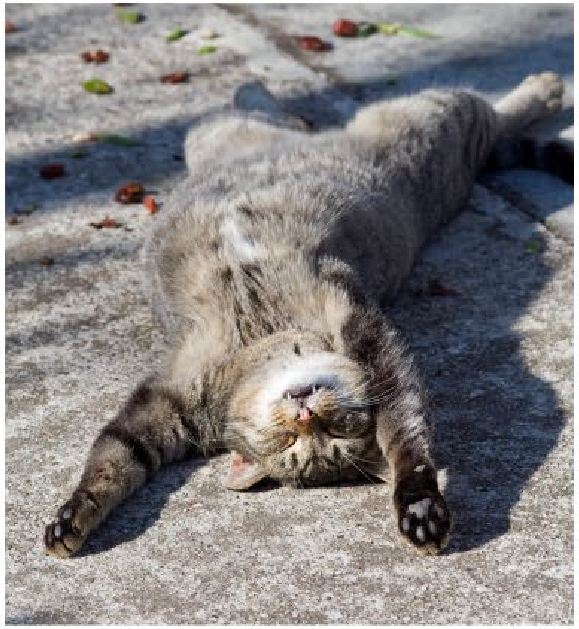
Credit: Justin Johnson, EECS 498 F20 (link)

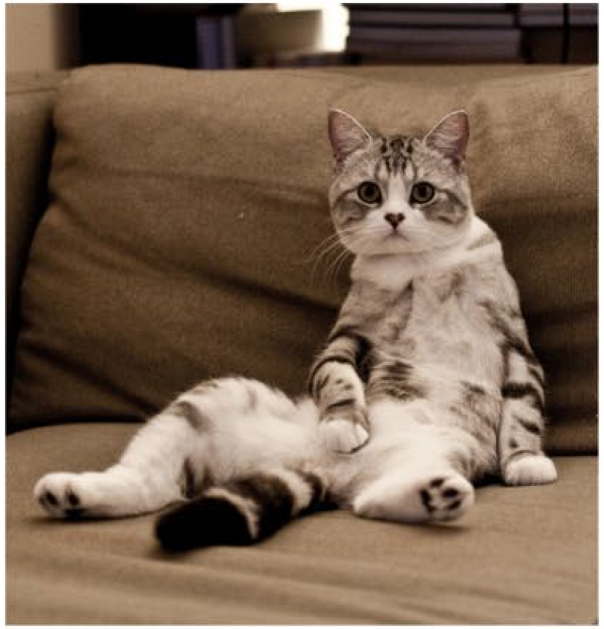
Challenges: Lighting Changes

Credit: Justin Johnson, EECS 498 F20 (link)

Challenges: Deformation







Credit: Justin Johnson, EECS 498 F20 (link)

Challenges: Occlusion

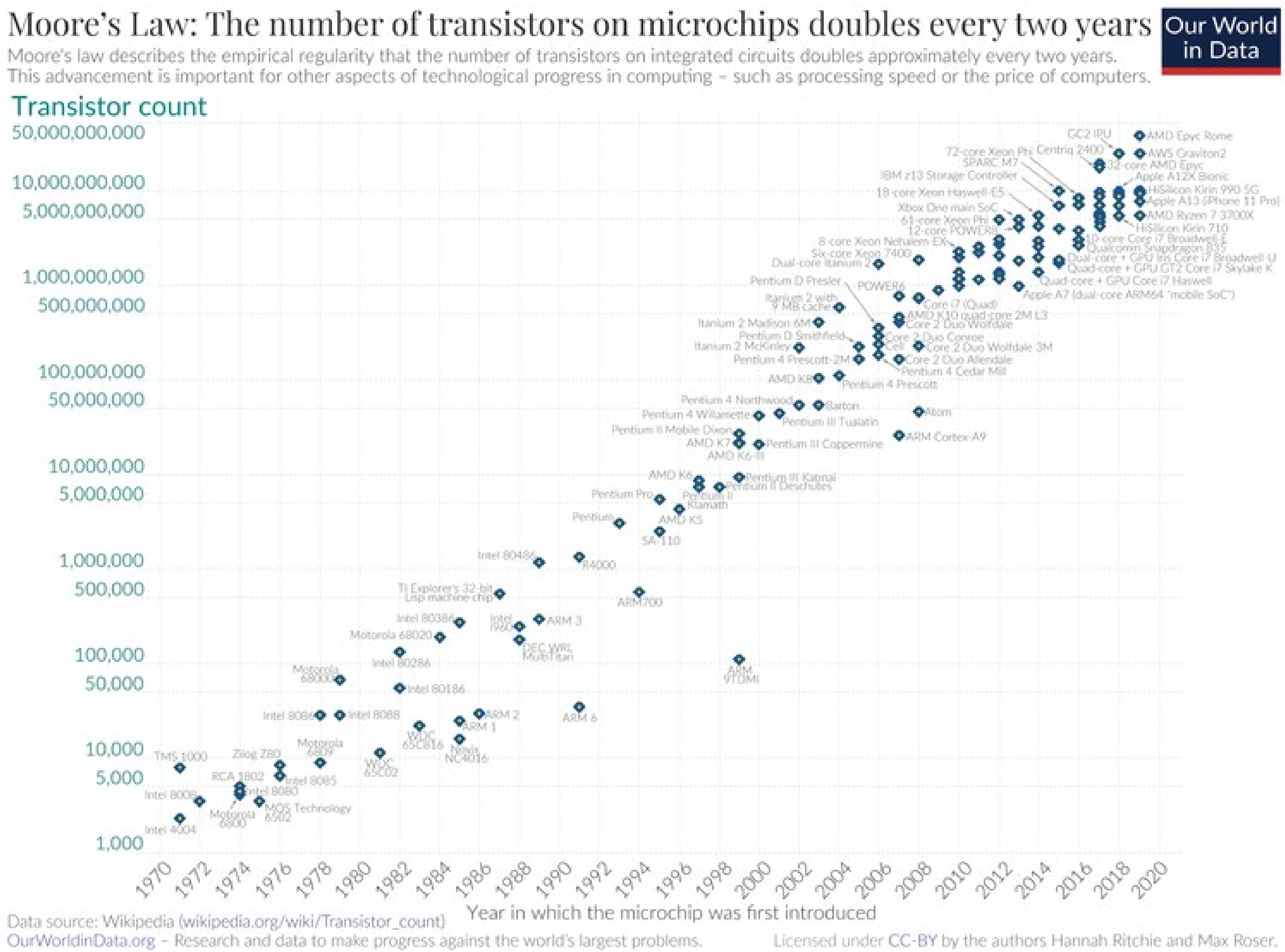
Credit: Justin Johnson, EECS 498 F20 (link)

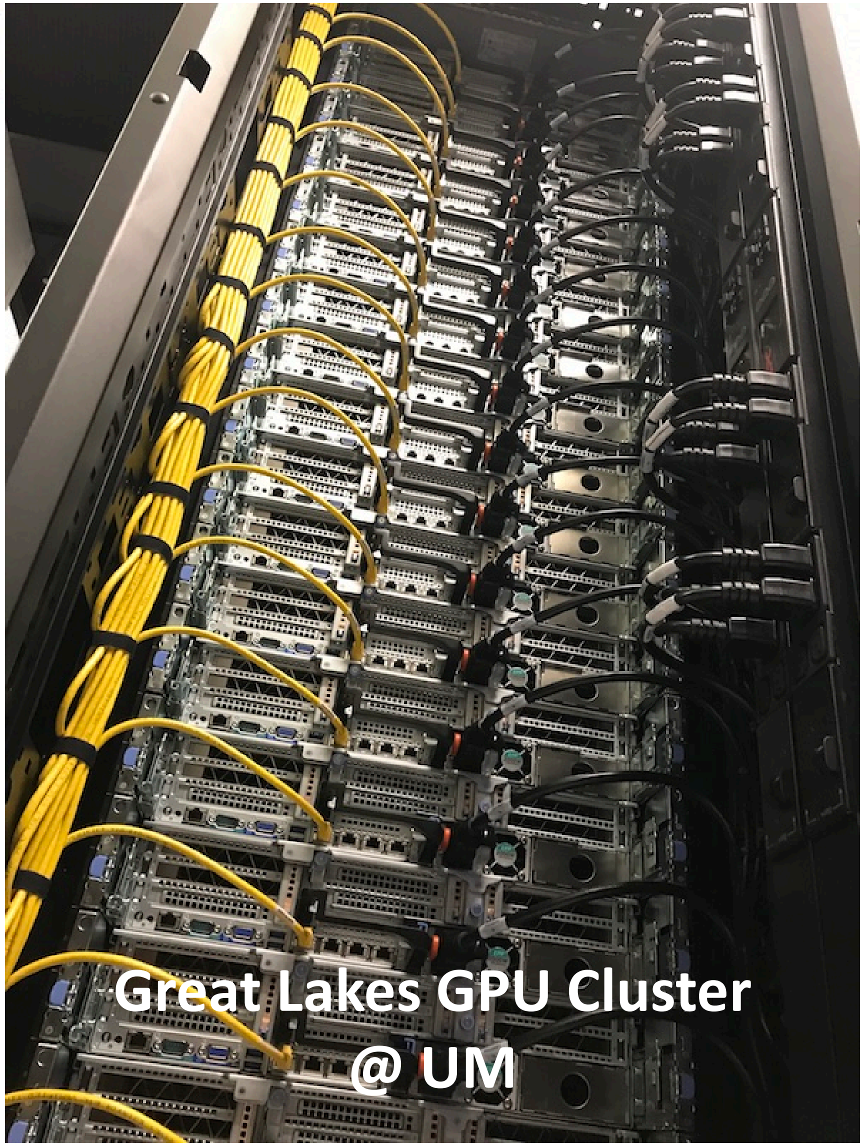
Data-Driven Approach

Idea: Use (lots & lots of) data to learn to classify images.

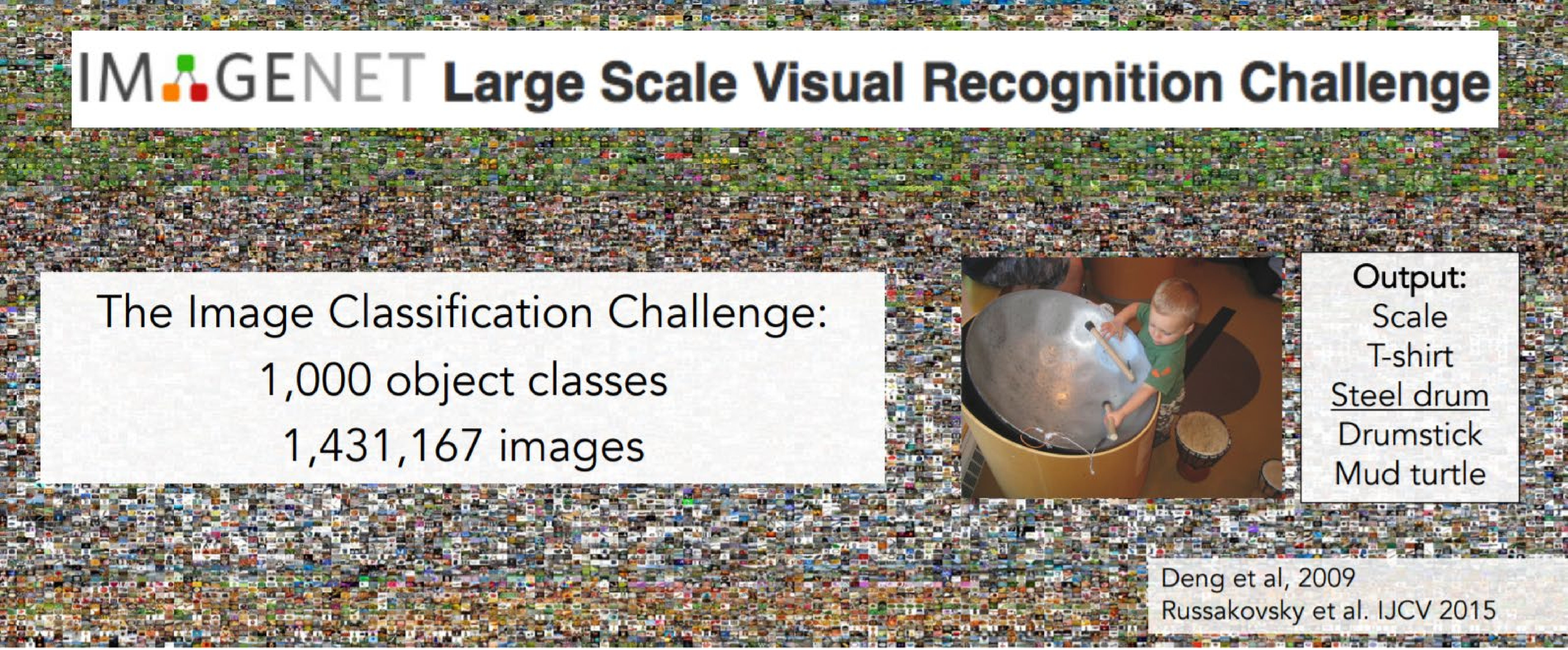
1. Get a bunch of labelled data.
2. Use Supervised Learning to train a classifier.
3. Use the classifier to label new images.

This will take lots and lots of <u>data</u> and <u>compute power</u>.





Computers have gotten very fast.



Huge open-source datasets are increasingly common.

IMAGENET Large Scale Visual Recognition Challenge

The Image Classification Challenge:

Output:
Scale

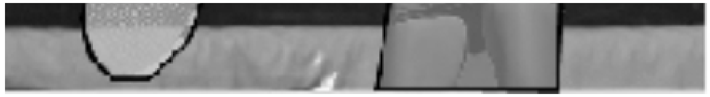
COCO 2020 Object Detection Task

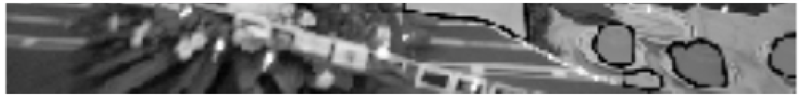
COCO Dataset (link)

Huge open-source datasets are increasingly common.

COCO 2020 Keypoint Detection Task

COCO Dataset (link)





Huge open-source datasets are increasingly common.

COCO 2020 Keypoint Detection Task

KITTI Vision Benchmark Suite ([link]) ([video])

Huge open-source datasets are increasingly common.

The MNIST dataset of images

Image: CC4.0 (link)

MNIST contains 70k images of handwritten digits. All of them are labelled as a digit from 0 to 9.

The images are 28x28 pixels (tiny!).

Machine Learning Algorithm

Training time:

Learn a prediction model by optimizing over a labeled dataset.

Testing time:

Use the model to perform prediction on new data.

Machine Learning Algorithm

Training time:

Learn a prediction model by optimizing over a labeled dataset.

Testing time:

Use the model to perform prediction on new data.

We have to pick the data, the model, and the optimization method.

Training & Testing Datasets

Training set: Labelled data used for training a machine learning algorithm.

Test set: Data used to test the accuracy of the machine learning algorithm.

- Usually smaller than the training set
- Also has labels, only used for measuring how good the algorithm is (no cheating!!)
- We don't look at this during training, so we are testing on images the algorithm has never seen before.

In MNIST, there are 60k training images and 10k test images.

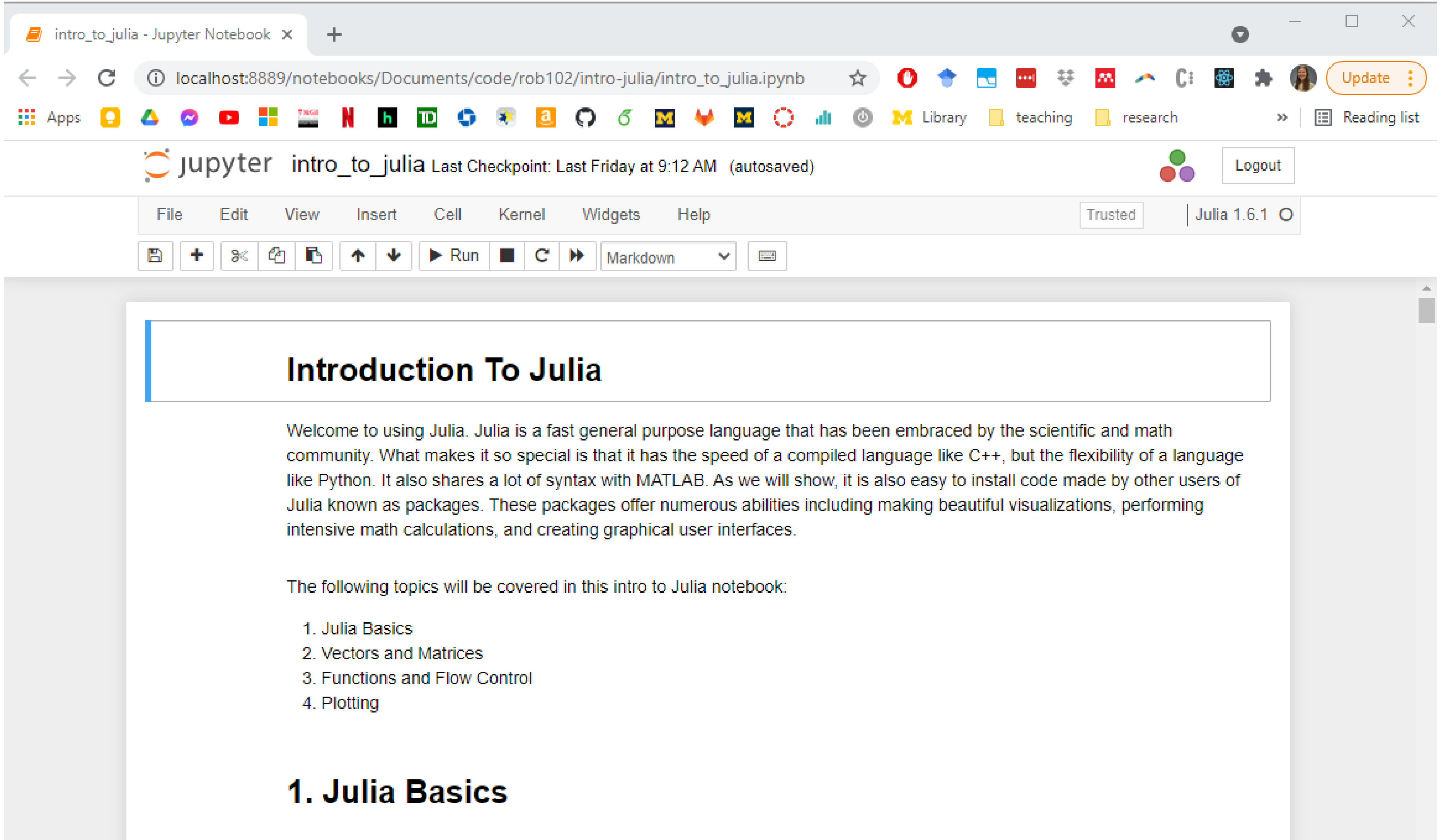
Project 4: Machine Learning

Implement three machine learning algorithms to classify images from the MNIST dataset.

1. Nearest Neighbors
2. Linear Classifier
3. Neural Network

Image: CC4.0 (link)

Julia & Jupyter Notebooks



Next time...

Implement three machine learning algorithms to classify images from the MNIST dataset.

1. **Nearest Neighbors** ← Next lecture!
2. Linear Classifier
3. Neural Network

Image: CC4.0 (link)