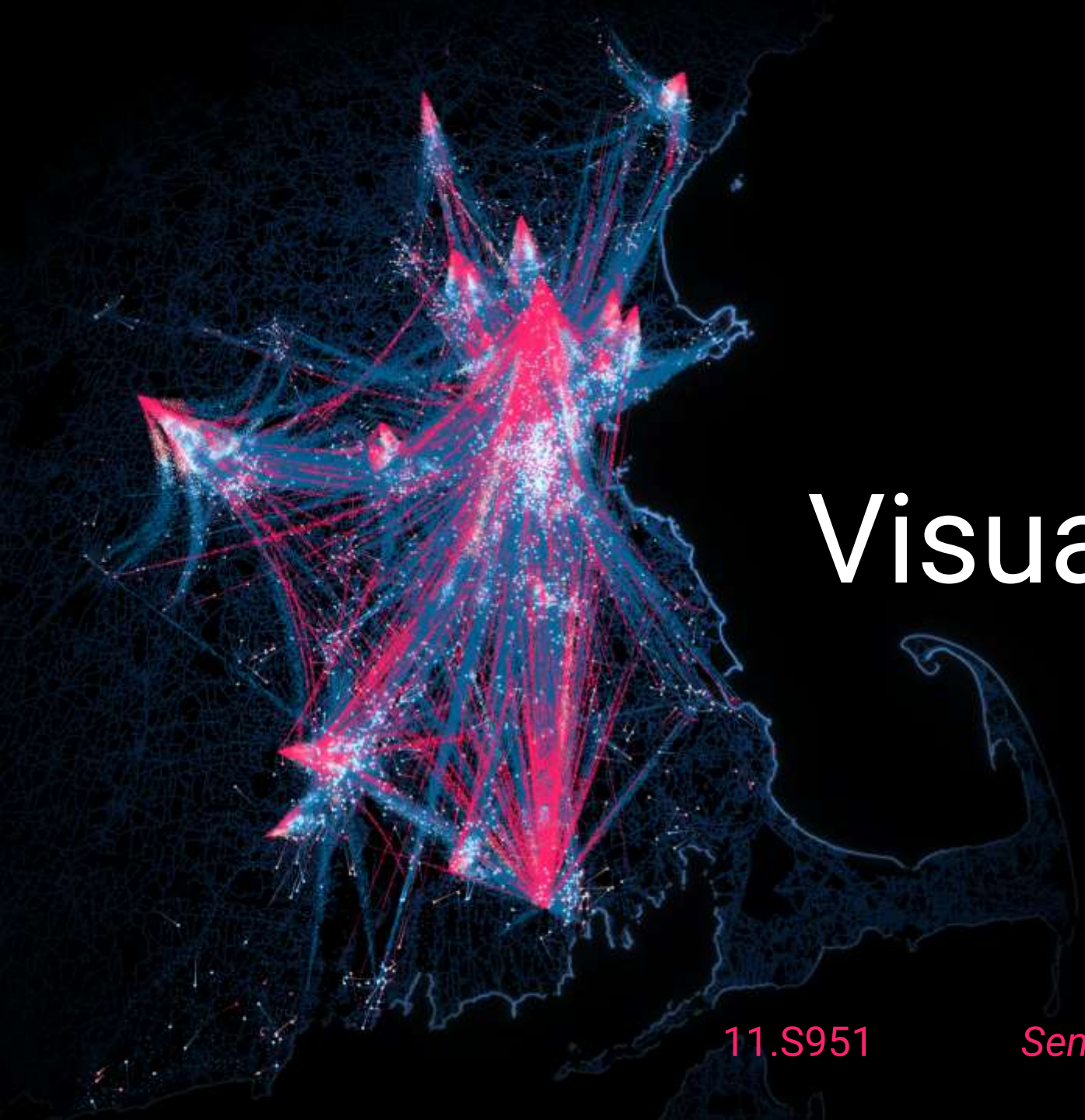


Week 6

Visual AI in Practice

FAN ZHANG, MAORAN SUN, NIKITA KLIMENKO

Review

Image Classification

Image Object Detection

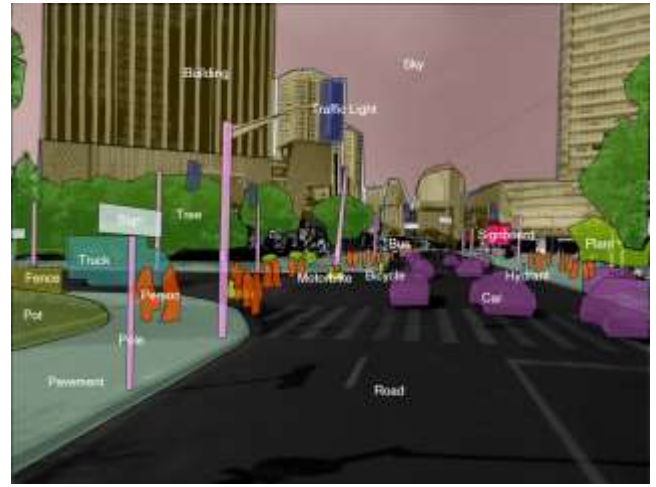


Image Segmentation

Review

- Treepedia

- AI Perception Map

- Infinite Corridor

- Smart Curbs

- [In]Distinct Cities

Outline

- City
  - Helsingborg, Trieste, Paris, & Stockholm
- Data
  - Passive Image Collection - GSV
  - Active Image collection – Mobile Cameras
- Method
  - Machine Learning & CNN
  - Hands-on Demonstration
    - Image Classification
    - Image Object Detection
    - Image Segmentation

City

Helsingborg, Trieste, Paris, Stockholm

Study Sites

Study Sites

Helsingborg, Sweden

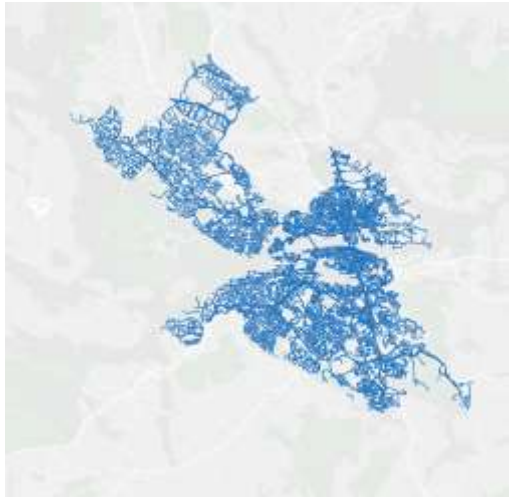
98, 693 panorama

Trieste, Italy

21, 201 panorama

Paris, France

354,400 panorama



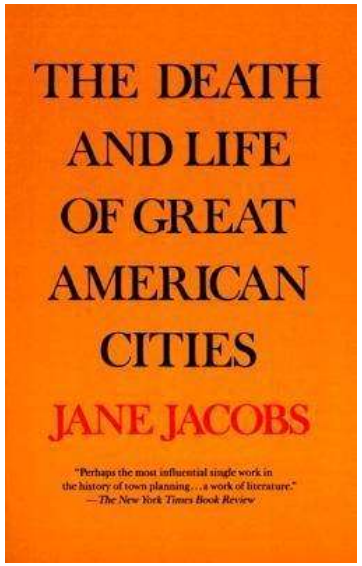
Stockholm, Sweden

252, 024 panorama

City – Helsingborg

SAFETY, TIPPING POINT

Motivation



The International Crime Prevention Through Environmental Design Association

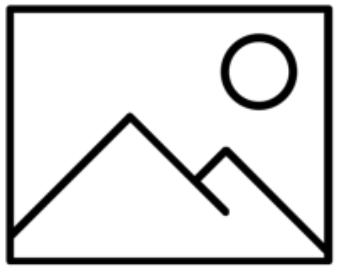
Crime and built environment

Hypothesis

Tipping Point Theory

Neighborhoods in bad physical condition will get progressively worse, whereas nicer areas will get better?

Data



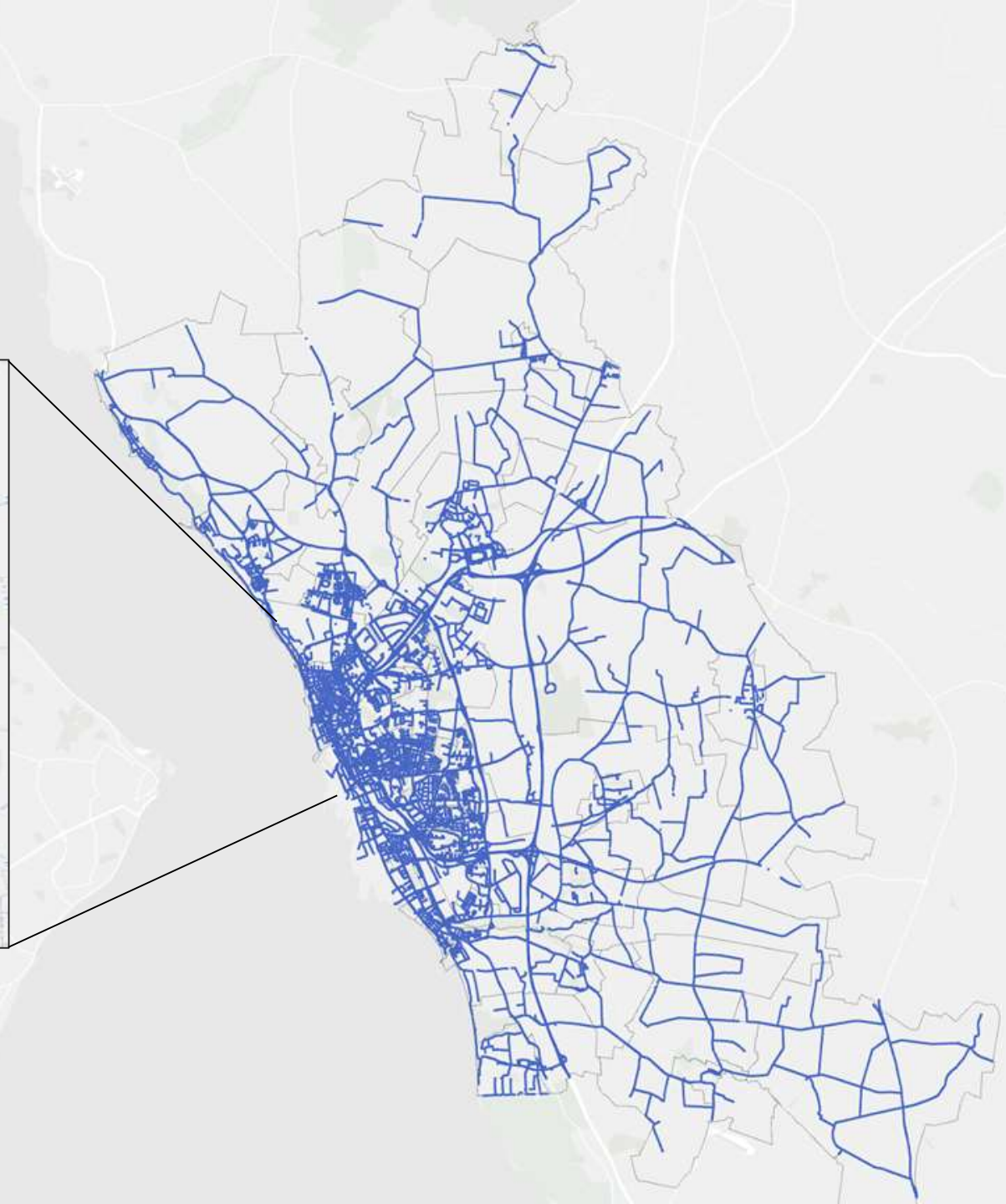
Street view image

Police Record

Census data

Data - GSV

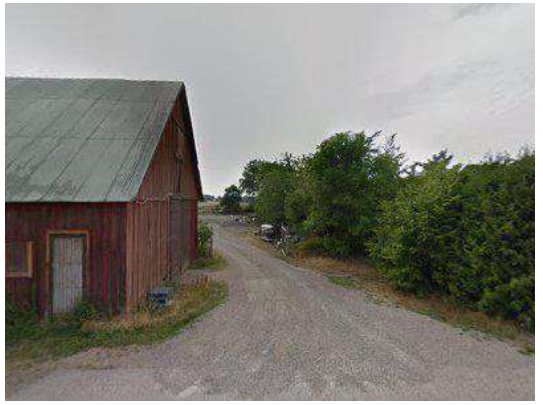
Overview of GSV data



Data - GSV

Perceived safe

Perceived unsafe



Data - GSV

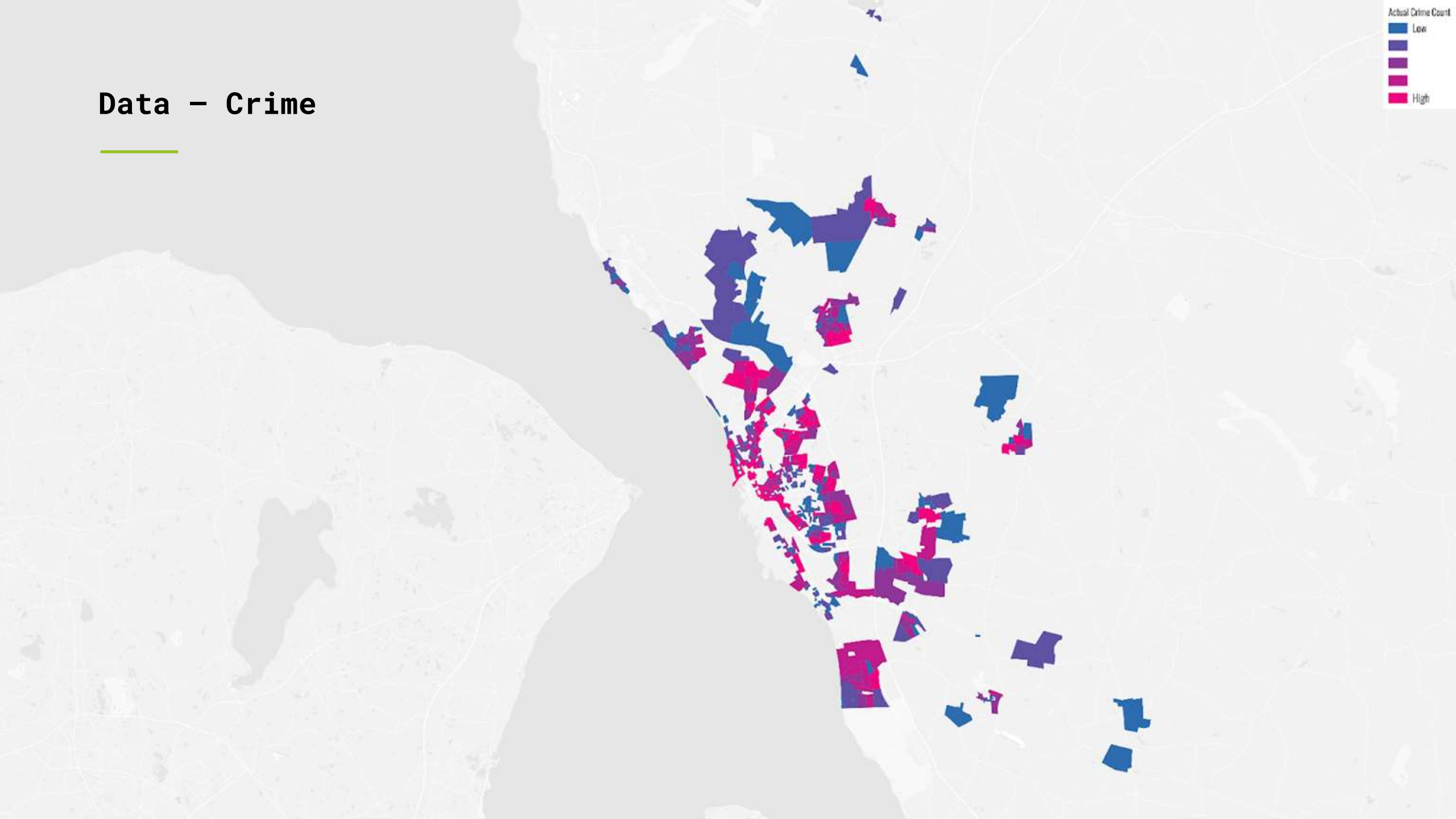
Filter GSV data

Data – Police Record

	BROTTSKOD	BROTTSTEXT	BROTTSTID_START	BROTTSDAG_START	BROTTSTID_SLUT	BROTTSDAG_SLUT	C_CODE	E_CODE	Quarter
0	414	Ofredande mot grupp	01:15	torsdag	01:30	torsdag	0213	21300	2015 Q1
1	1203	Skadegörelse, annan skadegörelse (ej klotter)	NaN	torsdag	NaN	lördag	0561	56110	2015 Q1
2	429	Ofredande mot pojke under 18 år	NaN	torsdag	NaN	fredag	1931	193141	2015 Q1
3	428	Ofredande mot flicka under 18 år	NaN	torsdag	NaN	fredag	1931	193141	2015 Q1
4	1203	Skadegörelse, annan skadegörelse (ej klotter)	00:35	torsdag	00:35	torsdag	0834	83410	2015 Q1

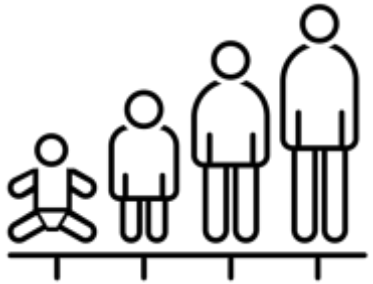
Data source: Helsingborg Police Department

Data – Crime

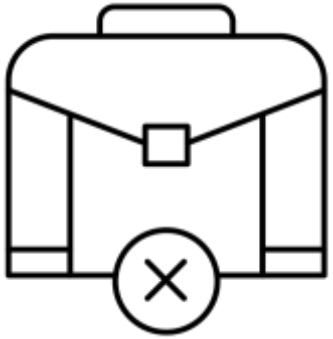


Data – Demographics

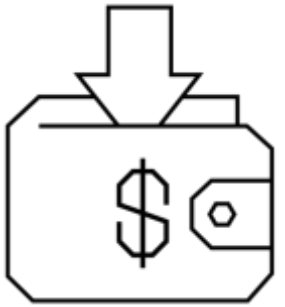
Daytime population



Age

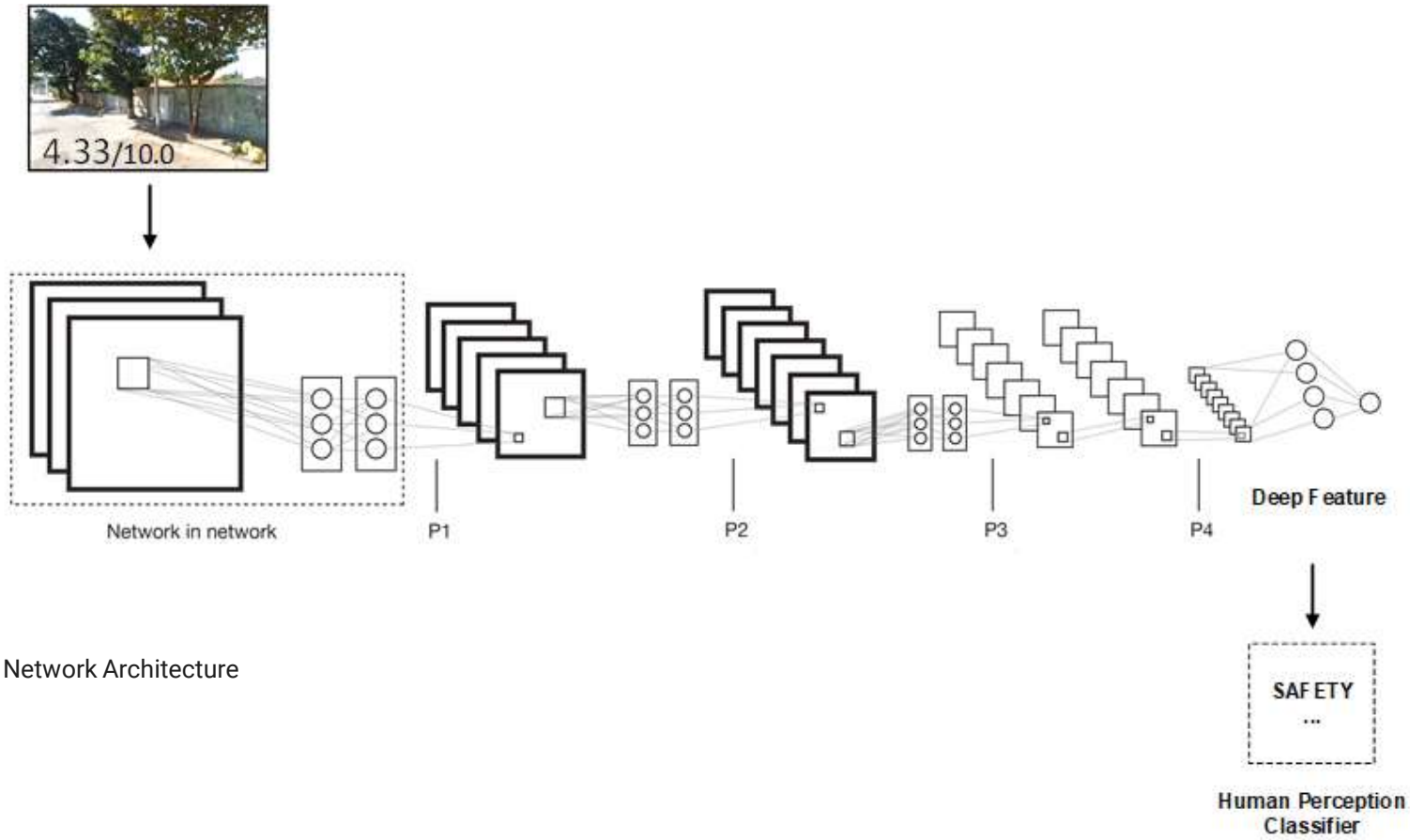


Unemployment rate



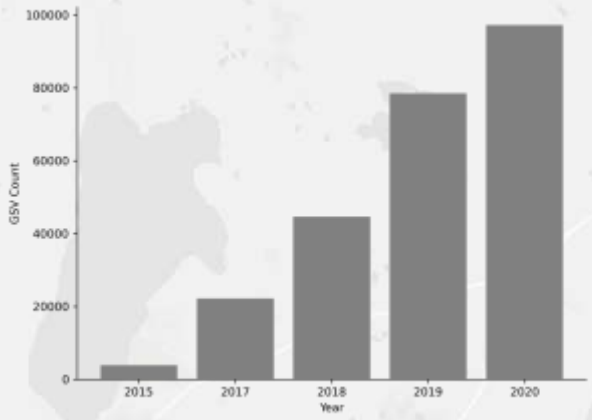
Average income

Methodology – Perception



Neural Network Architecture

Methodology – Perception



GSV numbers by year

Result – Tipping Point Theory

$$CrimeChange = \beta * GSVSafetyScore + \theta * BuiltEnvironment + \lambda * Demographics$$

$Crime_{2019} - Crime_{2015}$

$DistancetoCityCenter$

$DaytimePopulation$

$AverageIncome$

$UnemploymentRate$

Result – Tipping Point Theory

$$CrimeChange = \beta * GSVSafetyScore + \theta * BuiltEnvironment + \lambda * Demographics$$

OLS Regression Results

Dep. Variable:	crime_change_standardize	R-squared:	0.092
Model:	OLS	Adj. R-squared:	0.069
Method:	Least Squares	F-statistic:	3.994
Date:	Mon, 21 Feb 2022	Prob (F-statistic):	0.00179
Time:	20:24:06	Log-Likelihood:	-279.16
No. Observations:	204	AIC:	570.3
Df Residuals:	198	BIC:	590.2
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-1.388e-17	0.068	-2.05e-16	1.000	-0.133	0.133
GSV_safety_standardize	-0.1872	0.076	-2.450	0.015	-0.338	-0.037
daytime_population	0.2134	0.068	3.135	0.002	0.079	0.348
aver_income_standardize	0.0517	0.082	0.633	0.527	-0.109	0.213
unemployment_rate	0.1213	0.081	1.489	0.138	-0.039	0.282
dist_to_center_standardize	-0.0290	0.076	-0.382	0.703	-0.179	0.121

Omnibus:	168.764	Durbin-Watson:	2.202
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6170.823
Skew:	-2.700	Prob(JB):	0.00
Kurtosis:	29.397	Cond. No.	1.91

Data – Passive Collection

GOOGLE STREET VIEW

Street View Images

Image credit: Google

How Google collects data

- Street View Car

- Street View Trolley

- Street View Trekker

Street View Service

Image credit: wikipedia

Google Street View

Image credit: Google

GSV Perspectives

Natural view

(a)

Panoramic view

(b)

GSV API

https://maps.googleapis.com/maps/api/streetview/metadata?location={}&key={}

Location	Key
lat,lng	Do not share with others!

[Example 1]

GSV API

https://maps.googleapis.com/maps/api/streetview?size={width}x{height}&pano={}&heading={}&pitch={}&fov={}&key={}

Size

Max size: 640 * 640

Pano

Unique panoID

location

Unique panoID

Heading

0 - 360 degree

North: 0

East: 90

South: 180

West: 270

Pitch

-90 - 90 degree

Straight up: 90

Stright down: -90

FOV

Max value: 120

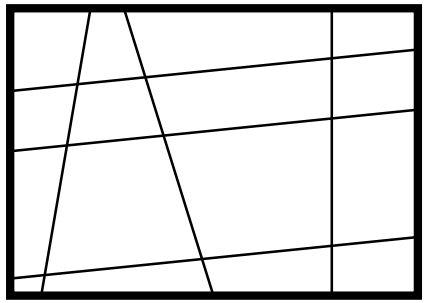
Default: 90

Key

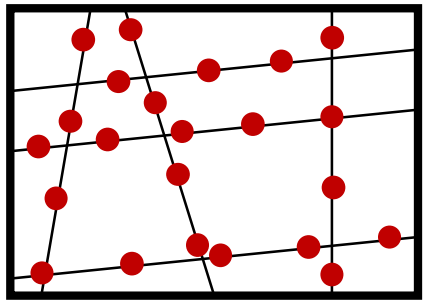
Do not share with others!

[Example](#)2

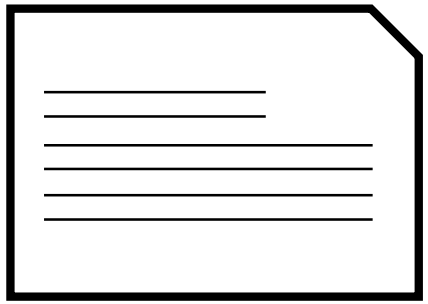
Large-scale GSV Download Process



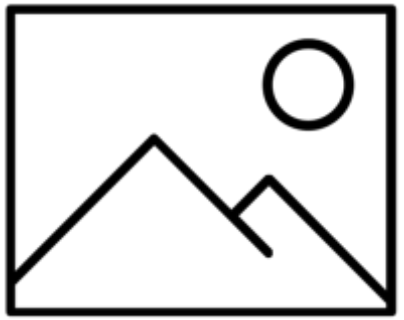
Download street network



Generate request point



Get image meta data



Download image

Large-scale GSV Download Example

Download street network

Generate request point

Tutorial for GSV collection

https://colab.research.google.com/drive/1o2cB5WuvF4vmsukeZsxTCx4ePr7jECpA?usp=sharing

Data – Active Collection

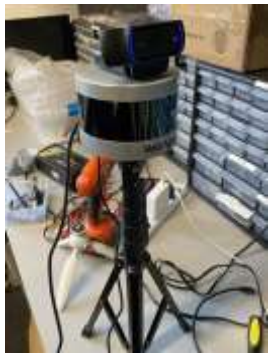
MOBILE SENSEING, CAMERA

Self-collection

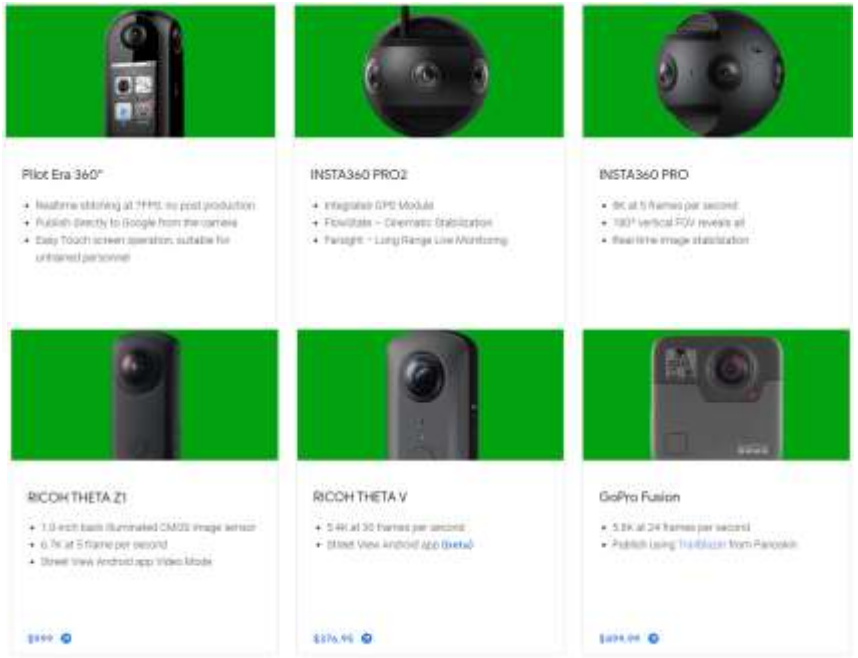
❑ Wearable Camera

GoPro

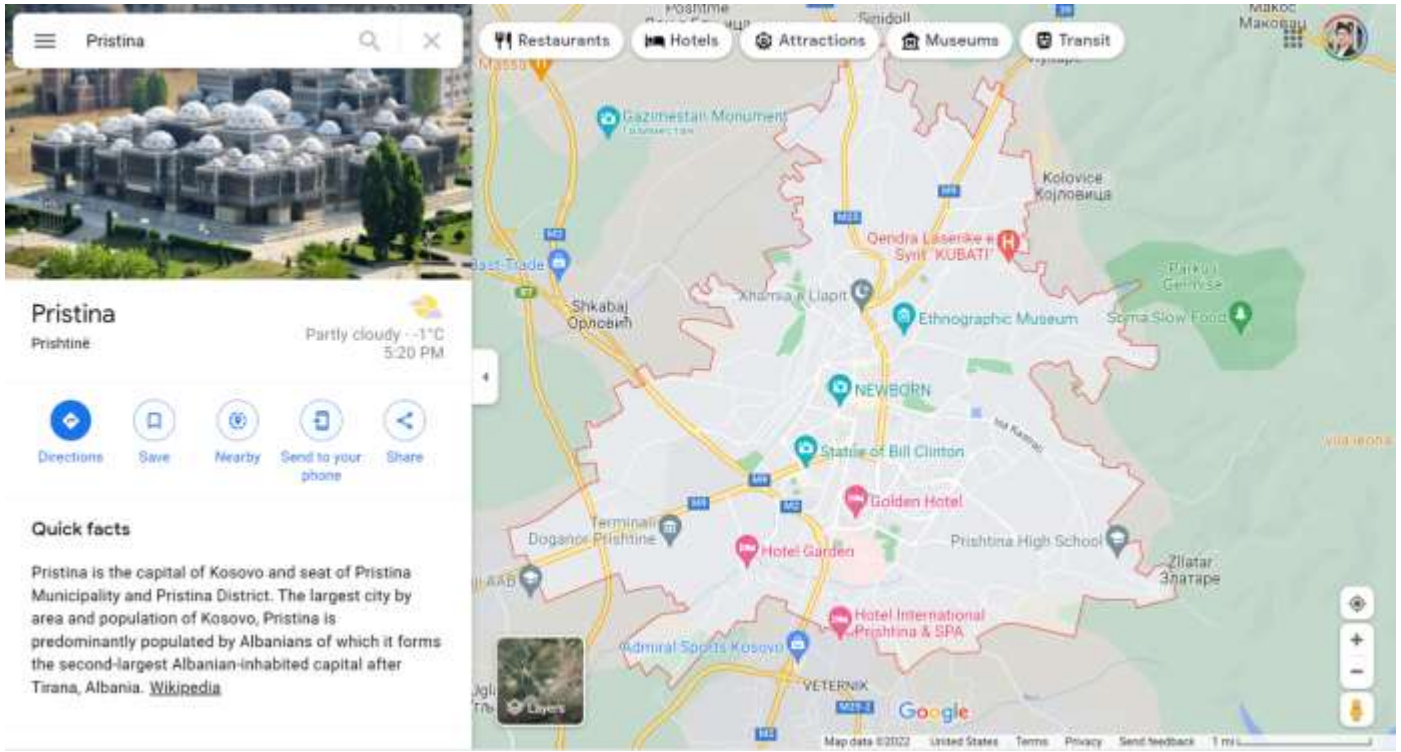
❑ Vehicle Mouted

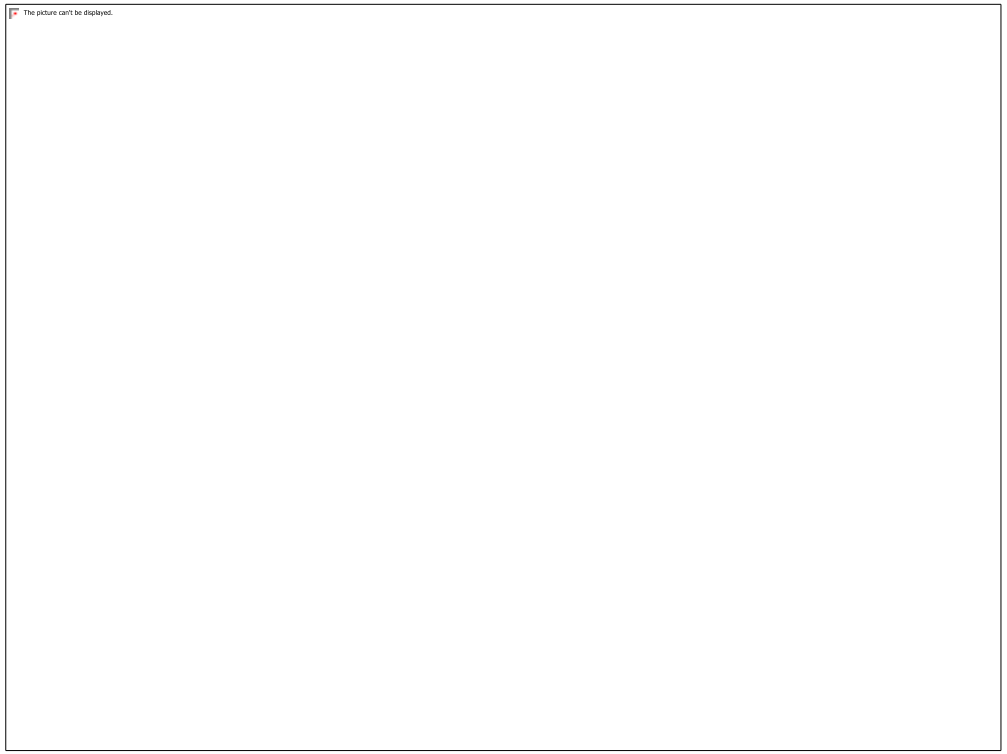


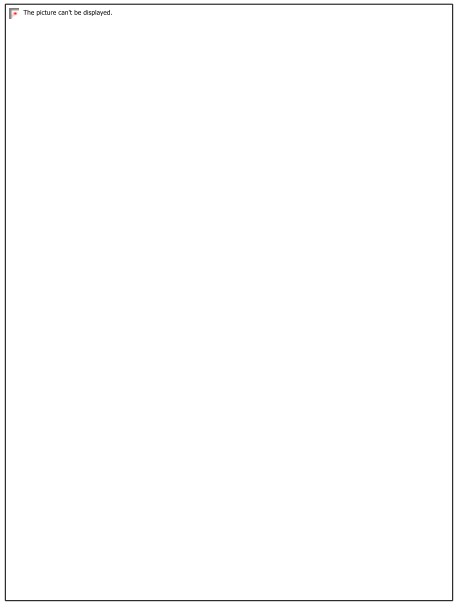
❑ Device

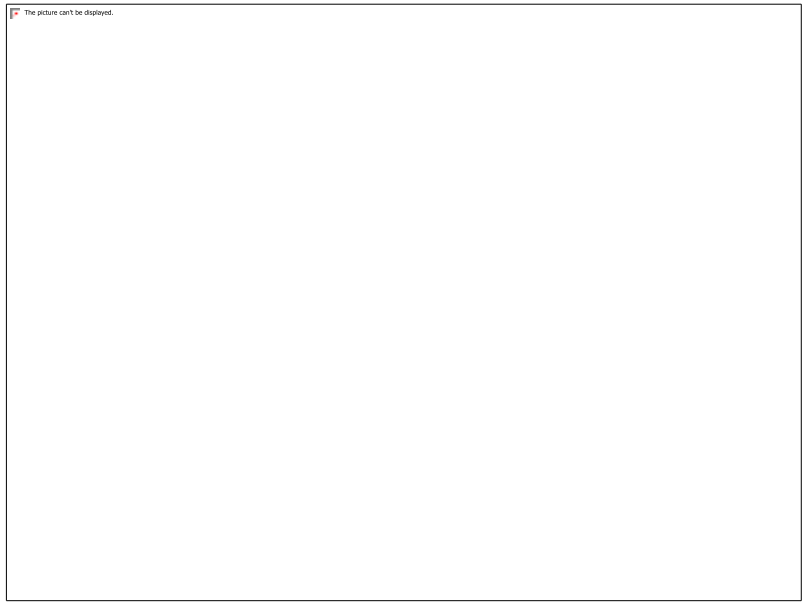


https://www.google.com/streetview/contacts-tools/


Pristina
Restaurants
Hotels
Attractions
Museums
Transit
Gazmendasti Monument
Kolovice
Qendra Laserske e Syrit "KUBATI"
Shkabaj
Ethnographic Museum
NEWBORN
Statue of Bill Clinton
Golden Hotel
Terminali Doganor Prishtine
Hotel Garden
Prishtina High School
Zllatar
Hotel International Prishtina & SPA
Admiral Sports Kosovo
VETERNIK
Google
Map data ©2022
United States
Terms
Privacy
Send feedback
1 mi
Pristina
Prishtinë
Partly cloudy · -1°C
5:20 PM
Directions
Save
Nearby
Send to your phone
Share
Quick facts
Pristina is the capital of Kosovo and seat of Pristina Municipality and Pristina District. The largest city by area and population of Kosovo, Pristina is predominantly populated by Albanians of which it forms the second-largest Albanian-inhabited capital after Tirana, Albania. Wikipedia


The picture can't be displayed.

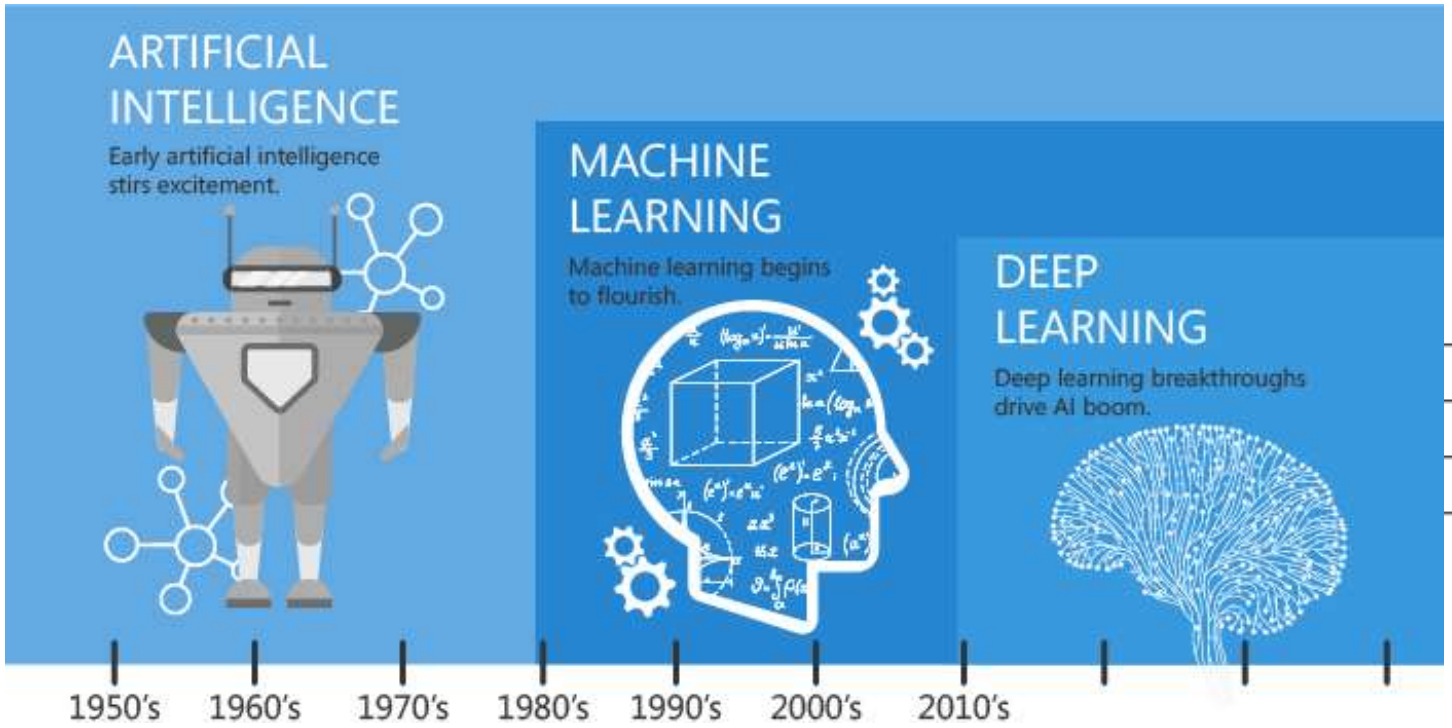

The picture can't be displayed.


The picture can't be displayed.

Method – Deep Learning

MACHINE LEARNING, NEURAL NETWORK

AI vs. ML. vs DL



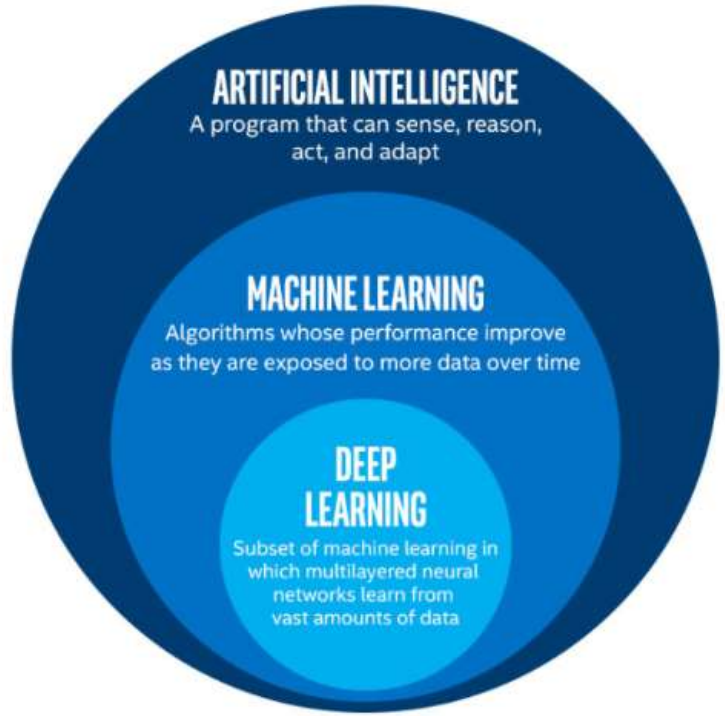
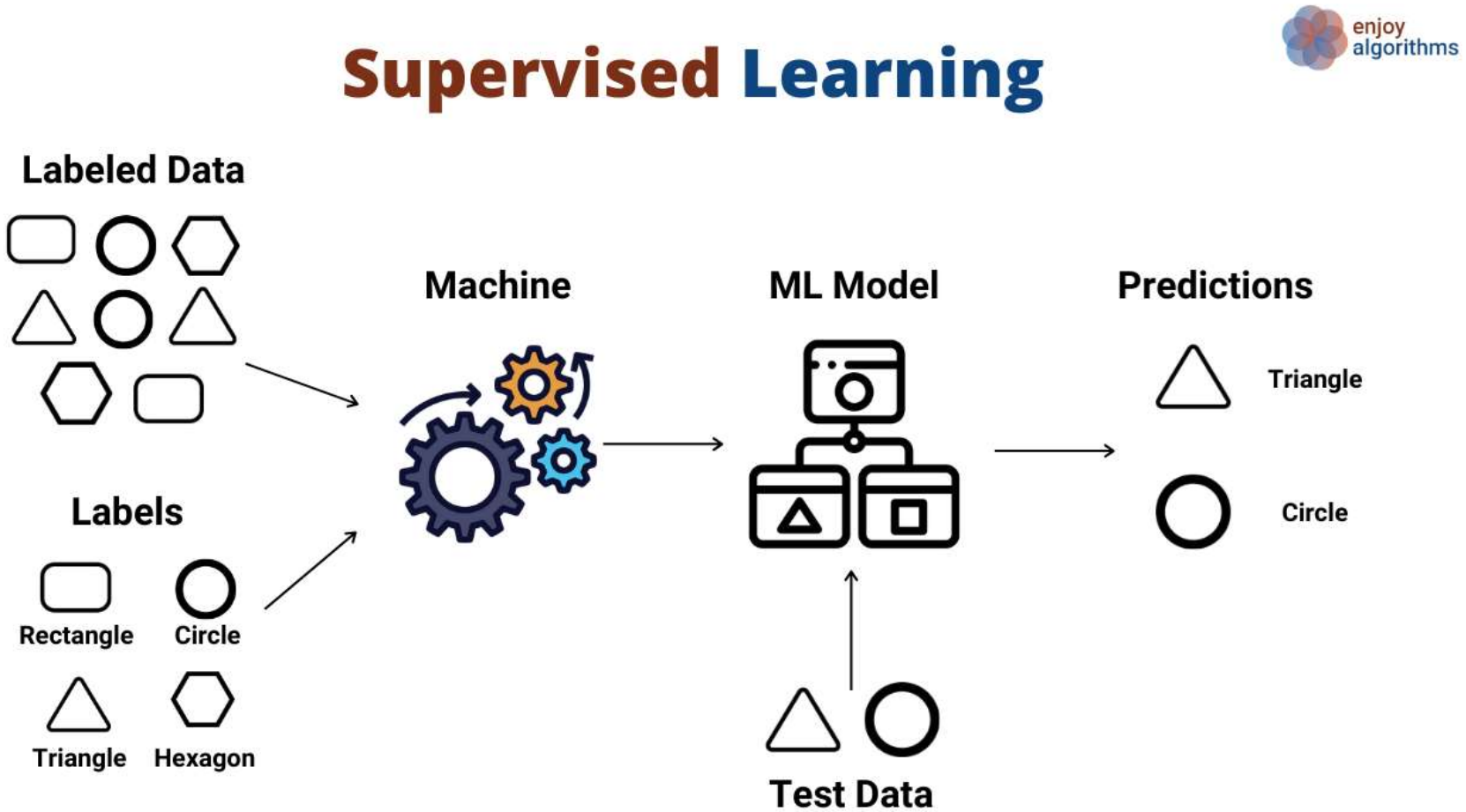
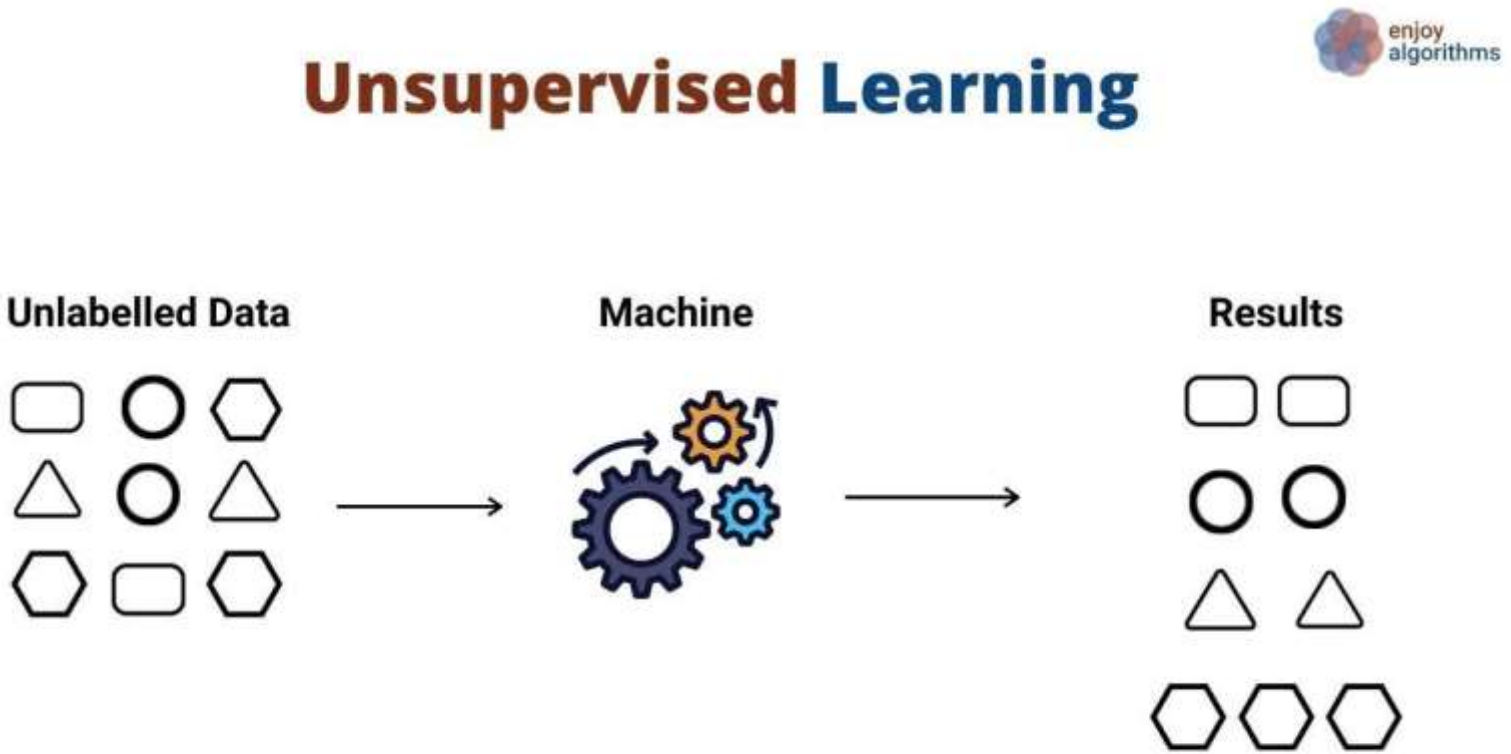


Image source: Artem Oppermann
https://towardsdatascience.com/artificial-intelligence-vs-machine-learning-vs-deep-learning-2210ba8cc4ac

Machine Learning



Machine Learning



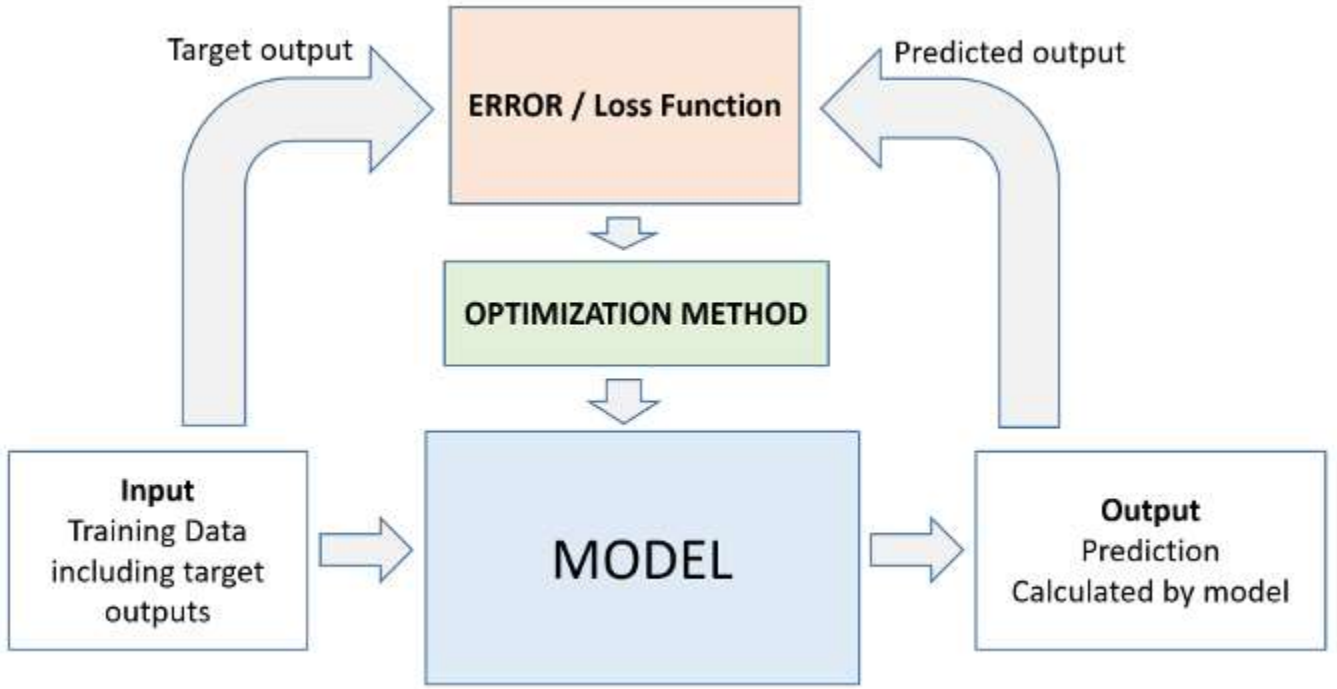
https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning

Machine Learning

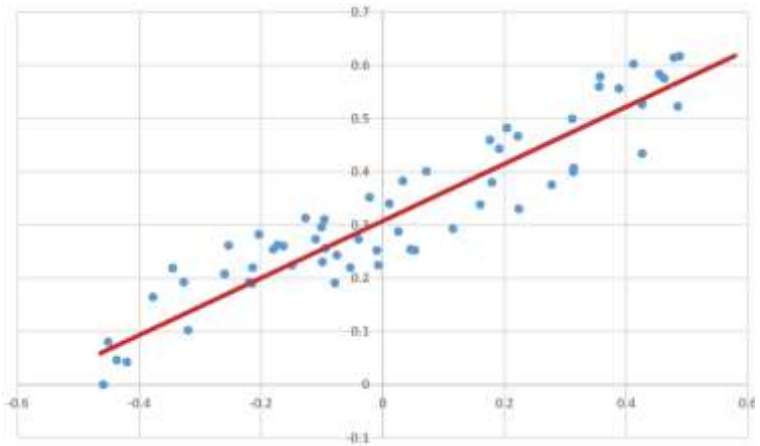
Three Key Components of ML:

1. Model
2. Loss Function
3. Optimization Method

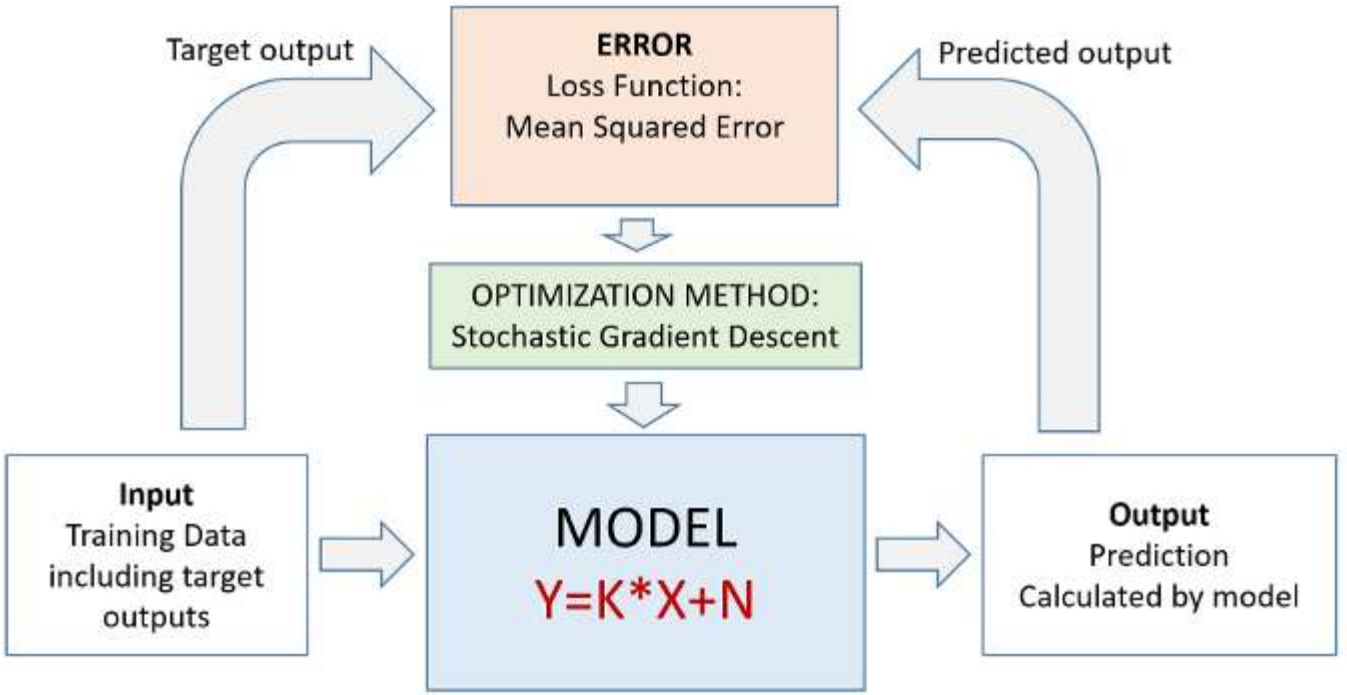
- **Learning** as loss minimization



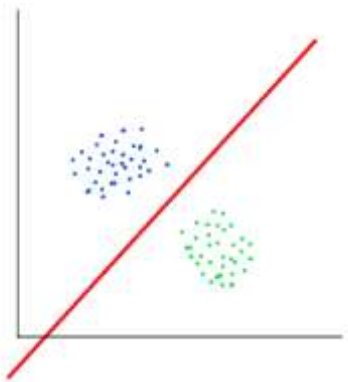
Machine Learning



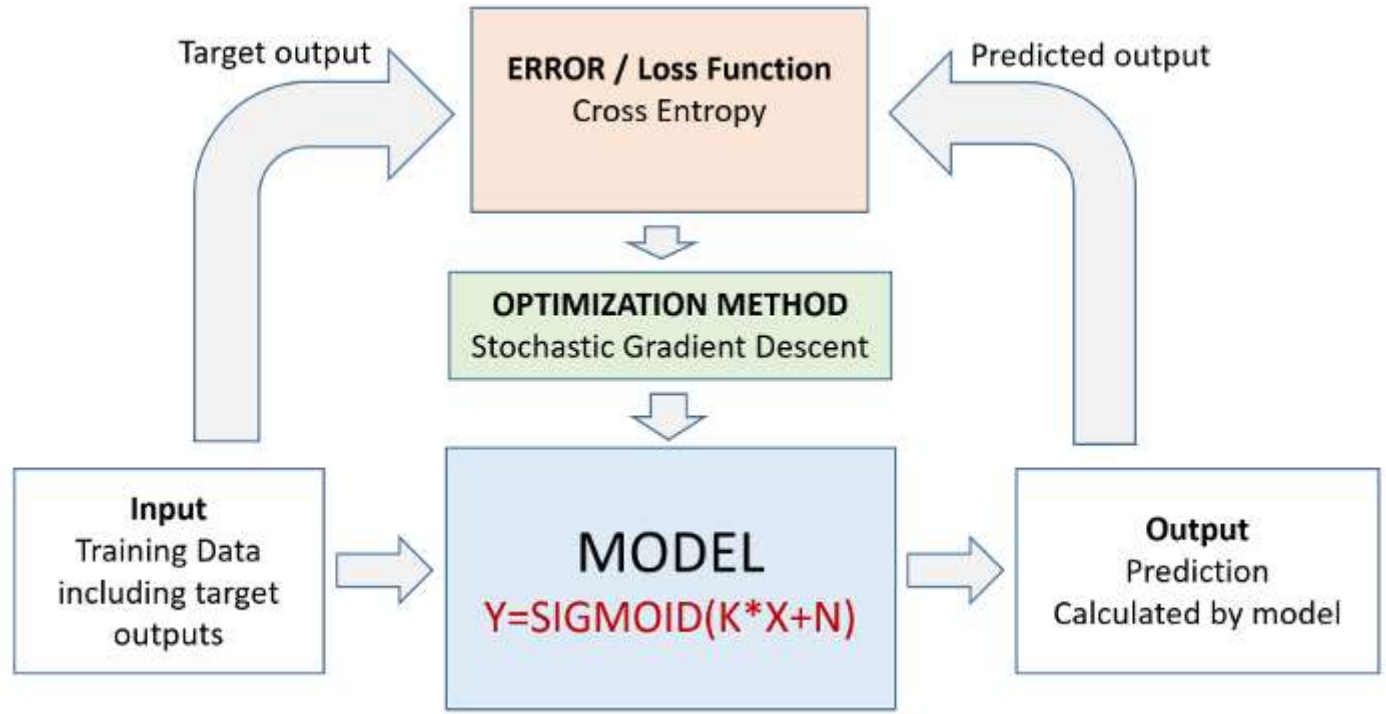
Linear Regression



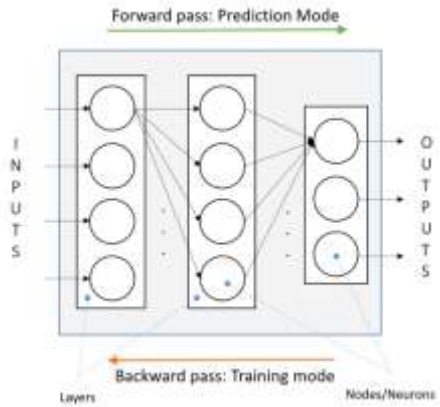
Machine Learning



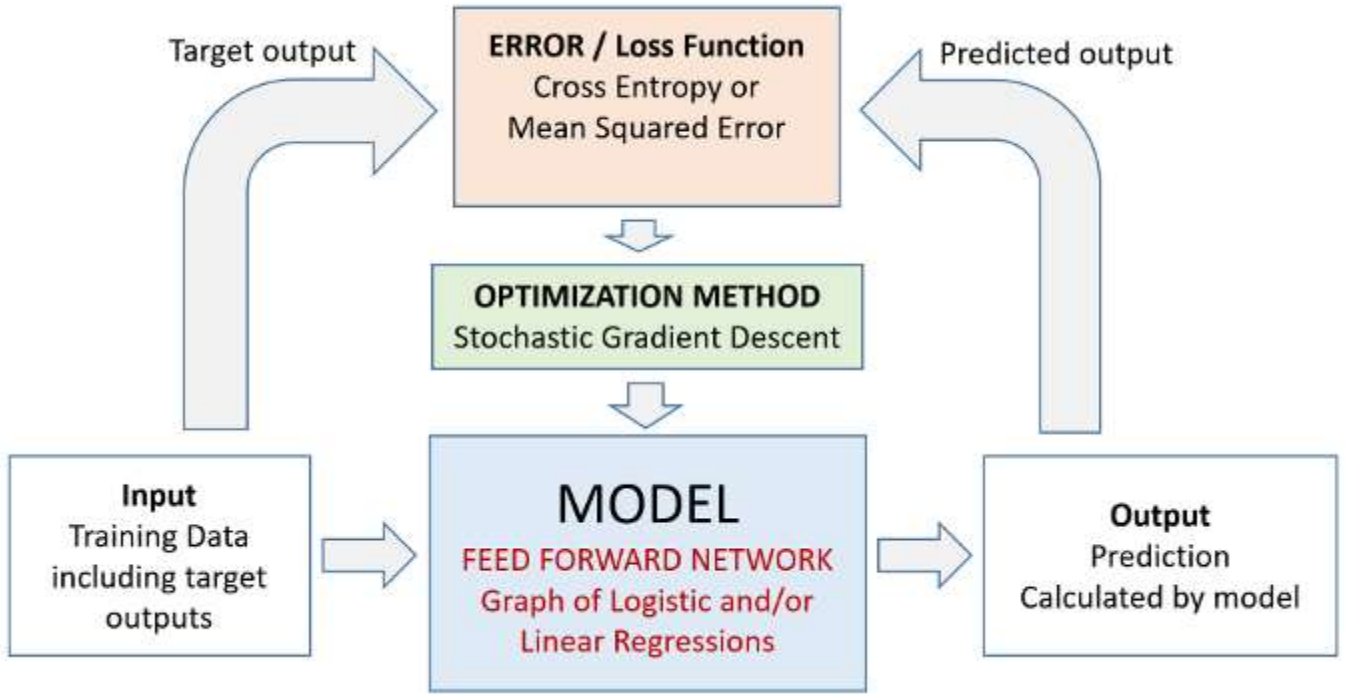
Logistic Regression



Machine Learning



Neural Network

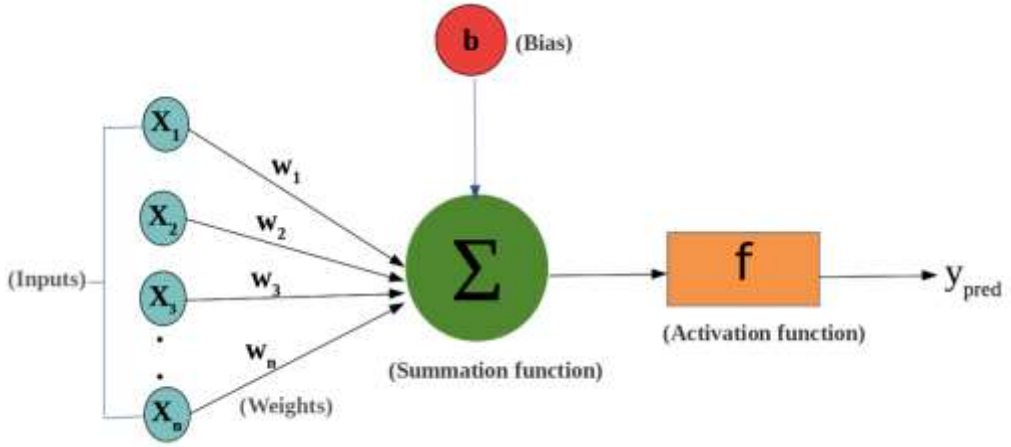


Training a neural network

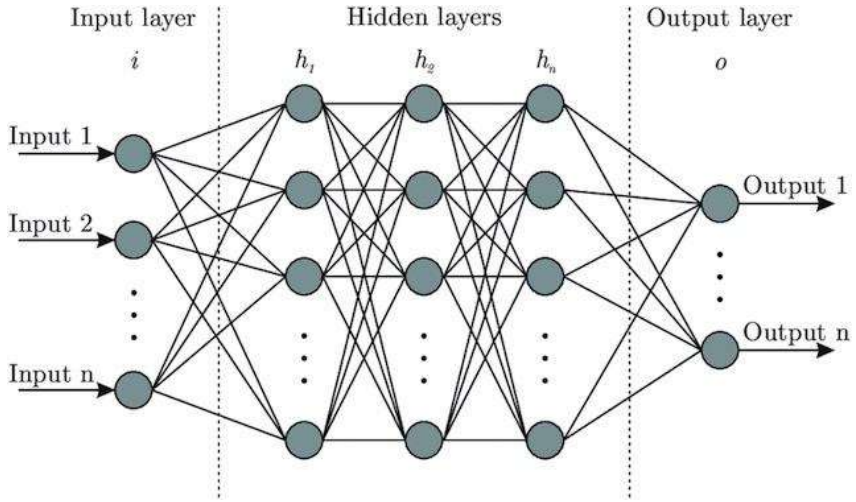
For a fixed architecture, a neural network is a function parameterized by its weights

- Given
  - A network architecture (layout of neurons, their connectivity and activations)
  - A dataset of labeled examples

- The goal: Learn the weights of the neural network

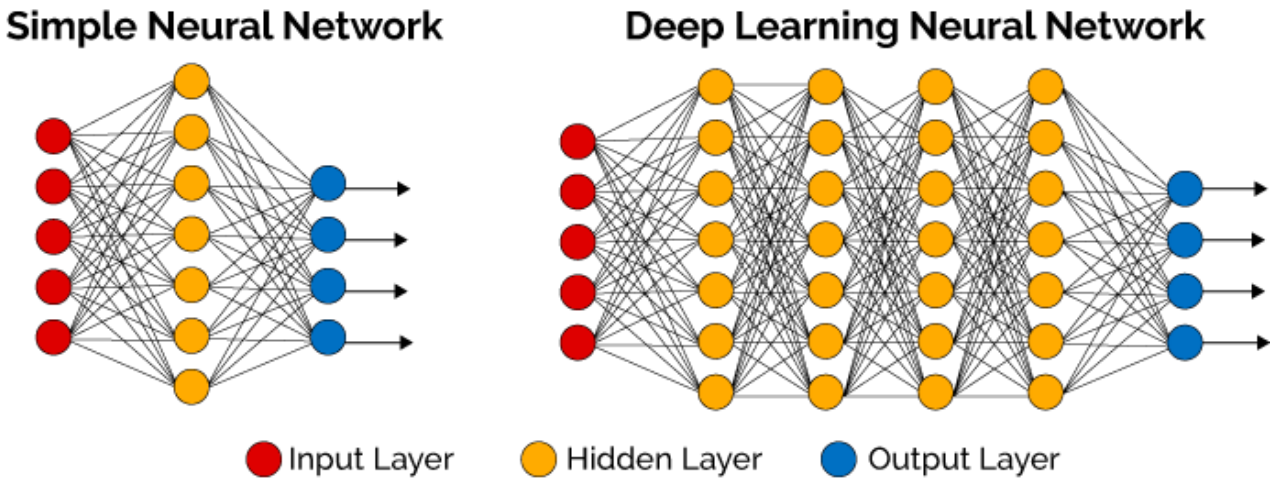


Perceptron - a single layer neural network

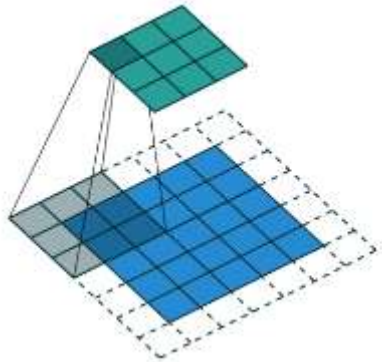


Neural Network - multilayer perceptron

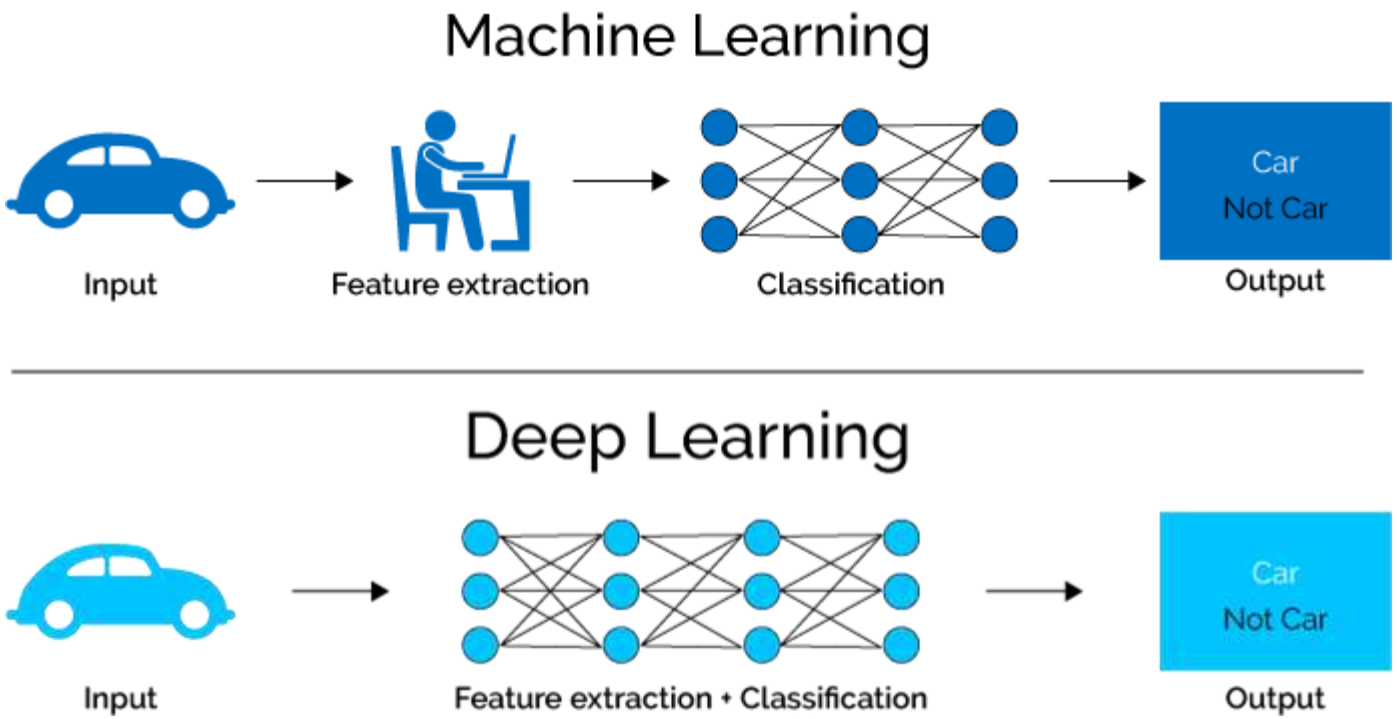
NN vs. DNN vs. DCNN



Convolution



Why Deep Learning?

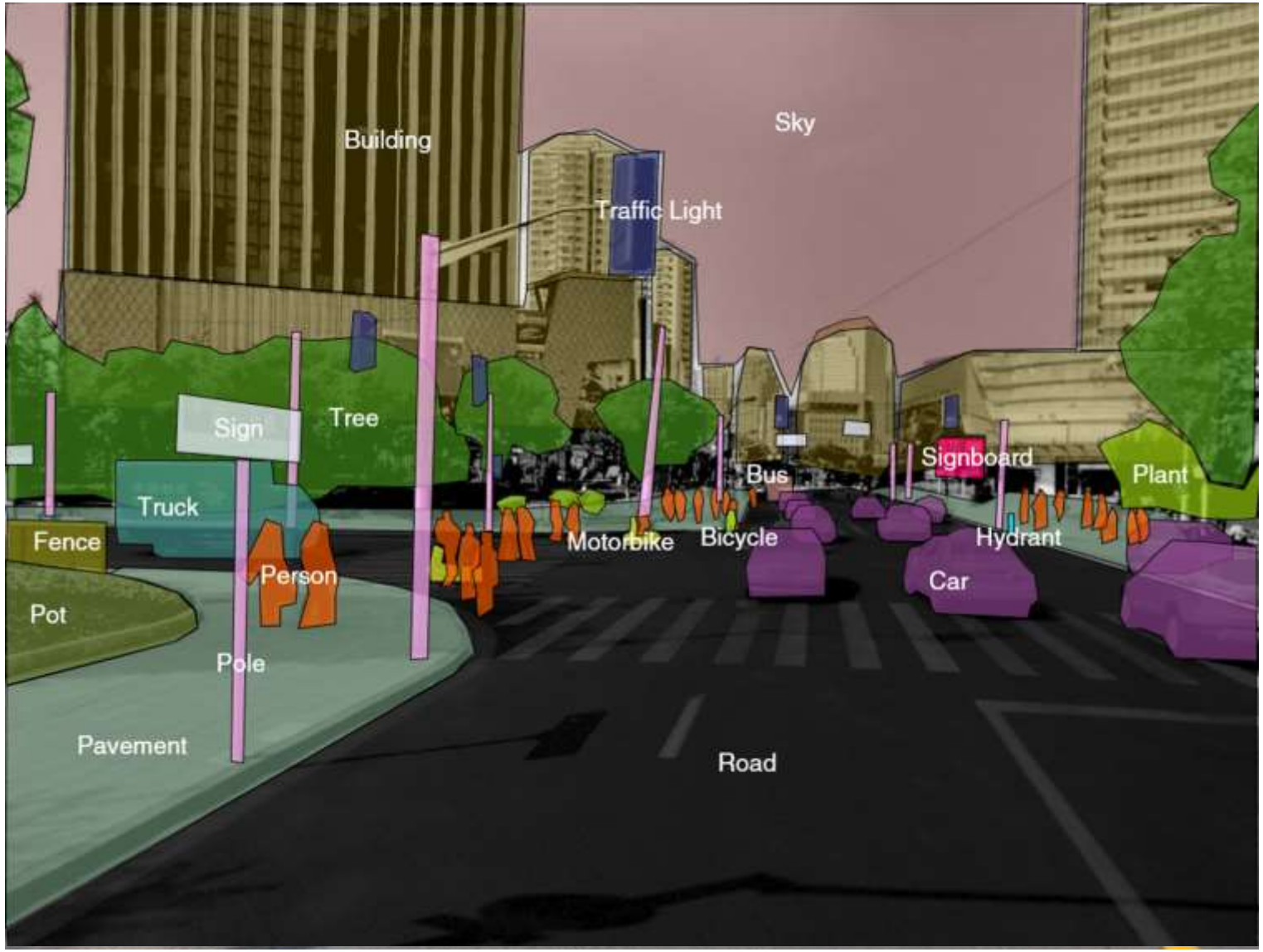


Computer Vision Applications

- Image Object Detection

Computer Vision Applications

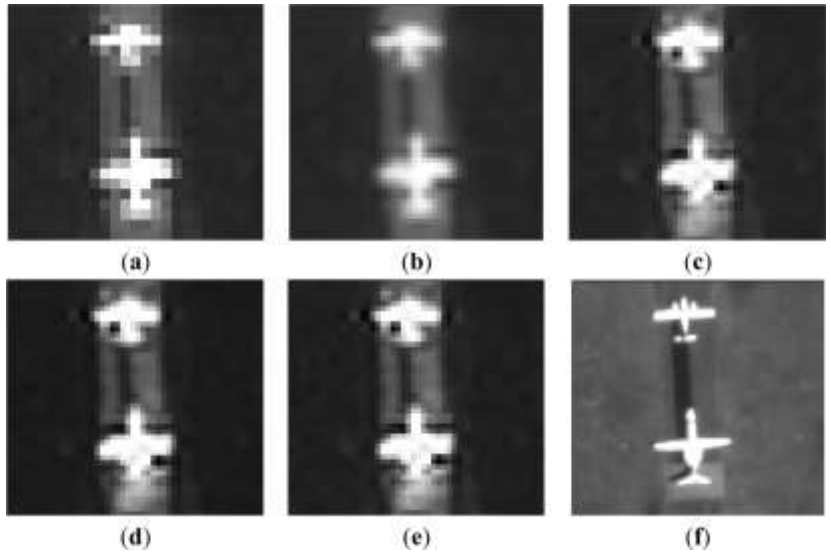
- Image Segmentation



Computer Vision Applications

❑ Image Super Resolution

- Reconstruction of the high resolution imagery from the observed low resolution image
- Applications: surveillance video, medical imaging, satellite imagery



Computer Vision Applications

❑ Style Transferring

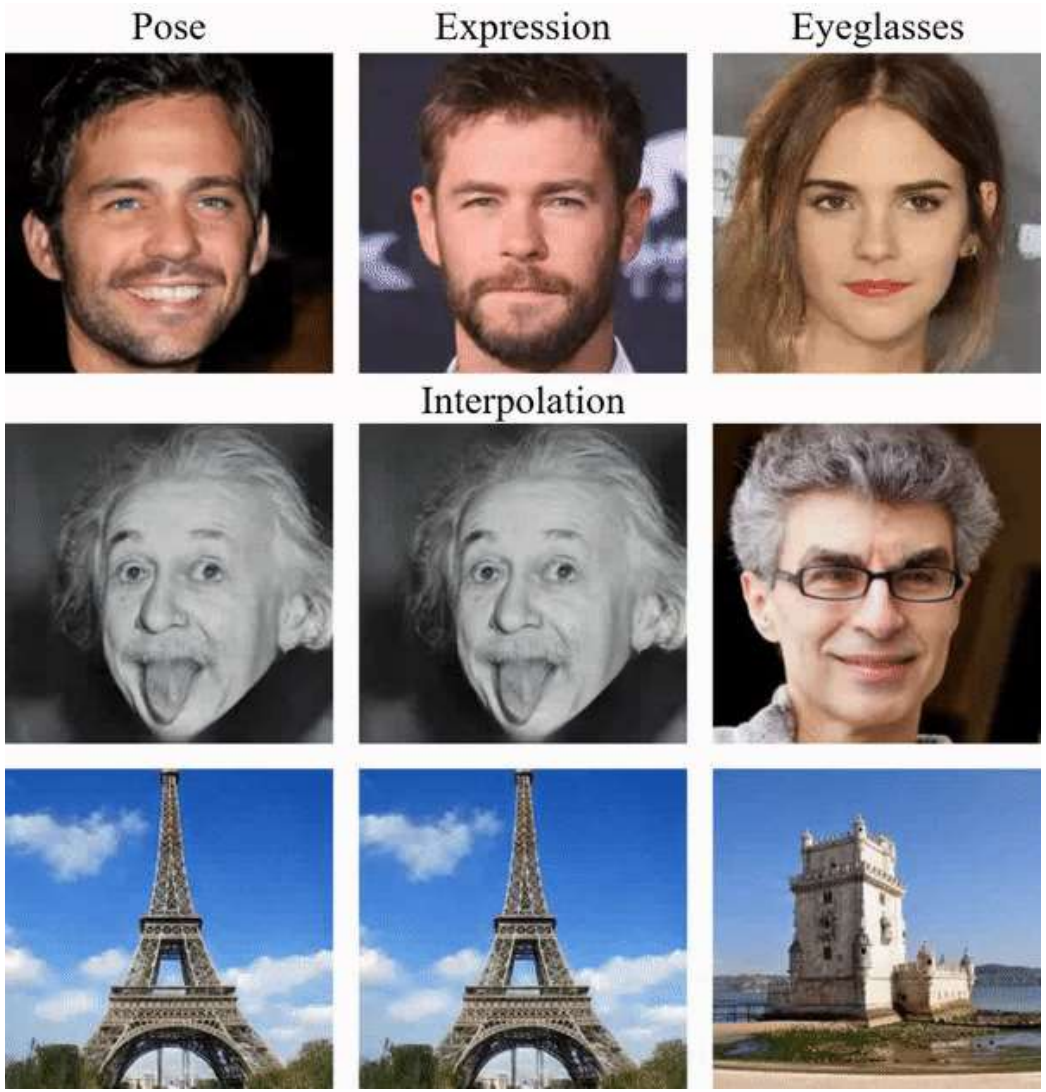
Style Extraction

Content Extraction

Style-content synthesis

Computer Vision Applications

- Style Transferring for Image Generation



Computer Vision Applications

- ❑ Image Captioning
- ❑ Text-to-image Generation

A computer screen with a Windows message about Microsoft license terms.

A can of green beans is sitting on a counter in a kitchen.

A photo taken from a residential street in front of some homes with a stormy sky above.

A blue sky with fluffy clouds, taken from a car while driving on the highway.

A hand holds up a can of Coors Light in front of an outdoor scene with a dog on a porch.

A digital thermometer resting on a wooden table, showing 38.5 degrees Celsius.

A Winnie The Pooh character high chair with a can of Yoohoo sitting on it in front of a white wall.

A cup holder in a car holding loose change from Canada.

https://vizwiz.org/tasks-and-datasets/image-captioning/

Learn more about ML/DL/CV?

- 6.036 Introduction to Machine Learning
- 6.S191 Deep Learning
- 6.801 Machine Vision
- 6.819 Advances in Computer Vision

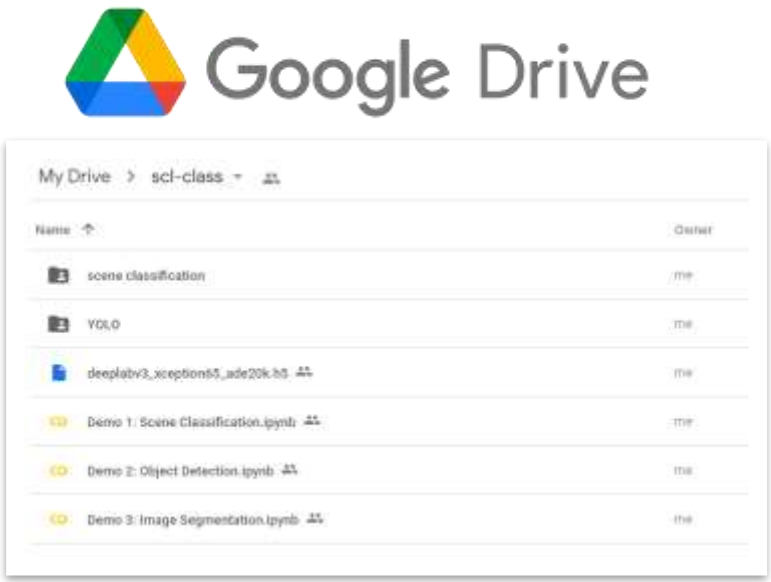
Hands-on session

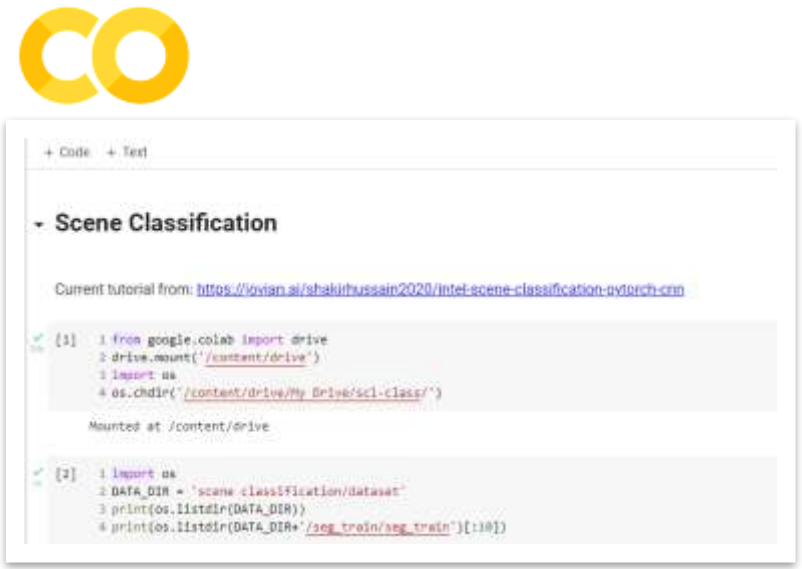
IMAGE CLASSIFICATION

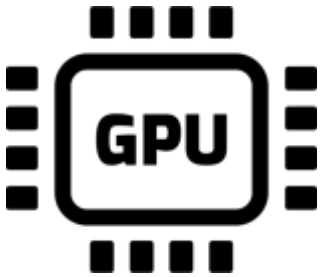
OBJECT DETECTION

IMAGE SEGMENTATION

Google Colab







Demo 1: Image Classification

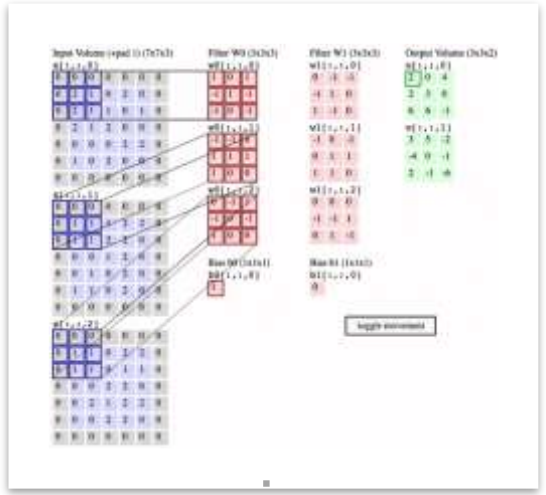
steel-frame

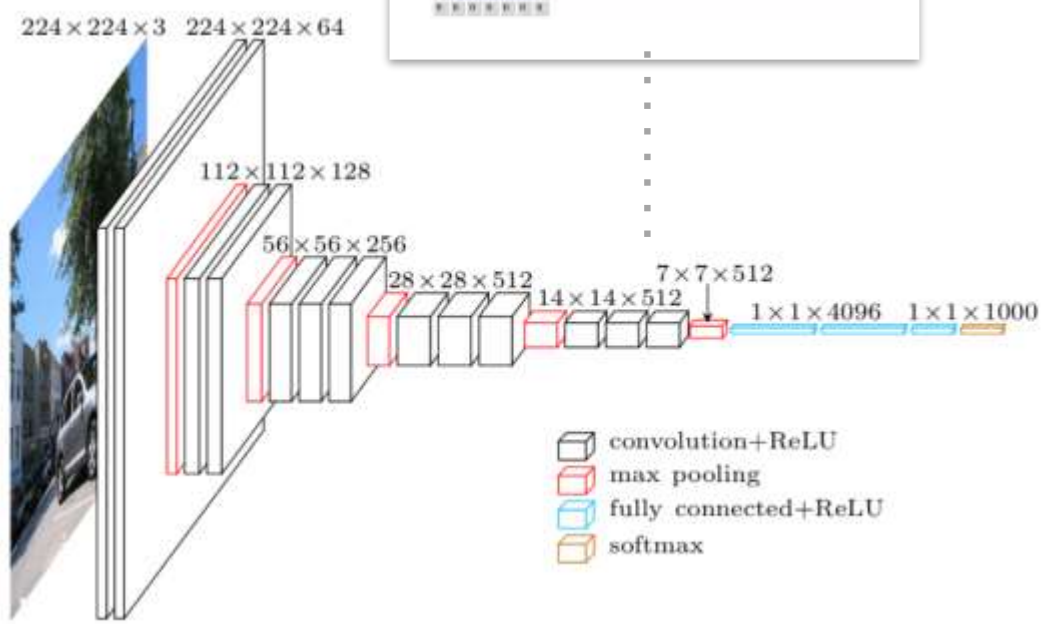
concrete

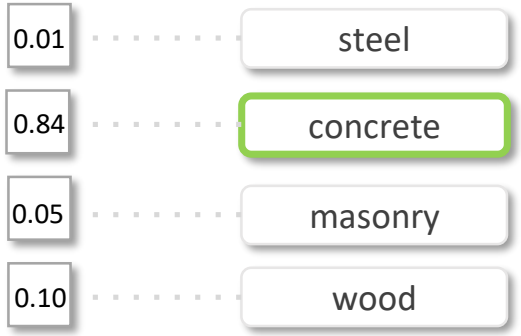
masonry

wood

concrete?

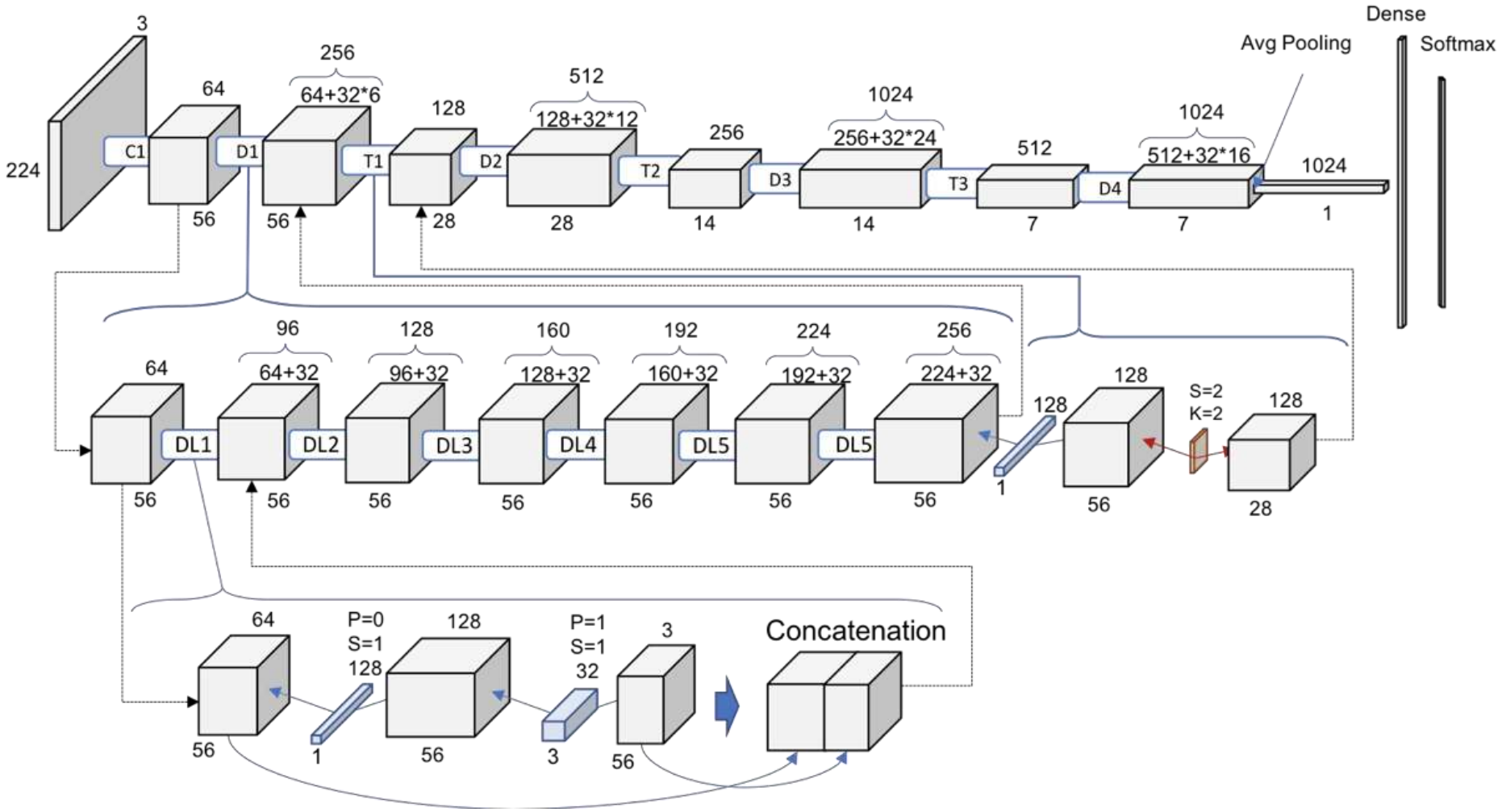






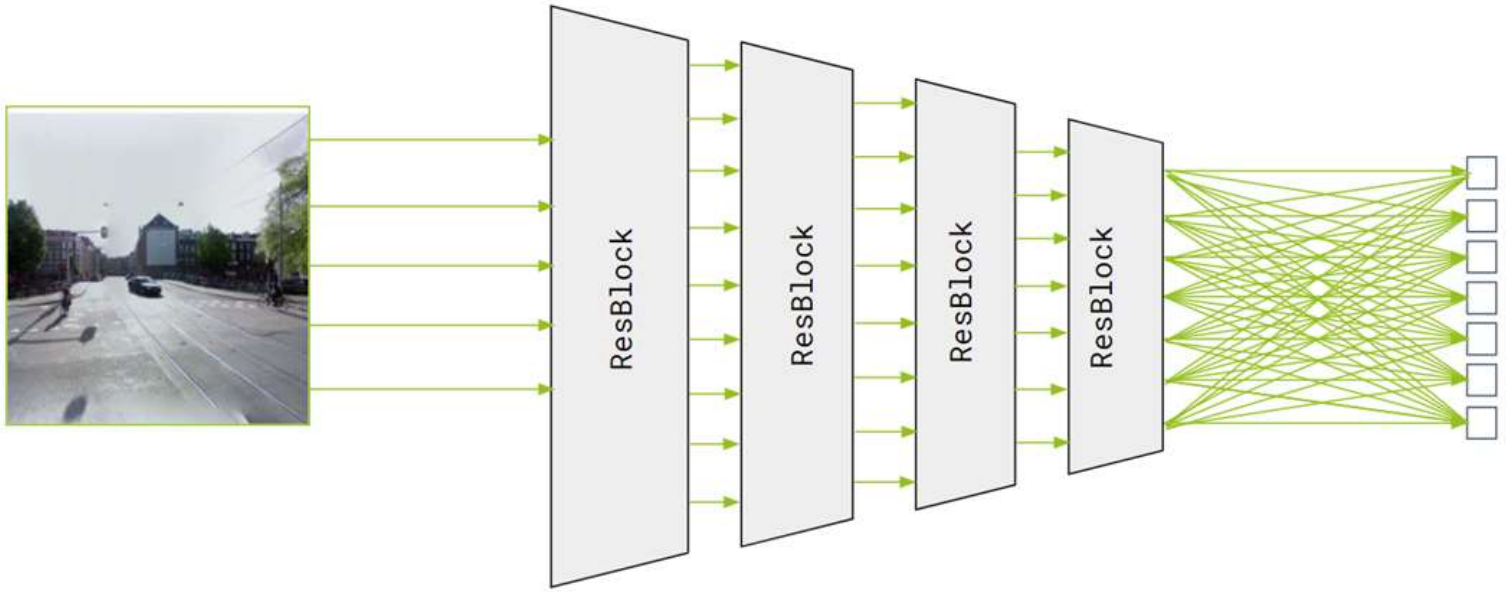
- Input: 224x224 images
- Output: 5x1 label vector

ResNet



ResNet 121 architecture

ResNet



Training Framework

PyTorch TensorFlow mxnet

Data Loader

Load images and labels from file and split them into batches

Model

Neural Network with weights ready to be trained

Train Loop

```
for epoch in range(NUM_EPOCHS):  # loop over the dataset multiple times

    running_loss = 0.0
    for i, data in enumerate(trainloader, 0):
        inputs, labels = data # get a batch from the dataloader
        optimizer.zero_grad()
        outputs = model(inputs) # compute labels
        loss = criterion(outputs, labels) # compute loss
        loss.backward() # update weights
        optimizer.step() # adjust learning rate

        running_loss += loss.item()
```

Loss Function

$$J = \frac{1}{2m}\sum_{i=1}^{m}(\hat{y} - y)^2$$

Difference between predicted and target value

Optimizer

Slow down or speed up learning rate

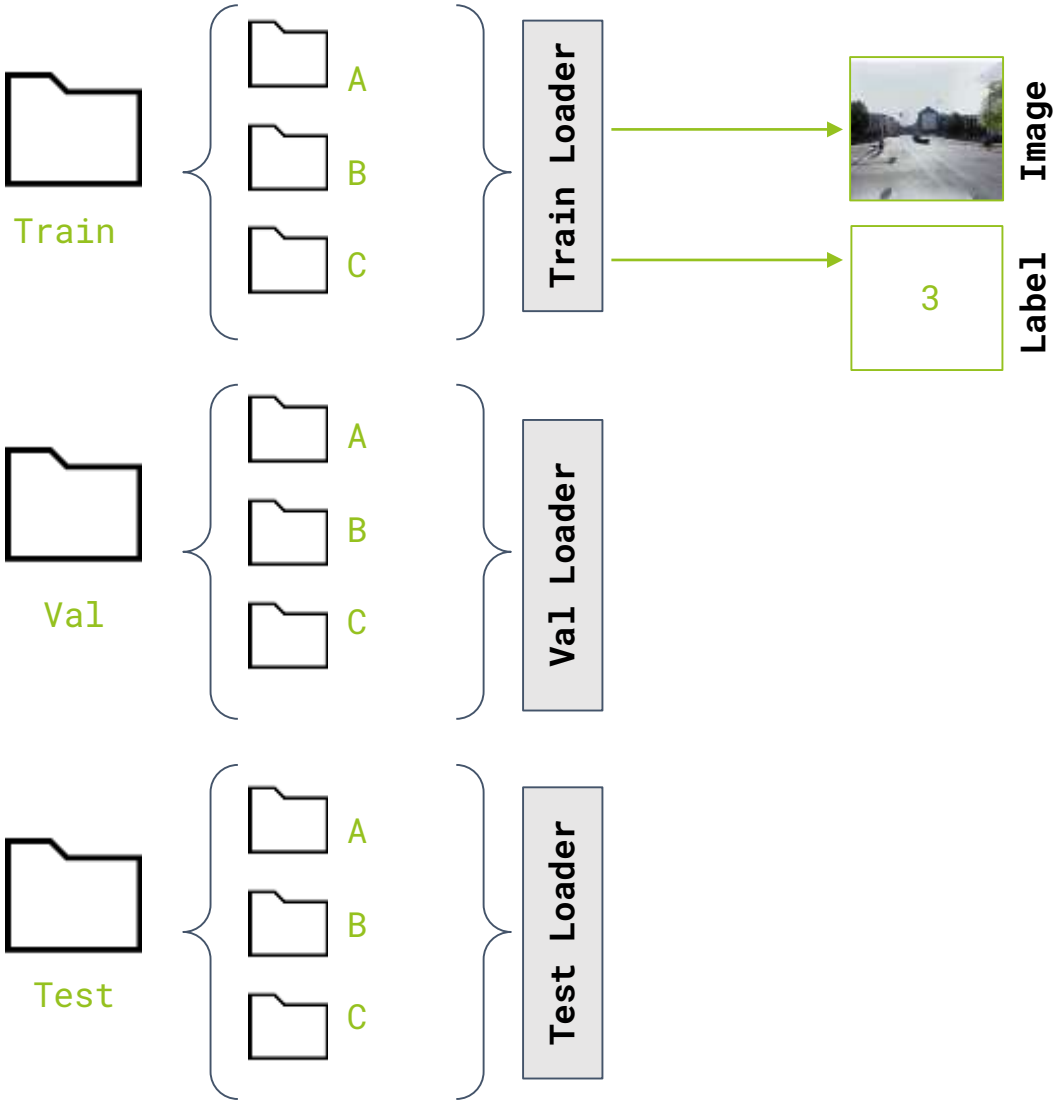
Backpropagation

Find weights in the model and update them

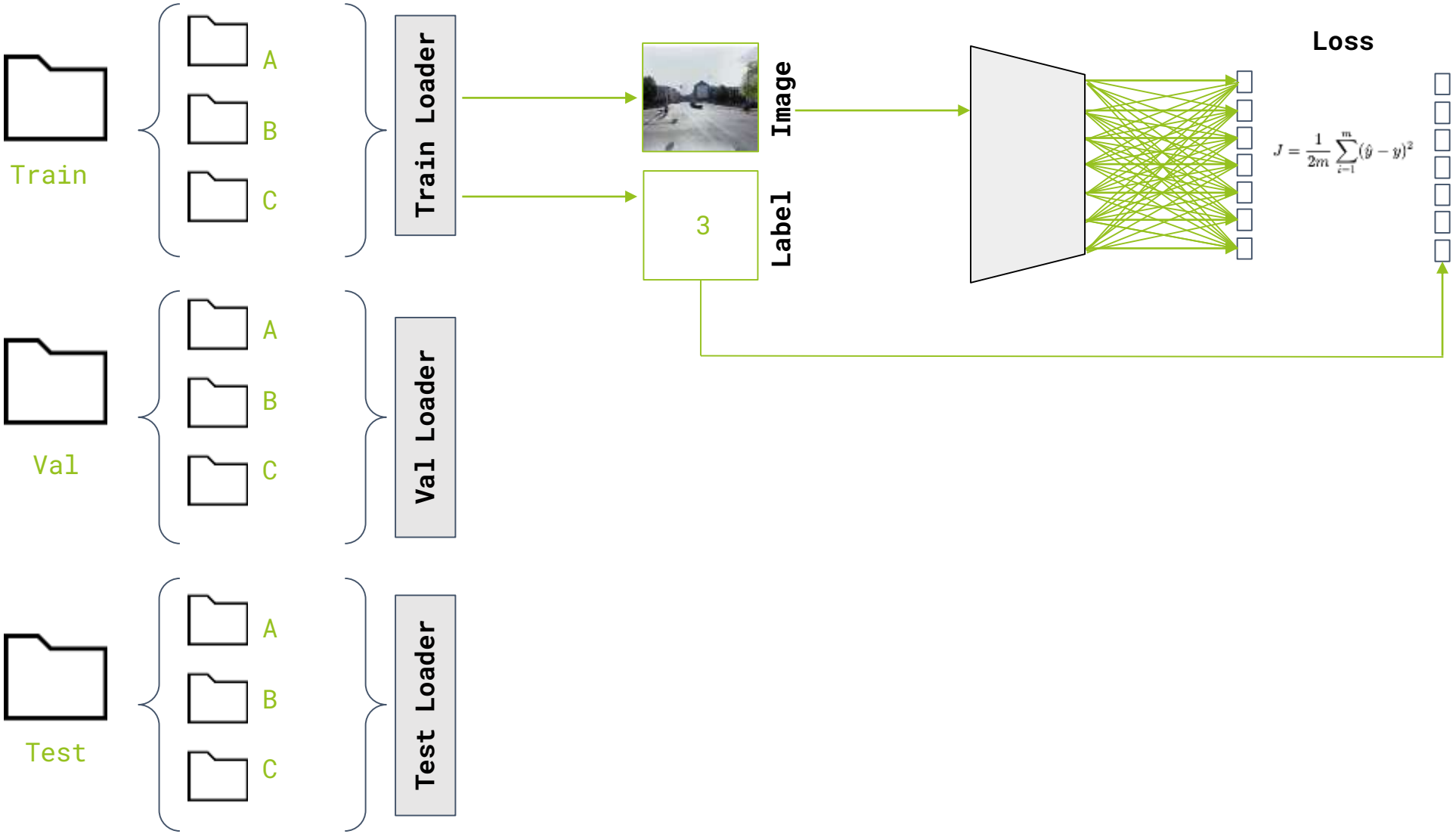
Manage GPU Usage

```
device = torch.device('cuda'
model = model.to(device)
```

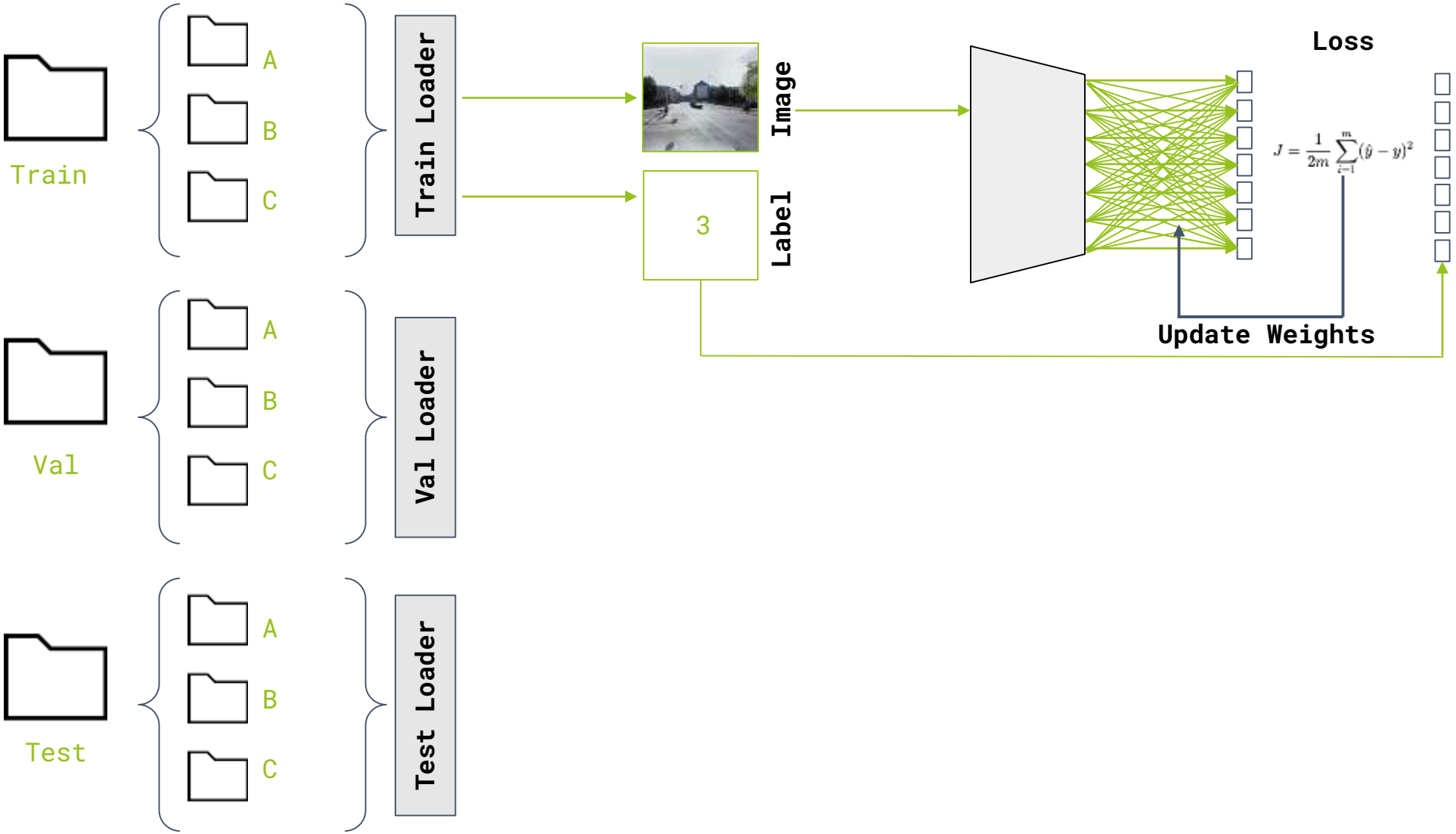
Training Workflow



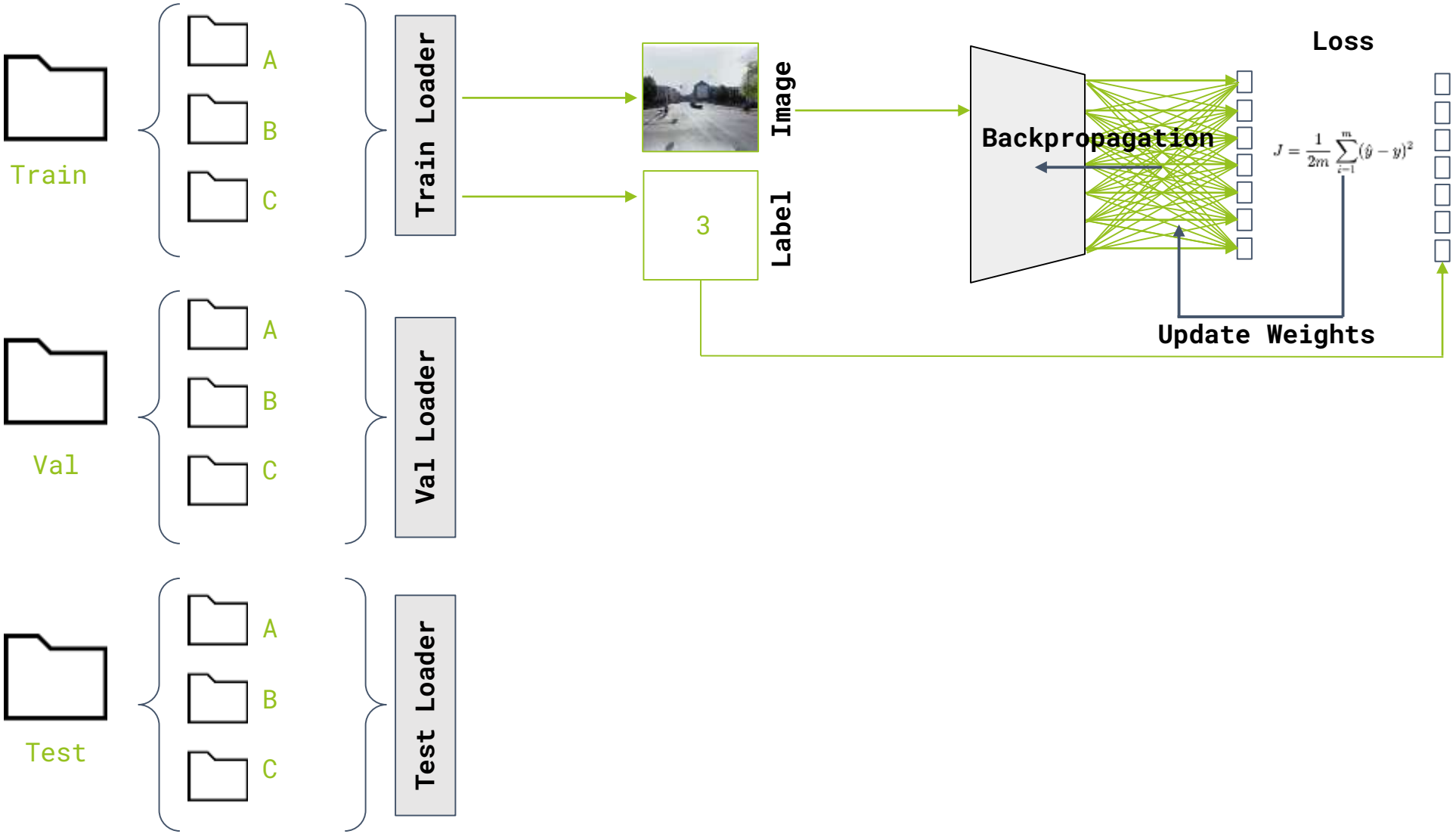
Training Workflow



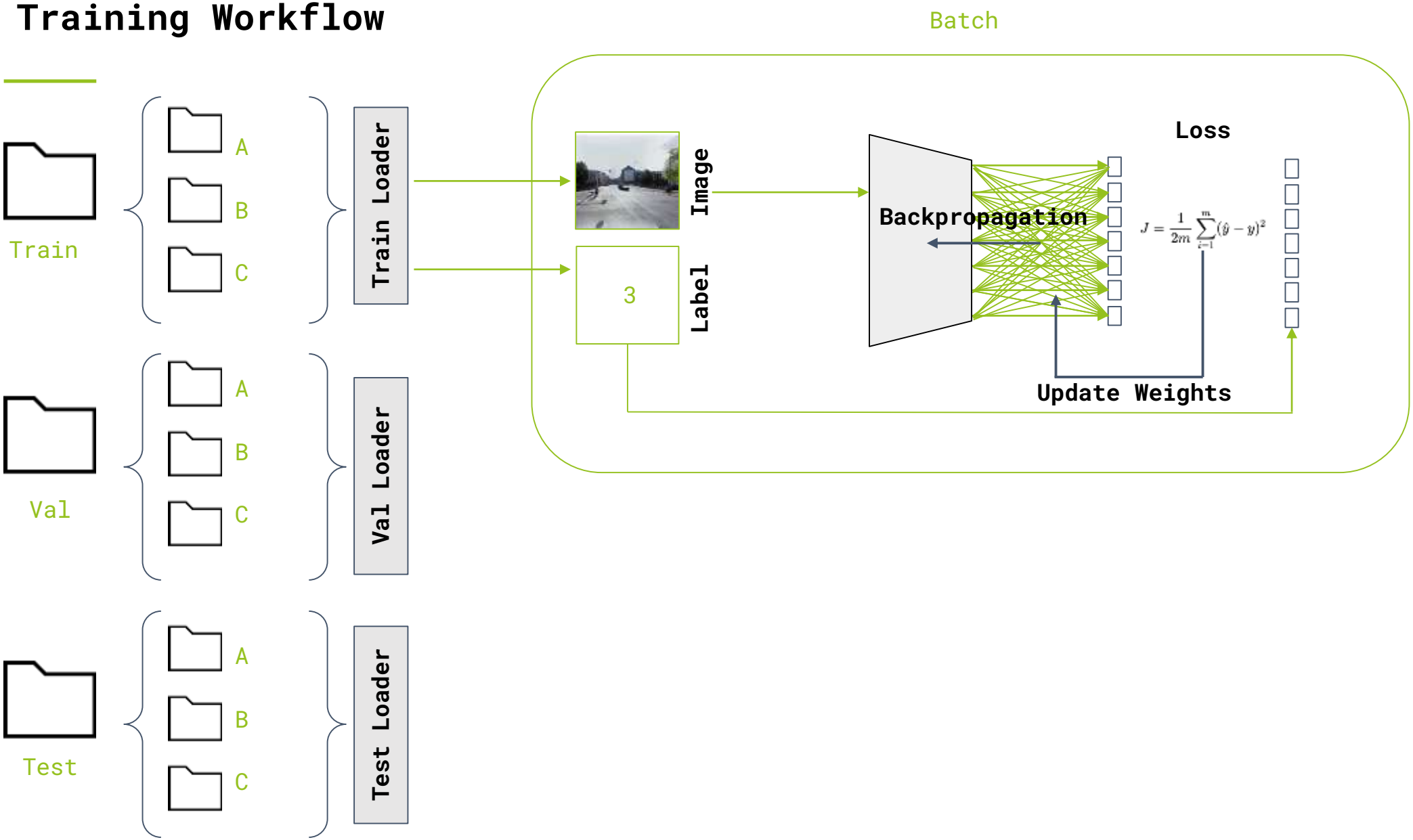
Training Workflow



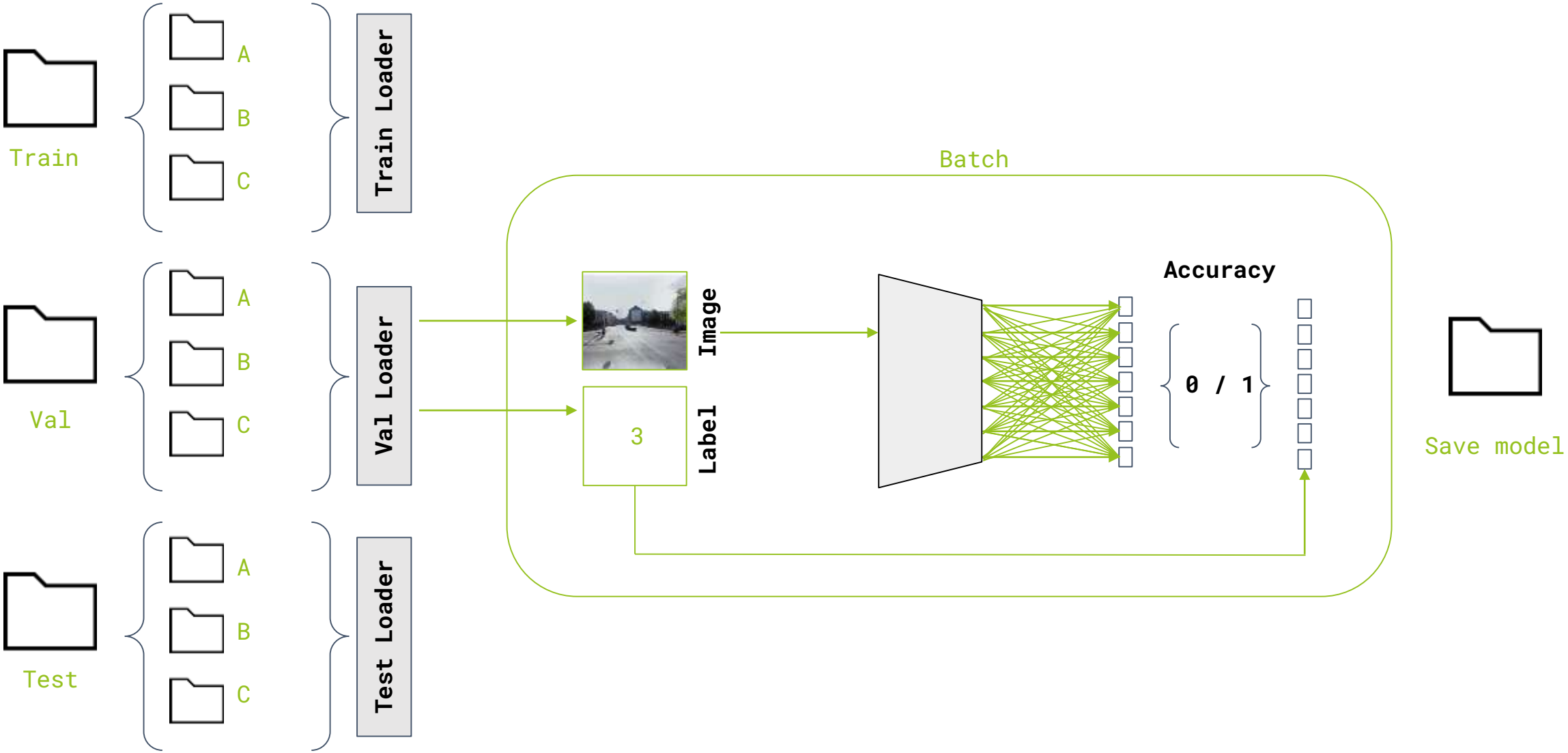
Training Workflow



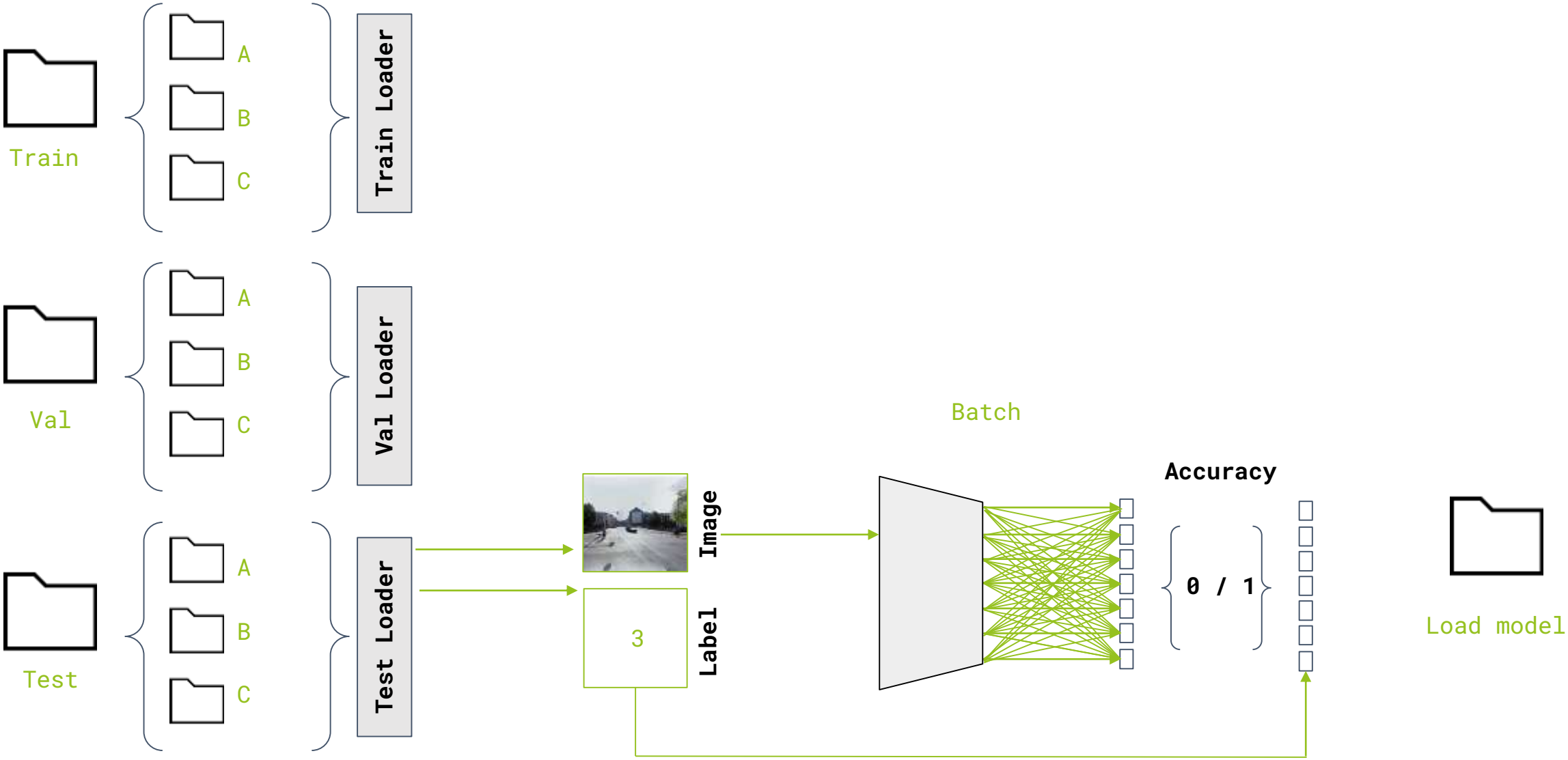
Training Workflow



Training Workflow



Training Workflow



Demo 1:Classifier

Go to Google Drive and Open Demo 1.ipynb

Demo 2: Image-to-Image Networks

- **Architecture**
  - Input and output are both images
  - Produced image compared to target image pixel-by-pixel
  - Has a narrow ‘bottleneck’ layer to encourage extraction of most important features
- **Applications**
  - Style transfer
  - Denoising
  - Image abstraction
  - Produce segmentations, masks

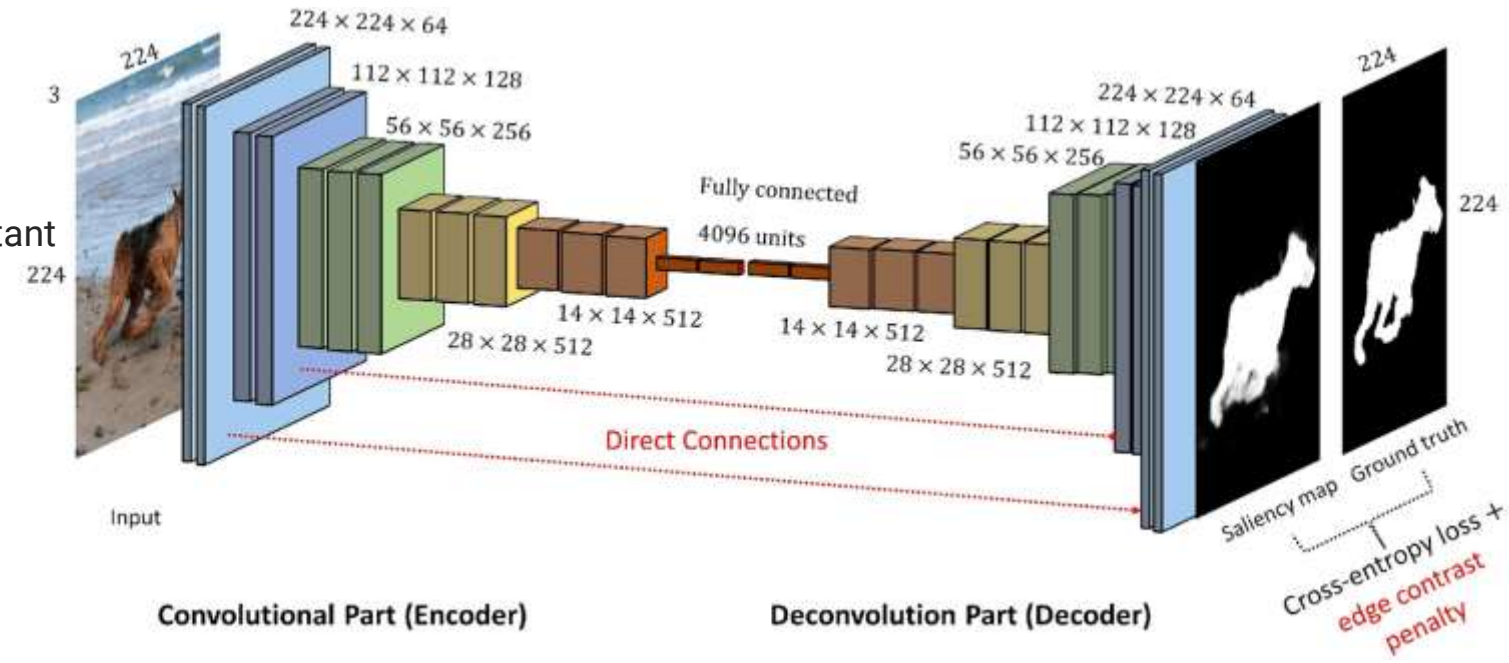
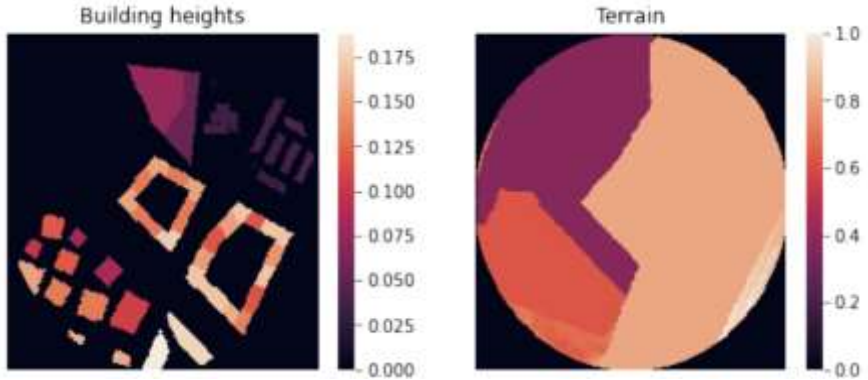
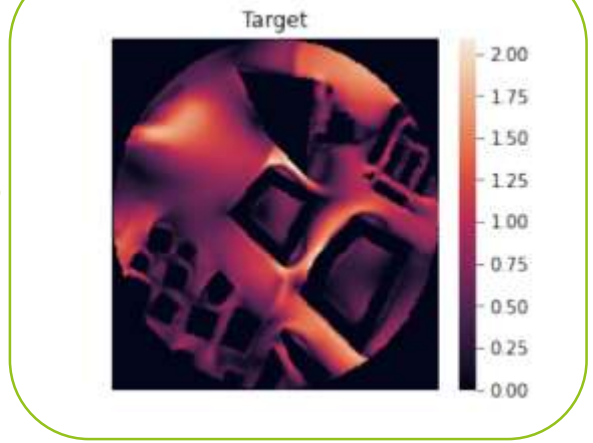


Image-to-Image Networks: Wind Heatmaps

Input



Target



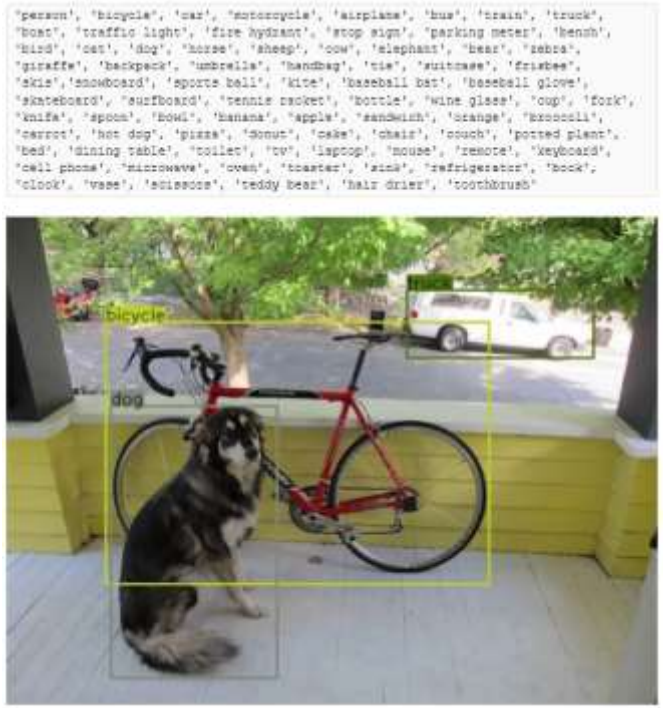
Network Outputs

Demo 2:Image->Image

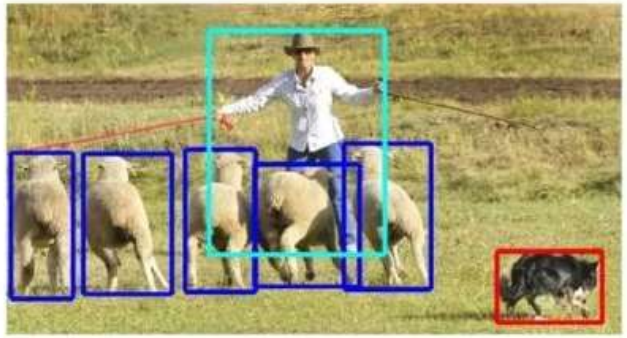
Go to Google Drive and Open Demo 2.ipynb

Demo 3: Object Detection

- **COCO Dataset**
  - 330 000 images
  - 91 object categories
  - 250 000 people



1) Image Classification

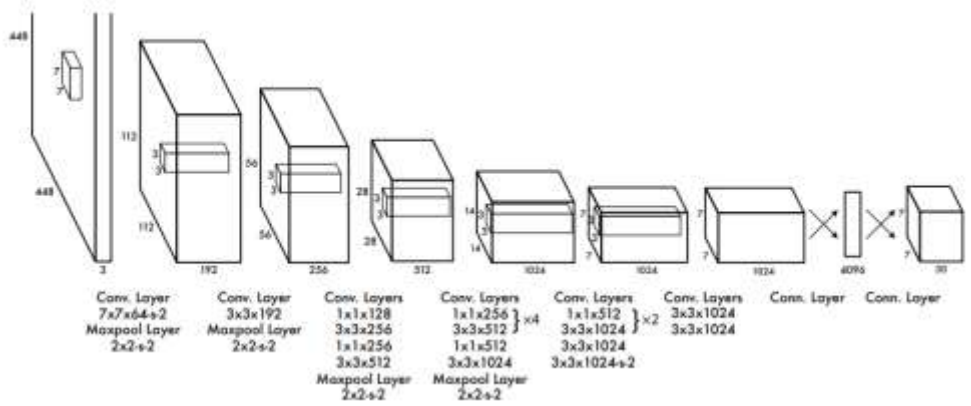


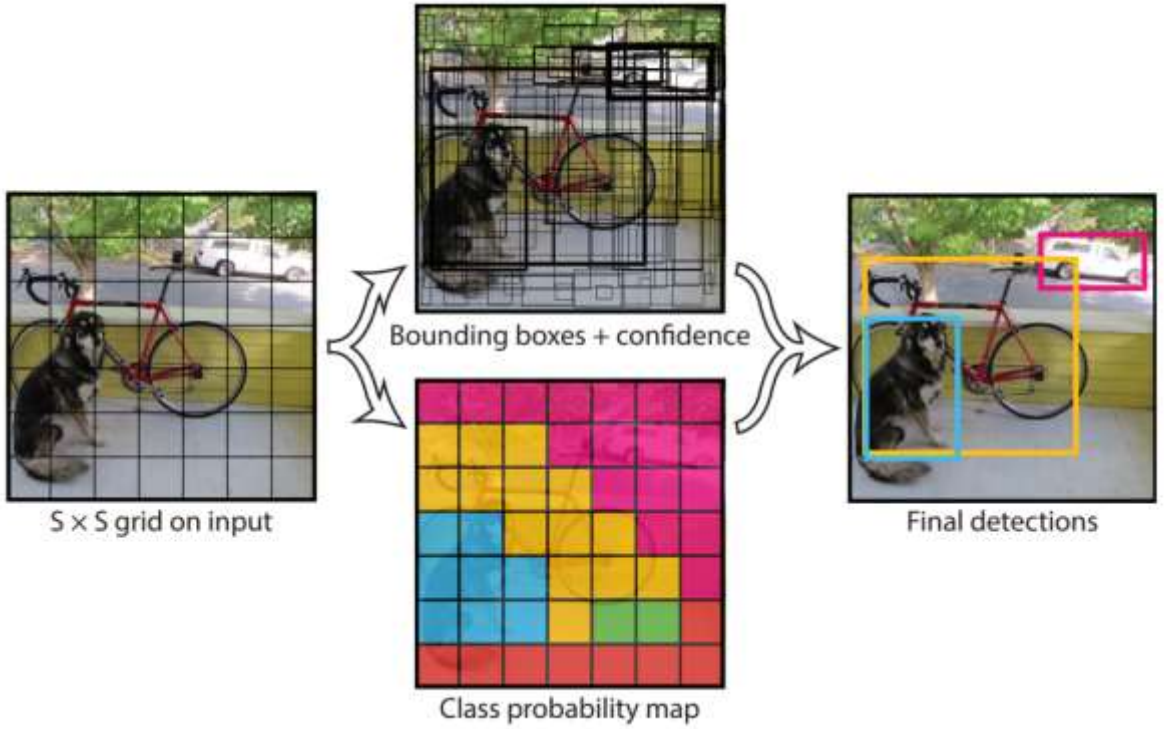
2) Object Localization

3) Semantic Segmentation

Demo 3: Object Detection

- **YOLO – 'You Only Look Once'**
  Convolutional Neural Network
  Input: 224x224 image
  Output: 7x7x30 tensor



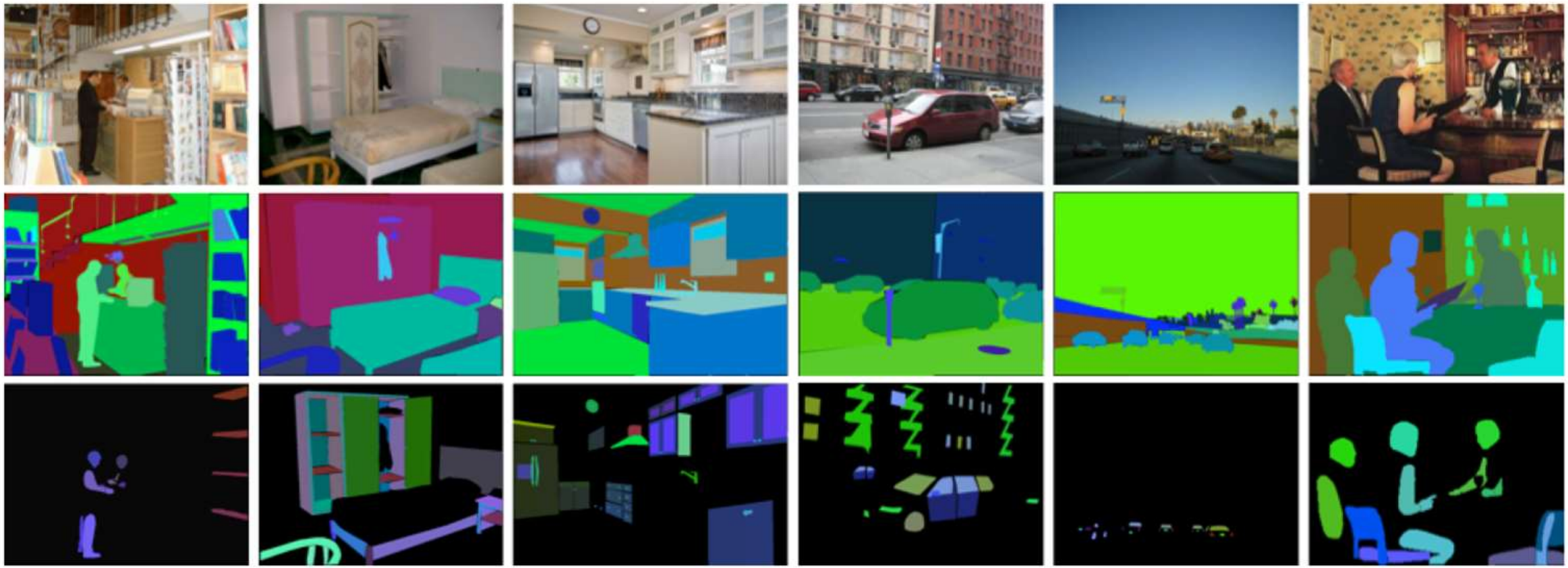


Demo 3: Detection

Go to Google Drive and Open Demo 3.ipynb

Demo 4: Image Segmentation

- **ADE20k Dataset**
  - 27 000 Images, 3000 Object Categories, 150 Semantic categories, 193,238 annotated objects, polygon annotations
  - By MIT CSAIL



Demo 4: Segmentation

Go to Google Drive and Open Demo 4.ipynb

In-Class assignment

GSV Downloading and Image Segmentation

In-class Practice

(In Google CoLab)
1. **Download GSV** of a small area (or use images from the last GSV downloading assignment);

2. Using **image segmentation** to process these images, show the data structure of the mask of one image sample;

3. Calculate the ratios of some objects (tree, cars, etc.), and **aggregate the results from all images** together. The final results could be, for example, the greenery ratio of all images of Mass. Ave., Cambridge.