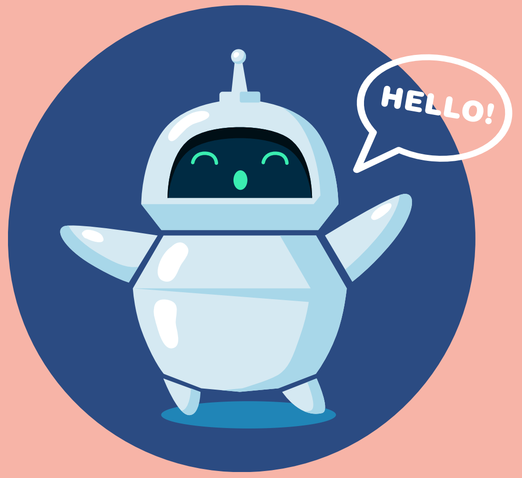
A RESEARCH PRESENTATION

# Pothole Severity Classification using Deep Learning Based Image Segmentation Methods

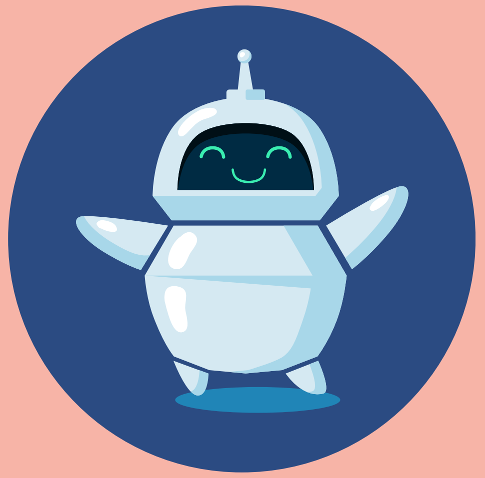
# The Team




**Ahmed Abul Hasanaath**
Robotics Society President

**Abdul Sami Mohammed**
Friend and Cool Dude

**Lara Alotaibi**
Robotics Society President



**Maria Alabdulrahman**
Robotics Society Vice President

# Presenter

**???**

# Agenda

**1**

**STEP**

The motivation to conduct this study

**2**

**STEP**

Review deep learning fundamentals

**3**

**STEP**

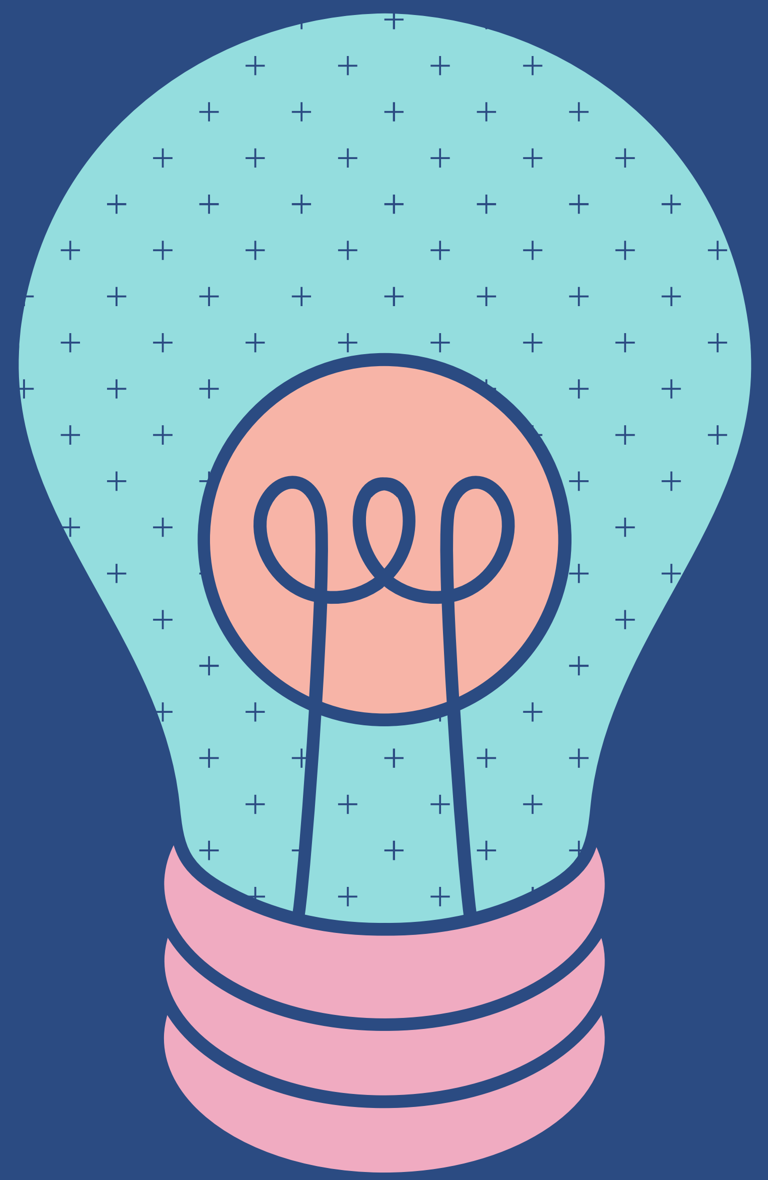
Discuss our PSC approach and results

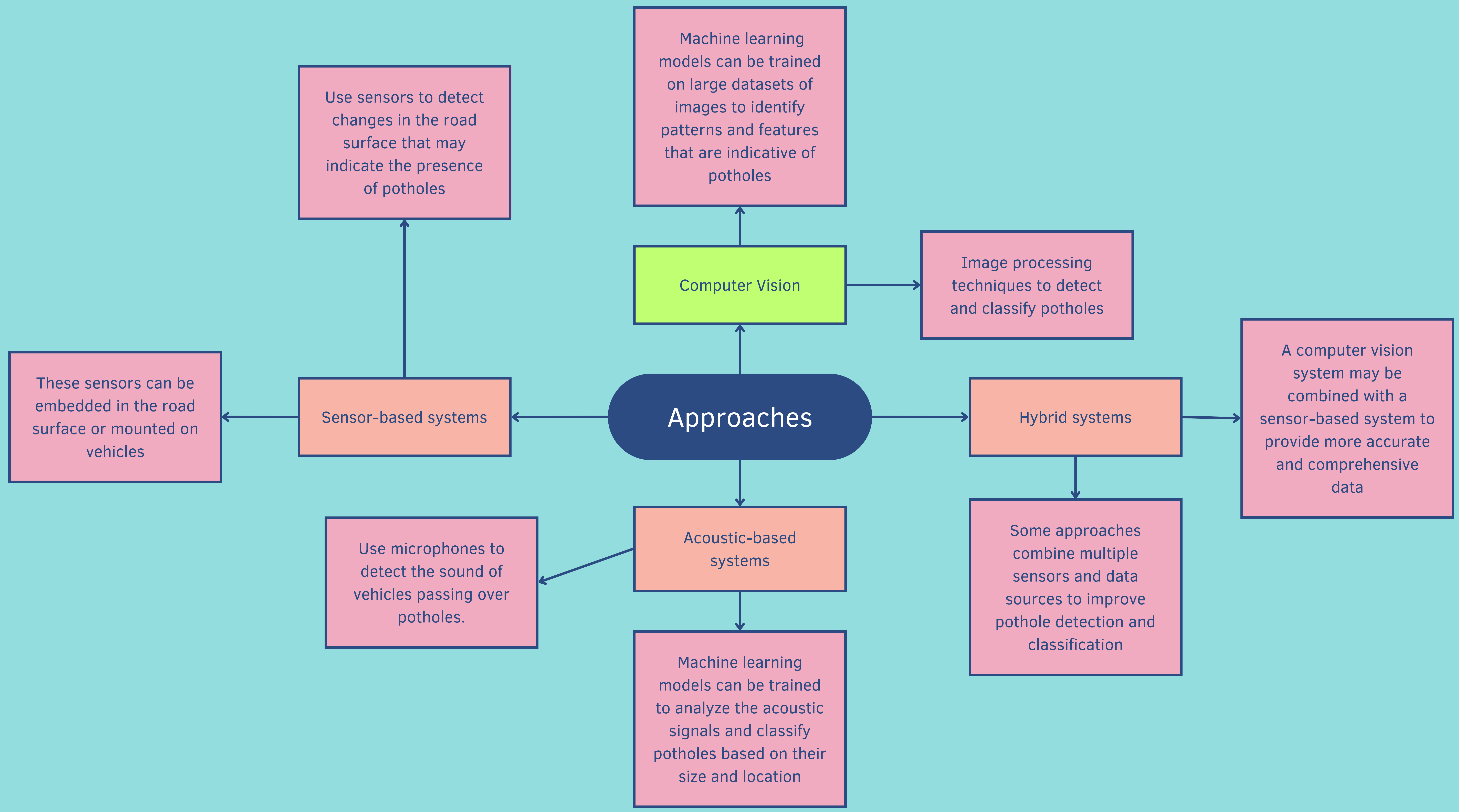
# MOTIVATION

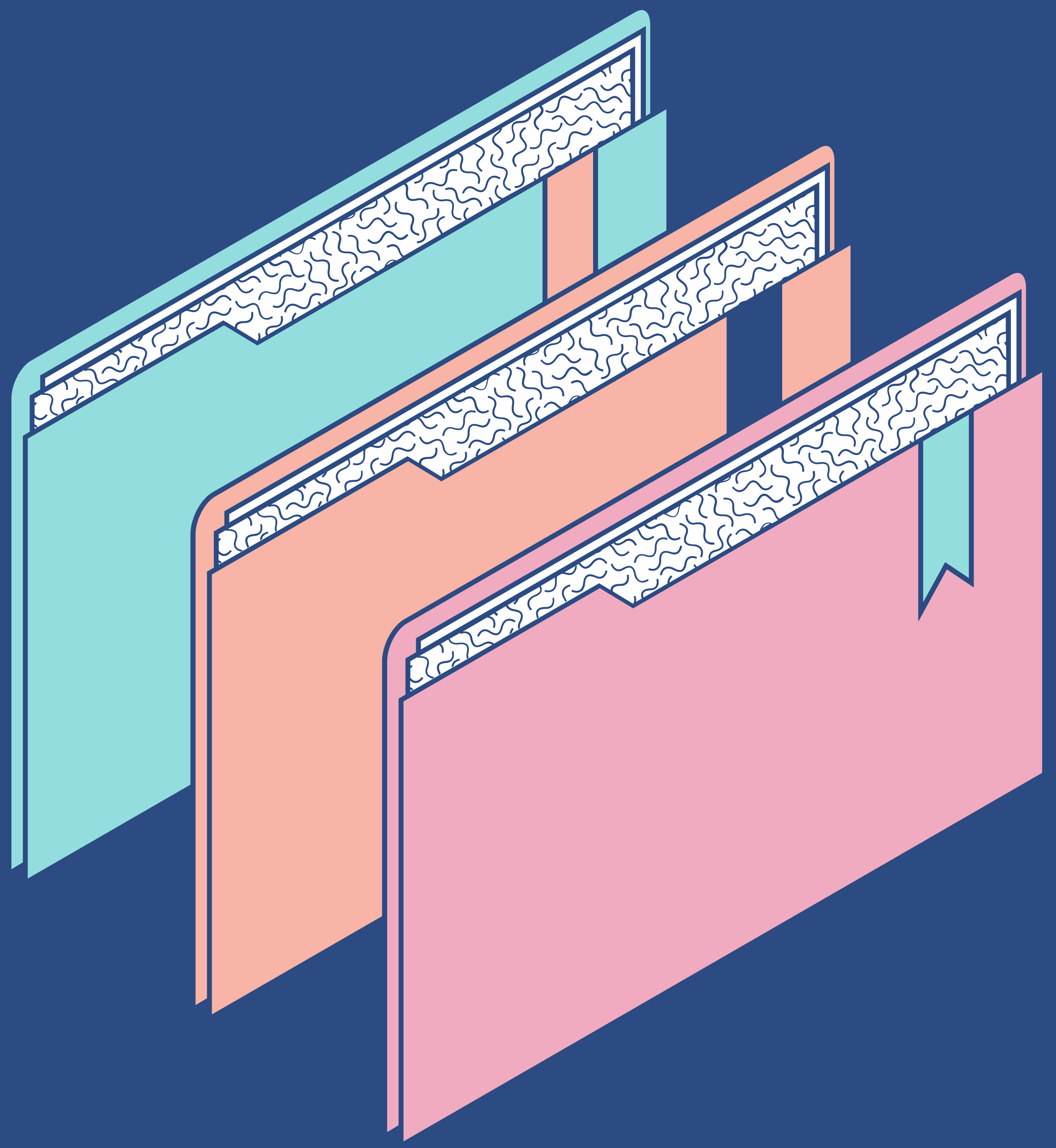
- SAFETY: POTHOLES CAN BE HAZARDOUS TO DRIVERS, PEDESTRIANS, AND CYCLISTS AS THEY CAN CAUSE ACCIDENTS AND INJURIES. IN EUROPE, 57% OF ROAD FATALITIES WERE CAUSED BY POOR ROAD CONDITIONS (2010)

- INFRASTRUCTURE MAINTENANCE: POTHOLES ARE A SIGN OF DETERIORATING ROAD INFRASTRUCTURE, AND IF LEFT UNADDRESSED, THEY CAN CAUSE FURTHER DAMAGE TO THE ROAD AND THE SURROUNDING ENVIRONMENT.

- ECONOMIC IMPACT: POTHOLES CAN CAUSE DAMAGE TO VEHICLES AND AFFECT THE TRANSPORTATION INDUSTRY, LEADING TO INCREASED COSTS FOR DRIVERS AND BUSINESSES.




Machine learning models can be trained on large datasets of images to identify patterns and features that are indicative of potholes
Use sensors to detect changes in the road surface that may indicate the presence of potholes
Computer Vision
Image processing techniques to detect and classify potholes
These sensors can be embedded in the road surface or mounted on vehicles
Sensor-based systems
Approaches
Hybrid systems
A computer vision system may be combined with a sensor-based system to provide more accurate and comprehensive data
Acoustic-based systems
Use microphones to detect the sound of vehicles passing over potholes.
Some approaches combine multiple sensors and data sources to improve pothole detection and classification
Machine learning models can be trained to analyze the acoustic signals and classify potholes based on their size and location



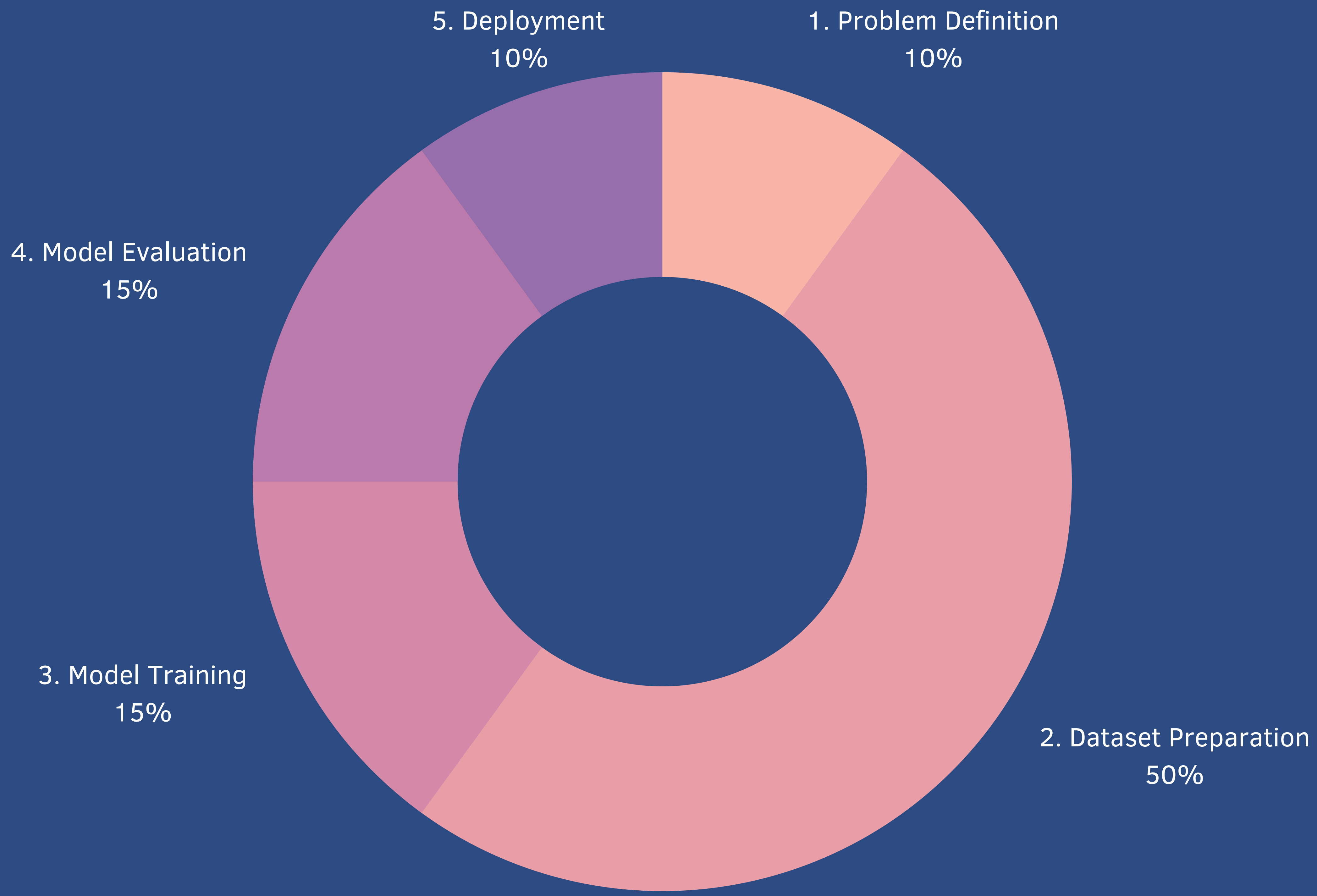
# Background

SOME FUNDAMENTALS BEFORE WE DISCUSS THE PROJECT

- 5 Steps of Machine Learning
- Deep Learning
- Convolutional Neural Networks
- Object Detection and Instance Segmentation

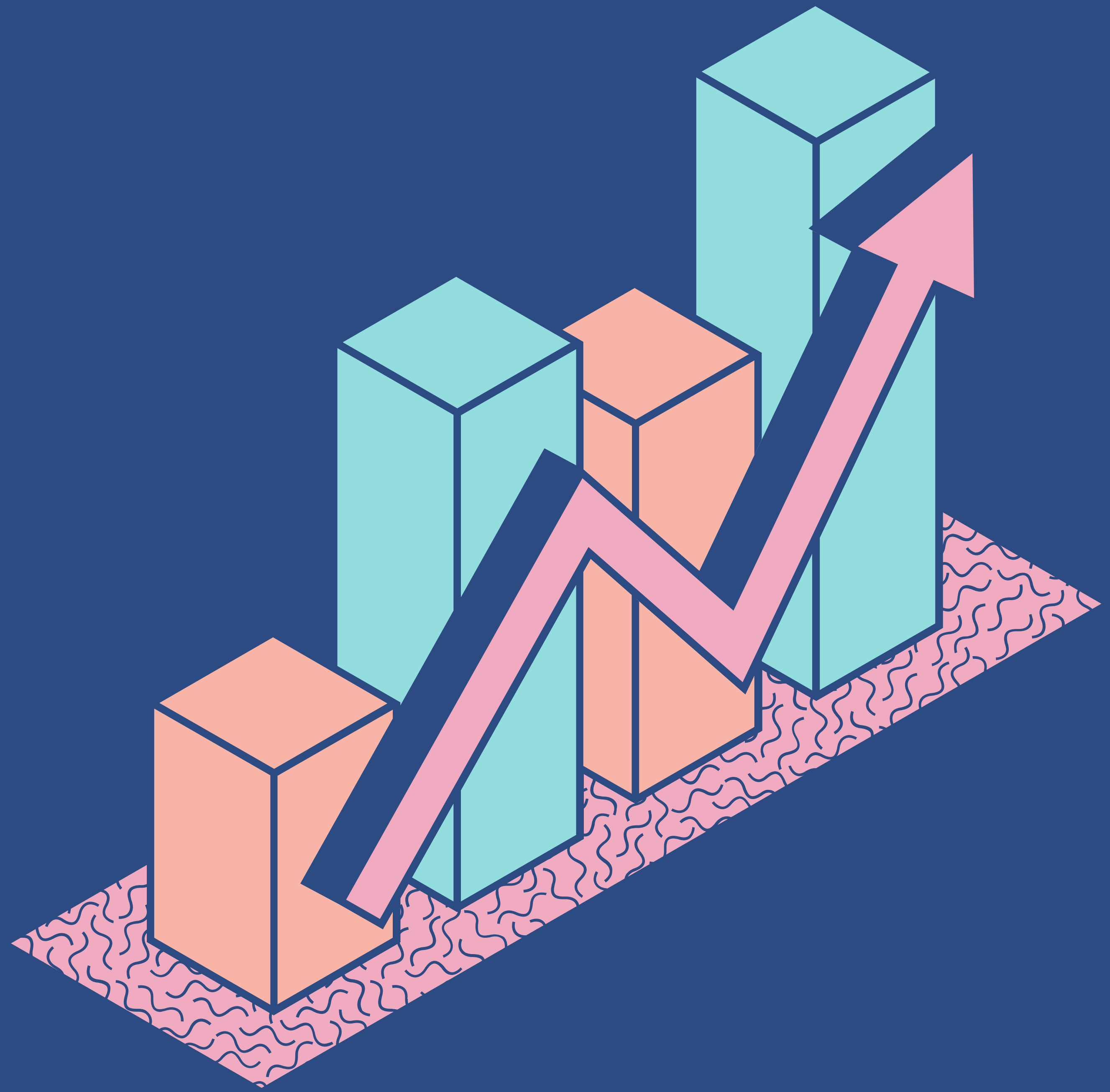
# 5 Steps of Machine Learning

THE FUNDAMENTAL STEPS COMMON BETWEEN ALL MACHINE LEARNING PROJECTS

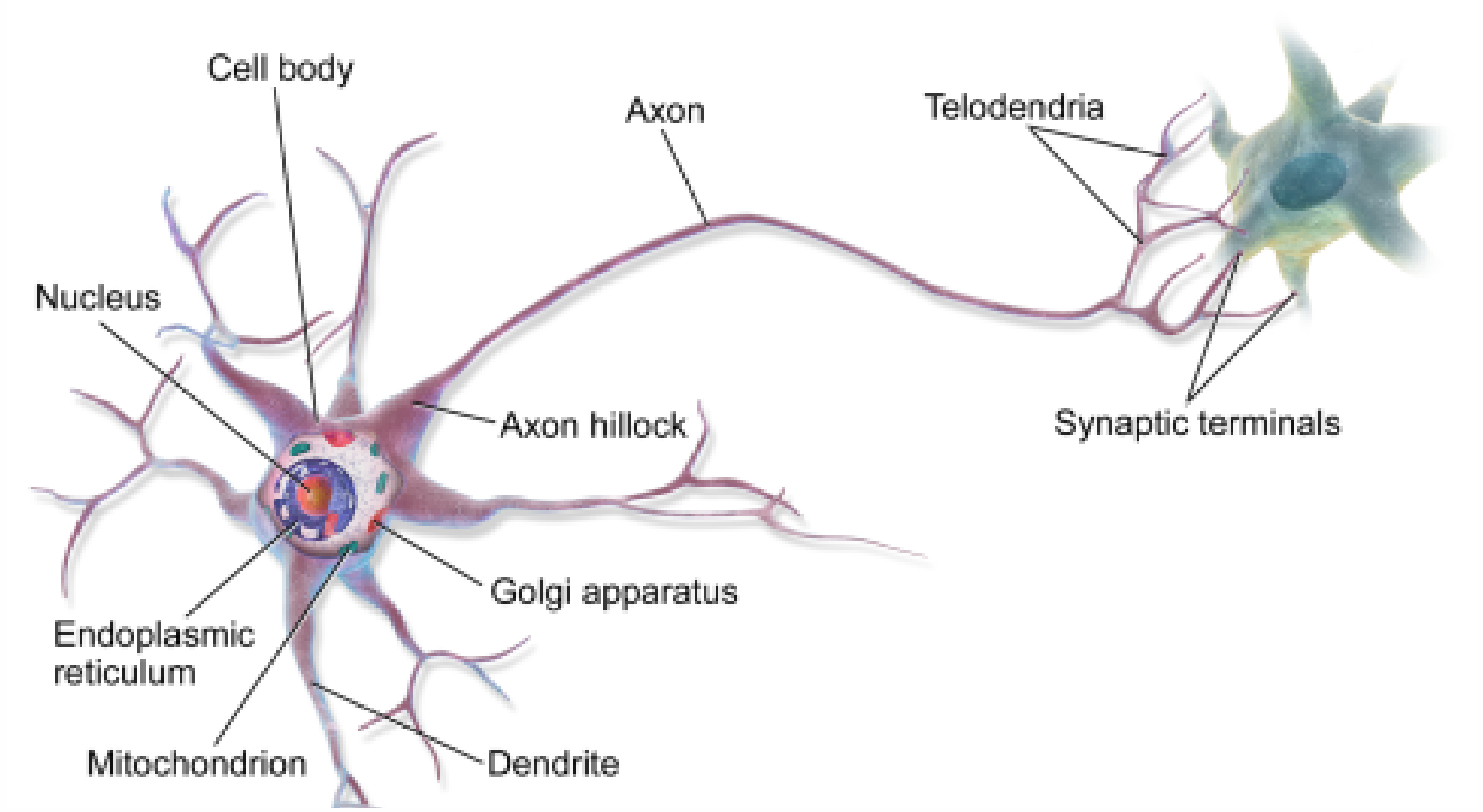


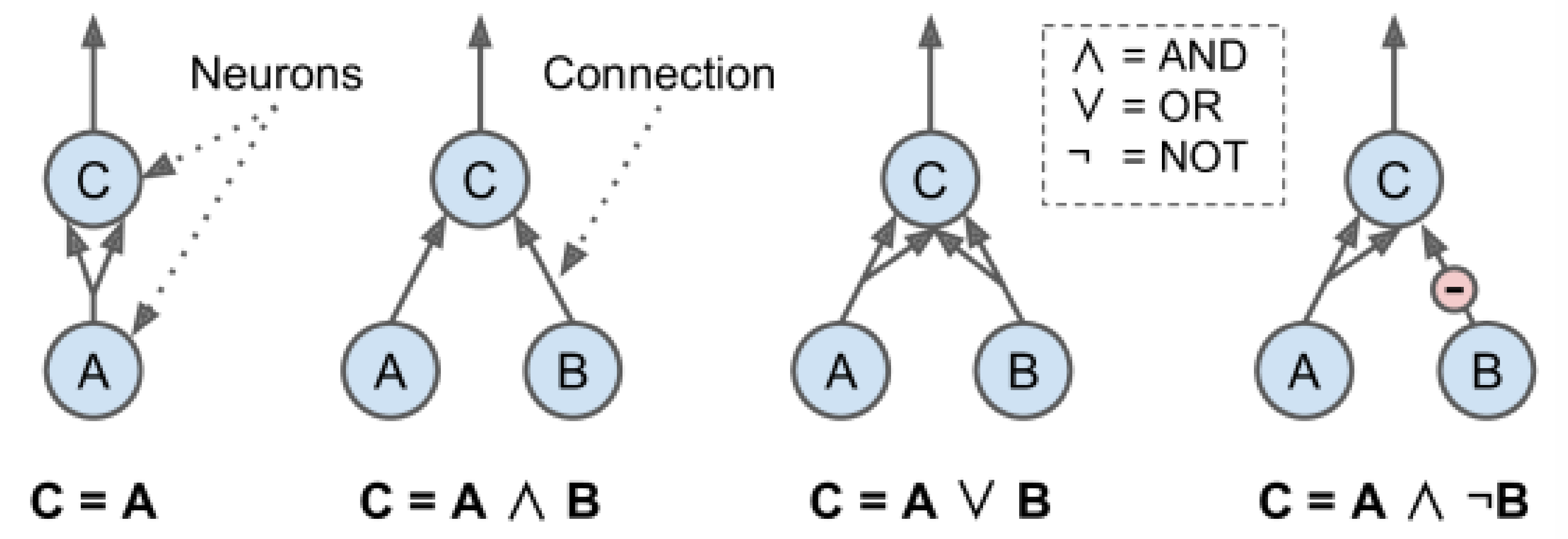
# Deep Learning

## A Primer



# Where It All Started



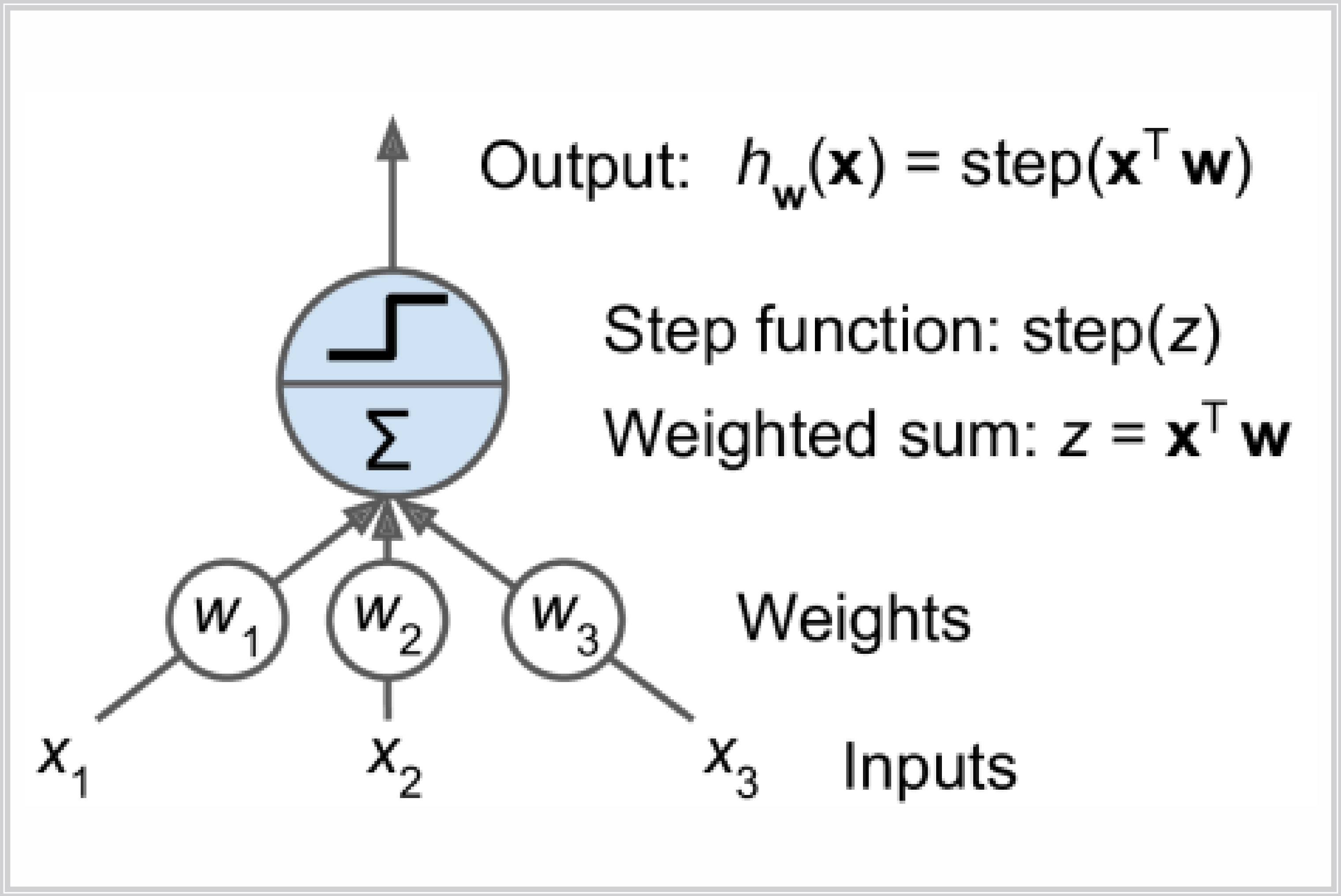




- Mostly found in animal cerebral cortexes (e.g., your brain)
- Biological neurons receive short electrical impulses called signals from other neurons via synapses.
- When a neuron receives a sufficient number of signals from other neurons within a few milliseconds, it fires its own signals.

- Has one or more binary (on/off) inputs and one binary output. The artificial neuron simply activates its output when more than a certain number of its inputs are active.
- With such a simple model it is possible to build a network of artificial neurons that can compute any logical proposition.

# The Humble TLU

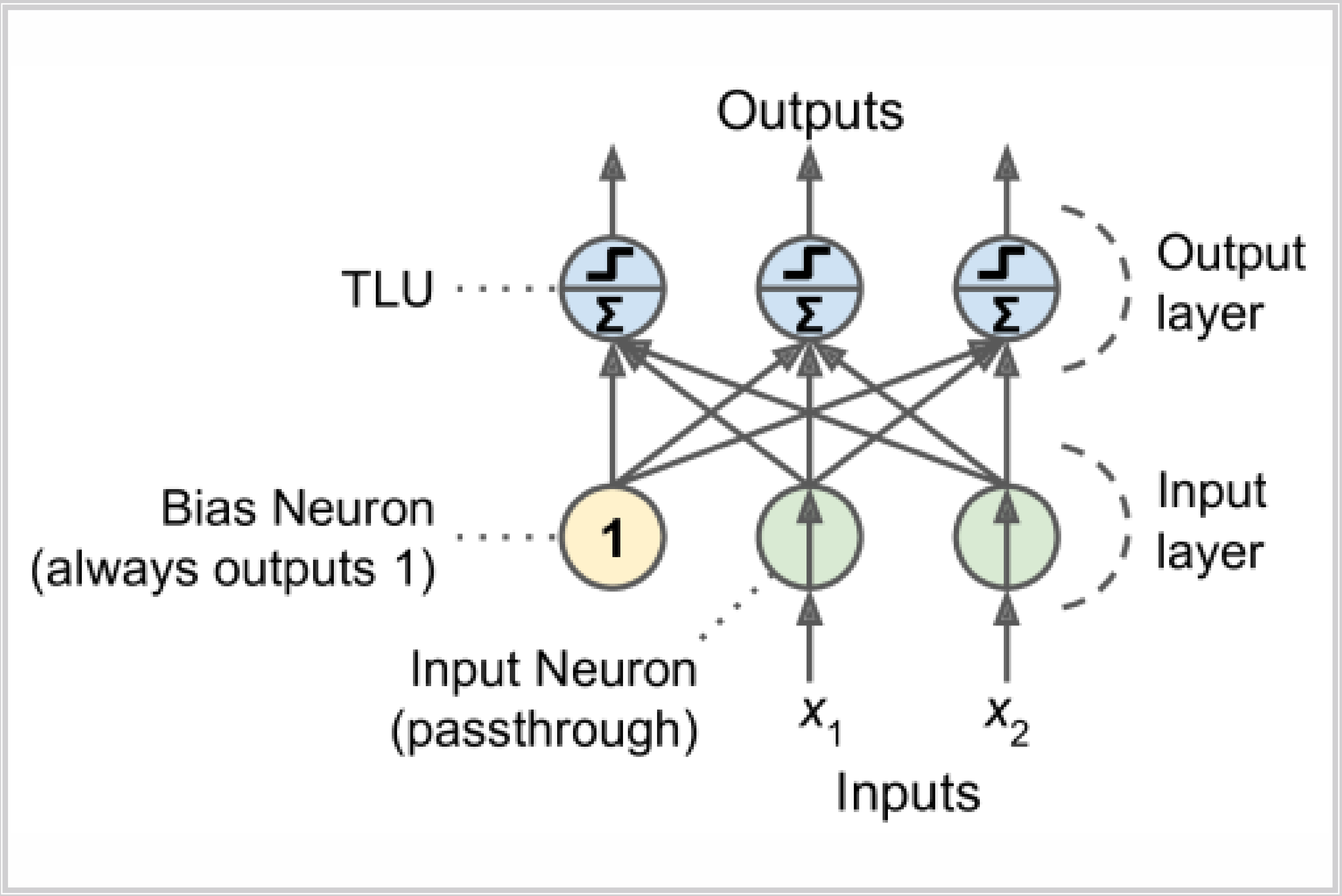
- The inputs and output are numbers and each input connection is associated with a weight.
- TLU computes a weighted sum of its inputs then applies a step function to that sum and outputs the result.
- A single TLU can be used for simple linear binary classification.
  - It computes a linear combination of the inputs;
  - If the result exceeds a threshold, it outputs the positive class; else outputs the negative class
  - Just like a Logistic Regression classifier or a linear SVM



# Perceptron

- A perceptron is composed of a single layer of TLUs.
- A Perceptron with two inputs and three outputs is represented in the image.
  - This Perceptron can classify instances simultaneously into three different binary classes.

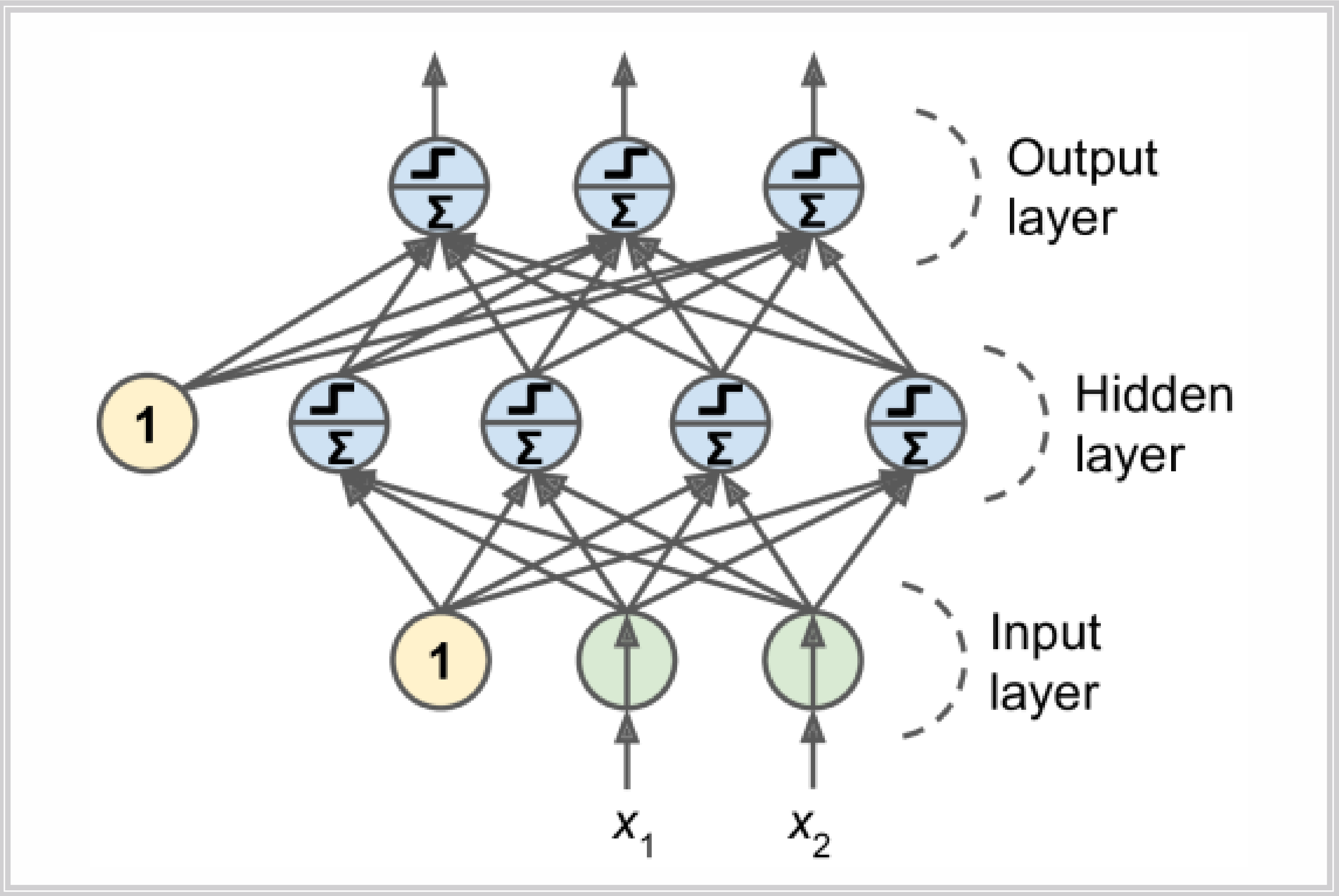
- Training a Perceptron:
  - The Perceptron is fed one training instance at a time, and for each instance it makes its predictions.
  - For every output neuron that produces a wrong prediction, it reinforces the connection weights from the inputs that would have contributed to the correct prediction.

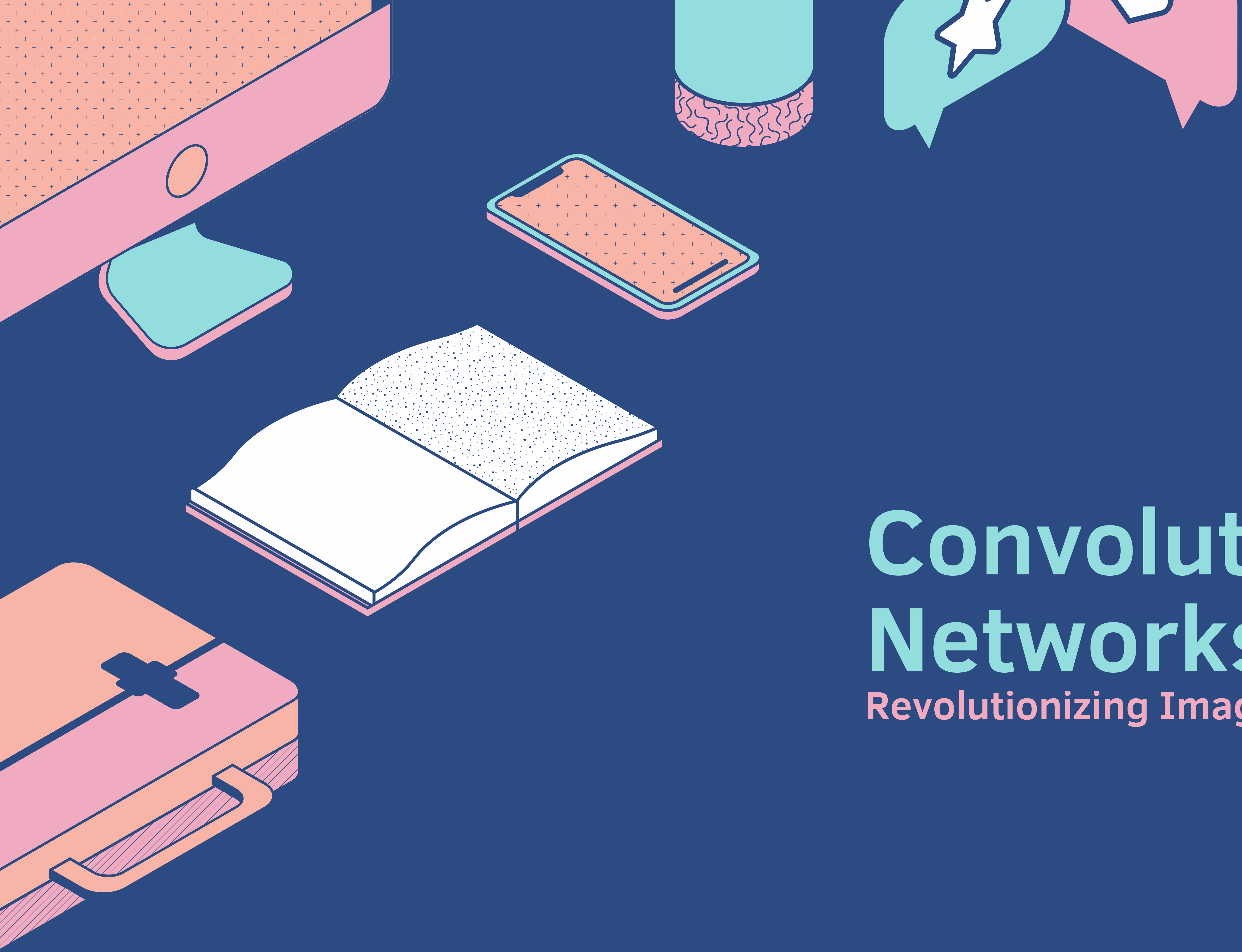


# Multi-Layer Perceptron

- An MLP is composed of:
    - Input layer
    - One or more layers of TLUs, called hidden layers
    - Final layer of TLUs called the output layer.

- This structure is the basis for modern artificial neural networks (ANNs).
- When an ANN contains a deep stack of hidden layers, it is called a deep neural network (DNN).
- The field of Deep Learning studies DNNs.





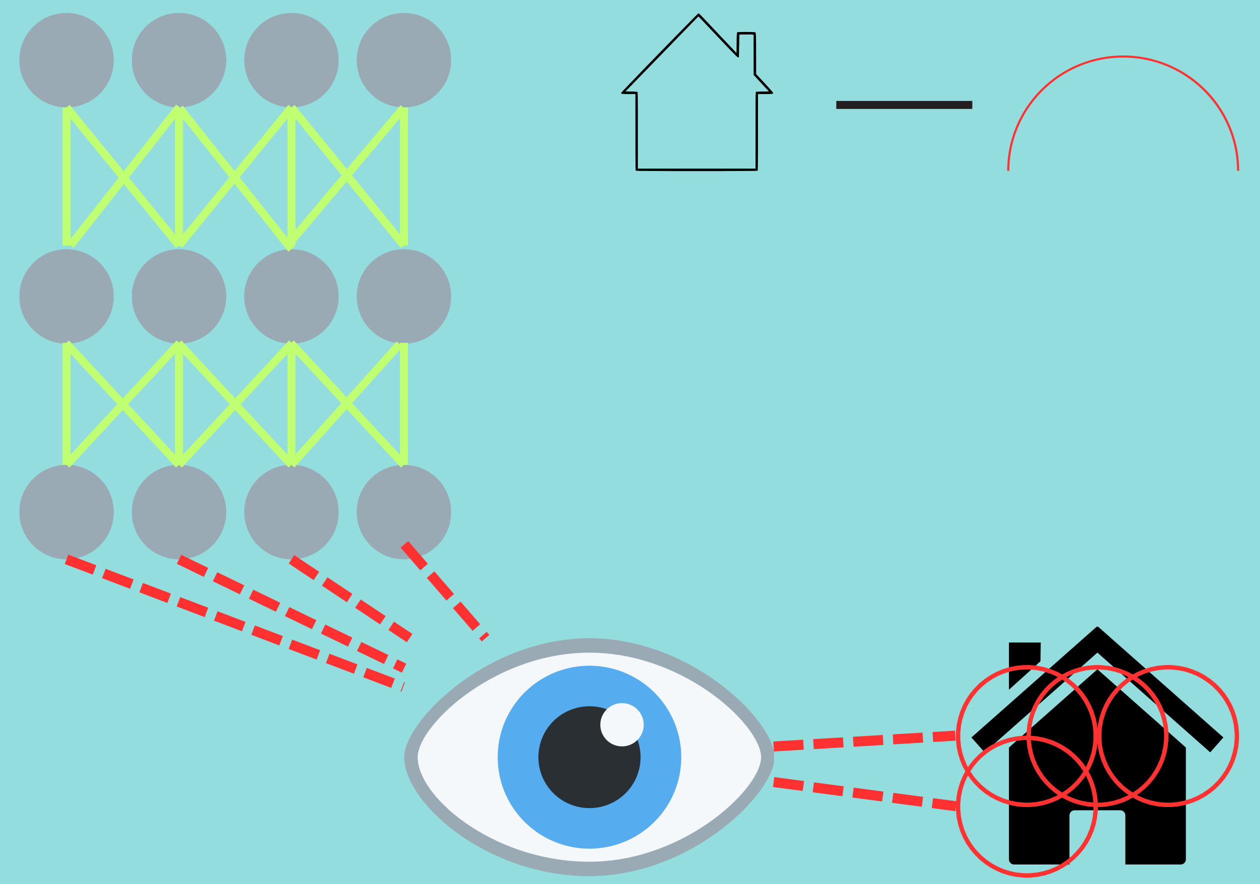
# Convolutional Neural Networks

## Revolutionizing Image Processing

# The Architecture of the Visual Cortex

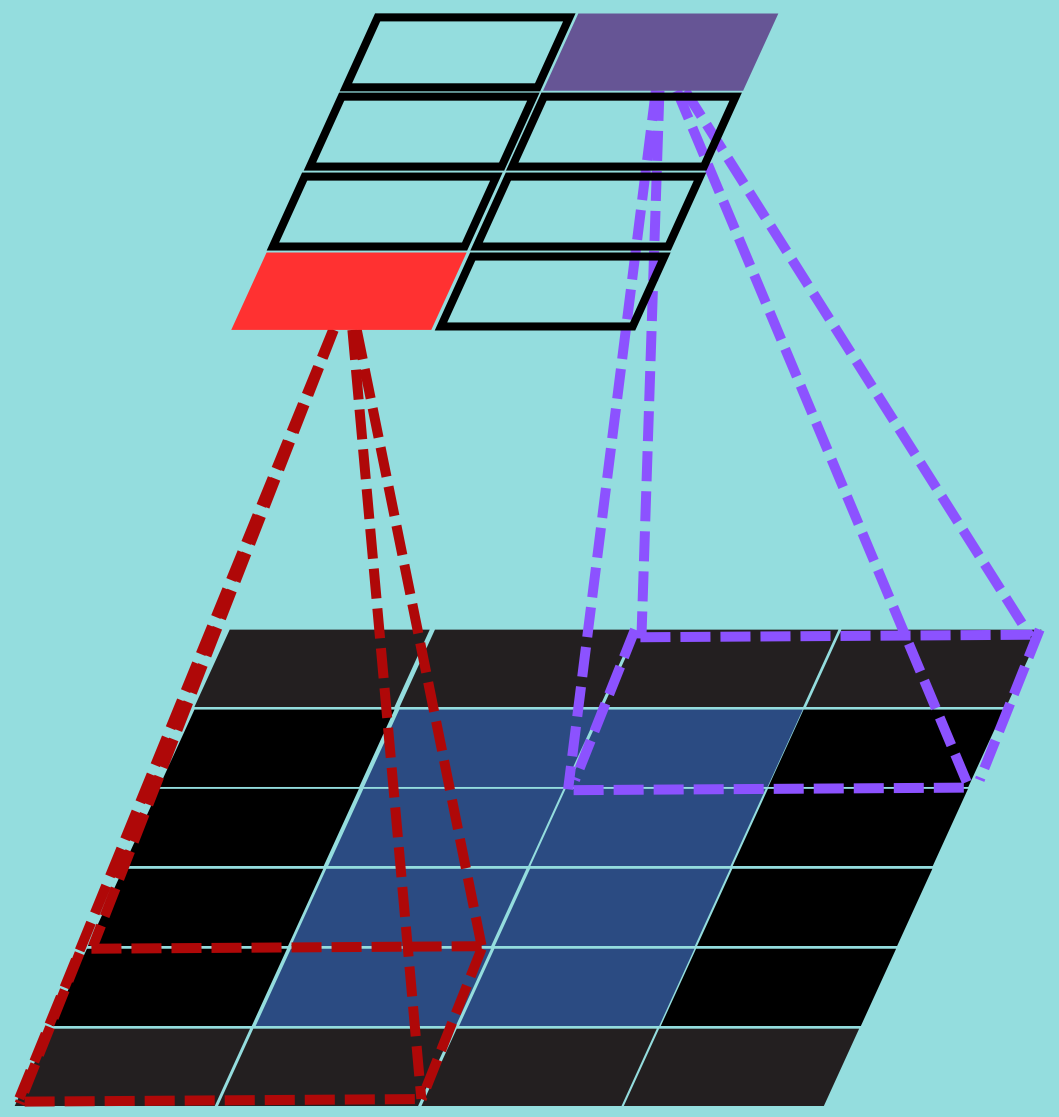
- Inspired by findings from Research experiments done on cats in the in the 1950s
  - Many neurons in the visual cortex have a small local receptive field
  - The receptive fields of different neurons may overlap, and together they tile the whole visual field.

- These studies gradually evolved into what we now call convolutional neural networks



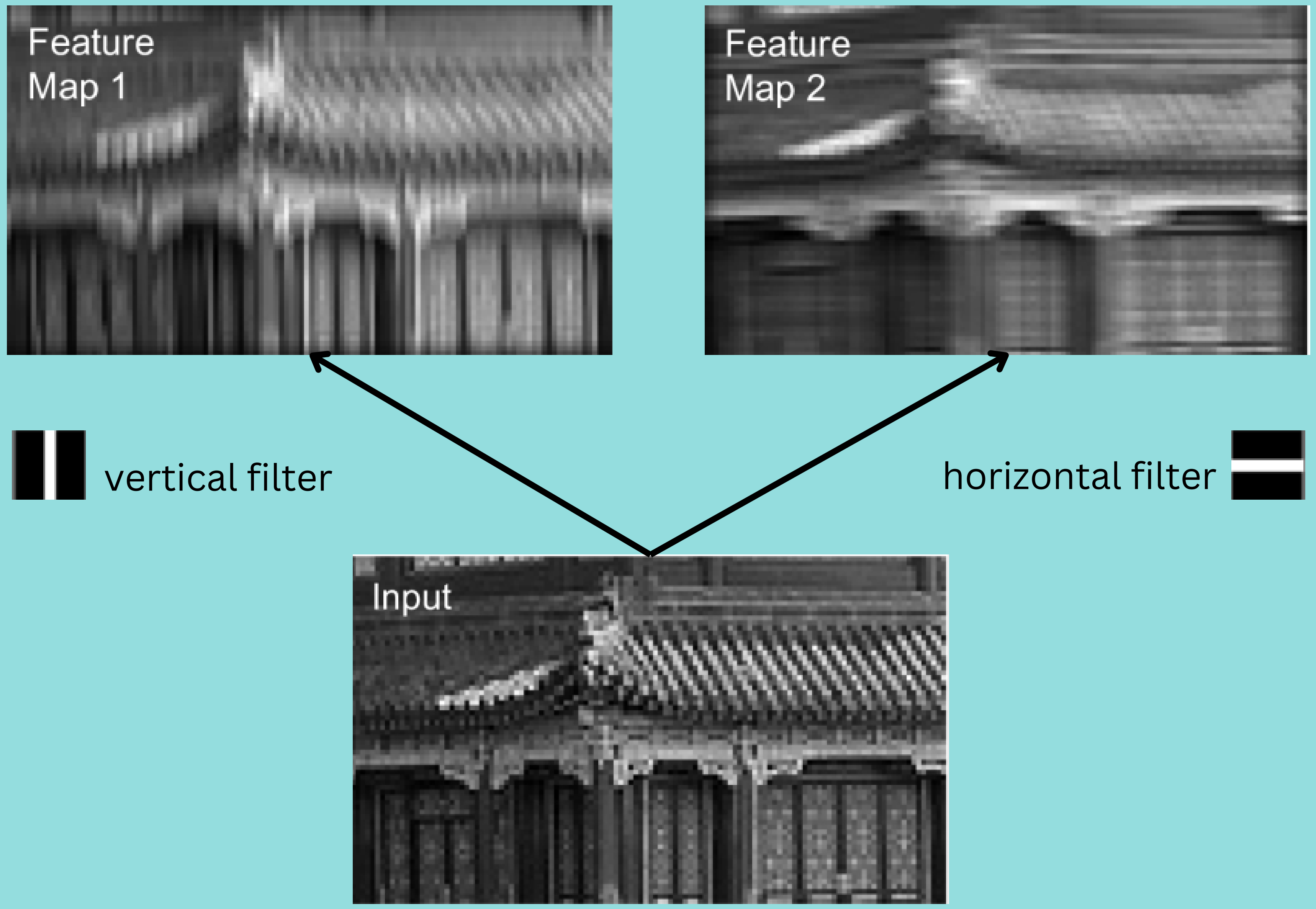
# Convolutional Layers

- Neurons in the first convolutional layer are not connected to every
- Single pixel in the input image, but only to pixels in the receptive field, and so on.
- This structure allows the network to focus on small low-level features in the next hidden layer and then assemble into larger higher level features.



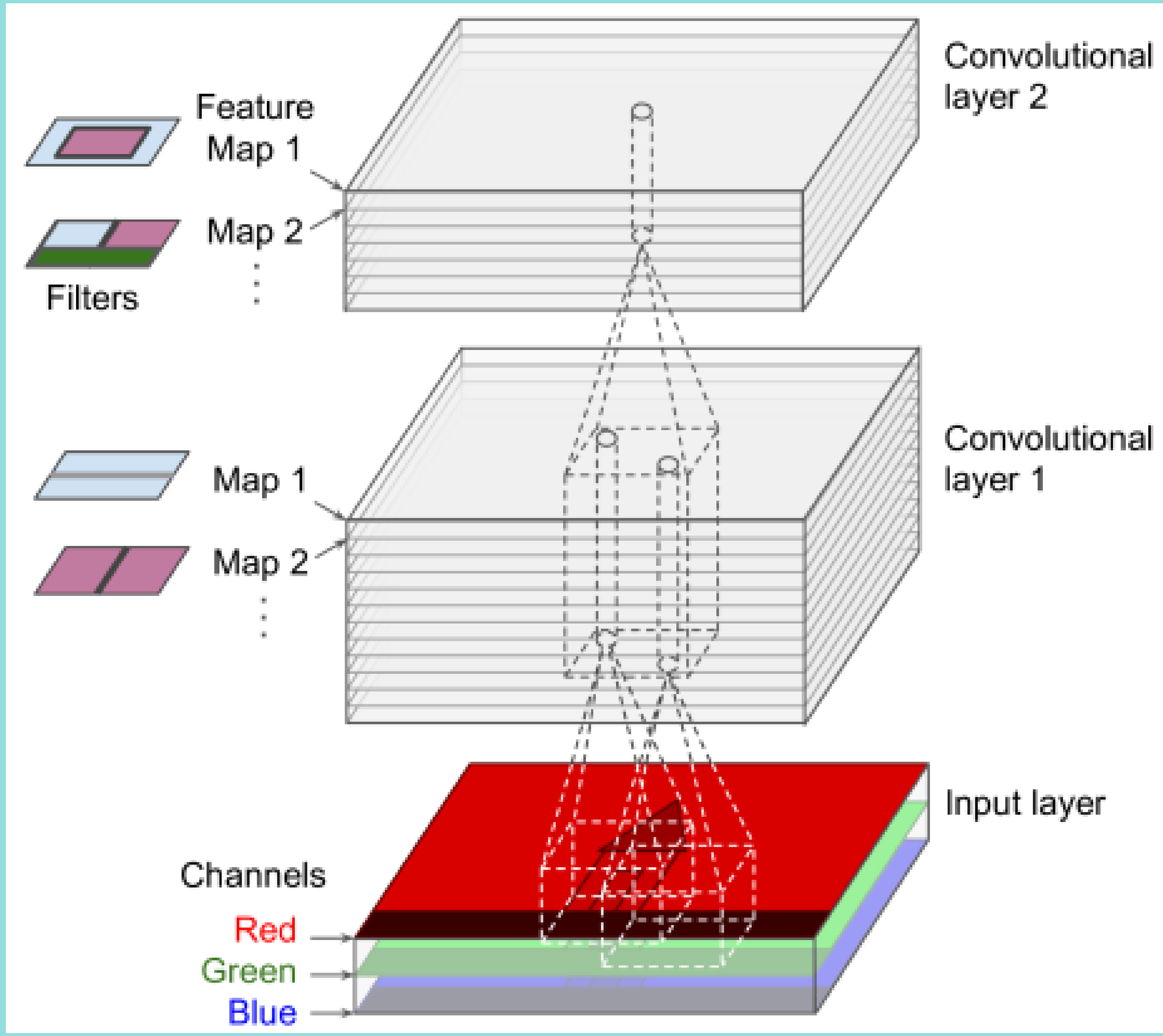
# Filters: optimizing weights in a CNN

- A neuron's weight can be represented as a small image the size of the receptive field; called filters.
- A layer full of neurons using the same filter outputs a feature map, which highlights the areas in an image that activate the filter the most.
- The filters are not defined manually
  - The convolutional layer will automatically learn the most useful features for its task. And the layers above will combine them into complex patterns.

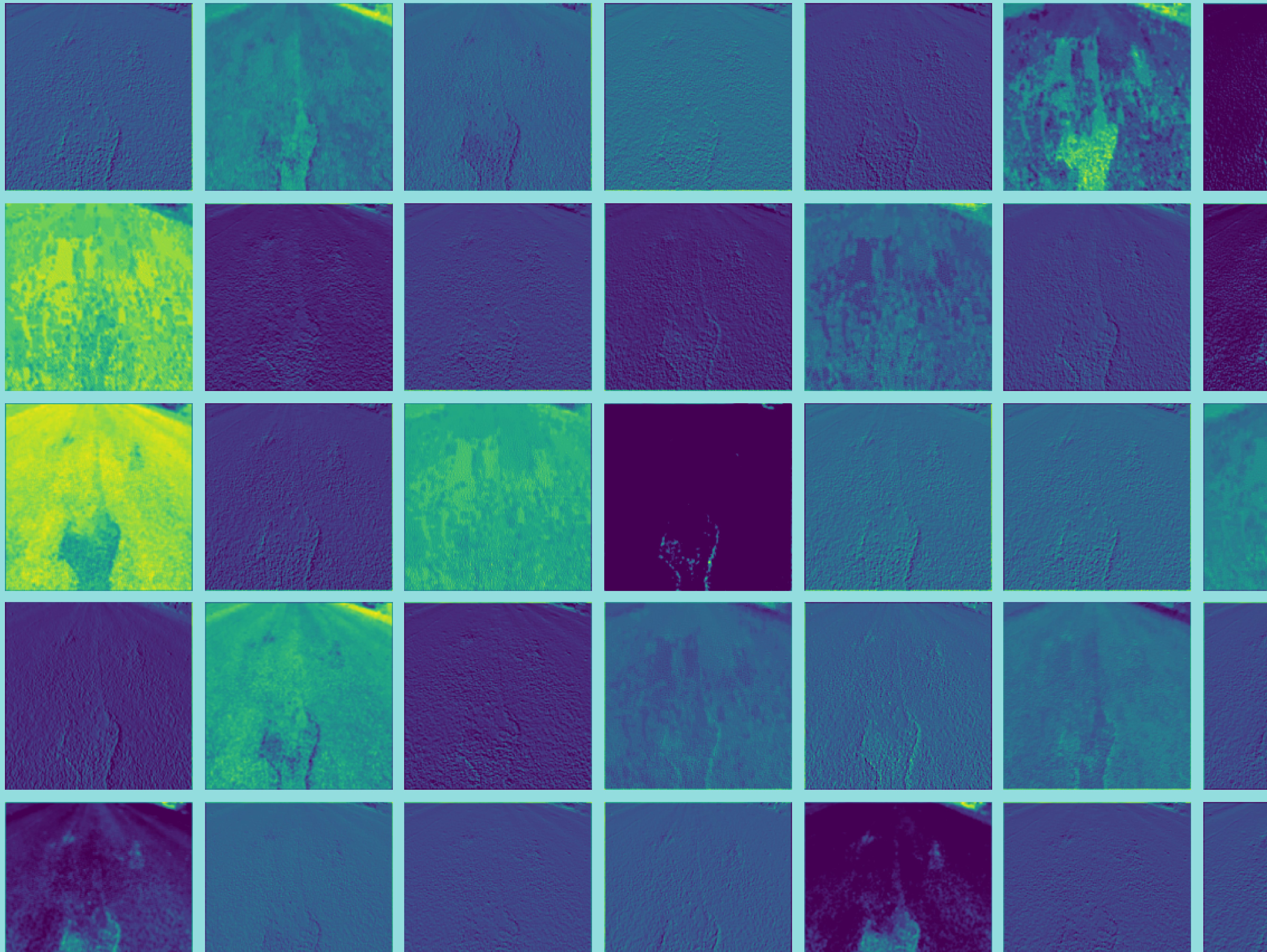


# Stacking Feature Maps

- A convolutional layer has multiple filters, and outputs one feature map per filter.
- A neuron's receptive field is the same as described extends across all feature maps.
- A convolutional layer simultaneously applies multiple trainable filters to its inputs, making it capable of detecting multiple features anywhere in its inputs.



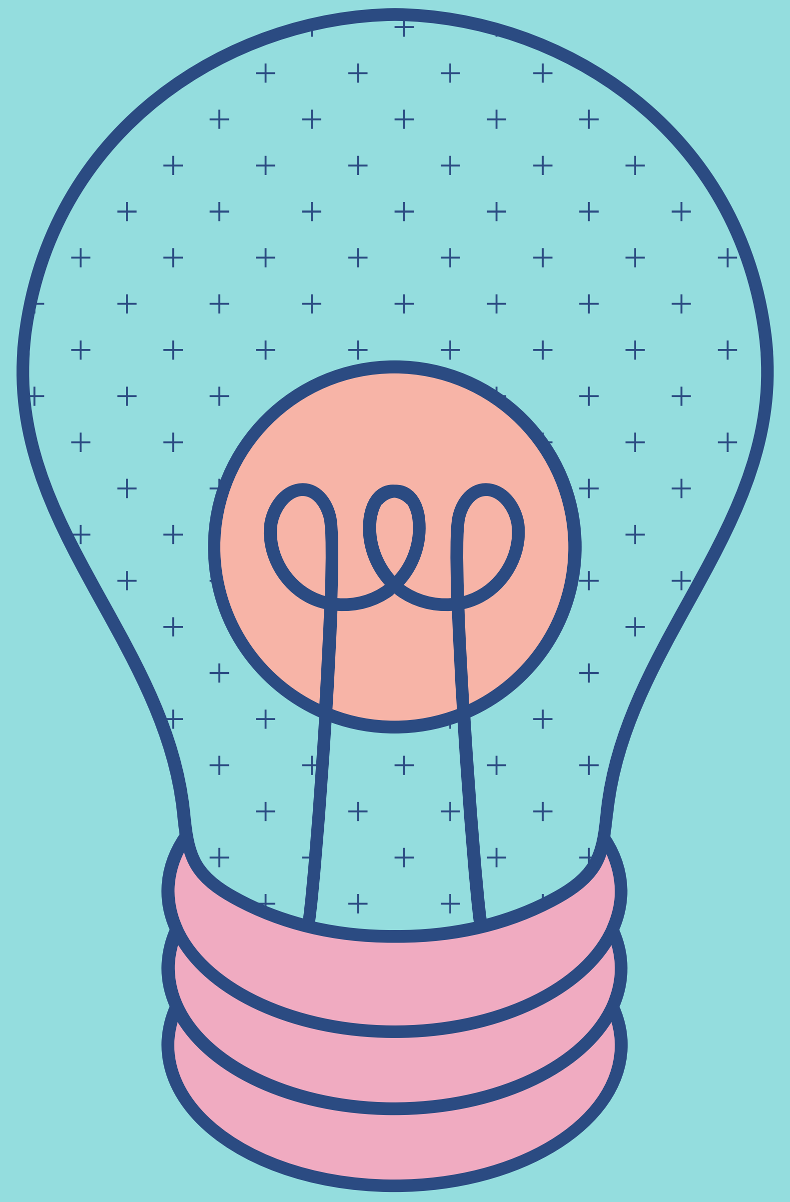
Feature map output of the first convolutional block of the VGG16 model



# Why Convolutional Layers?

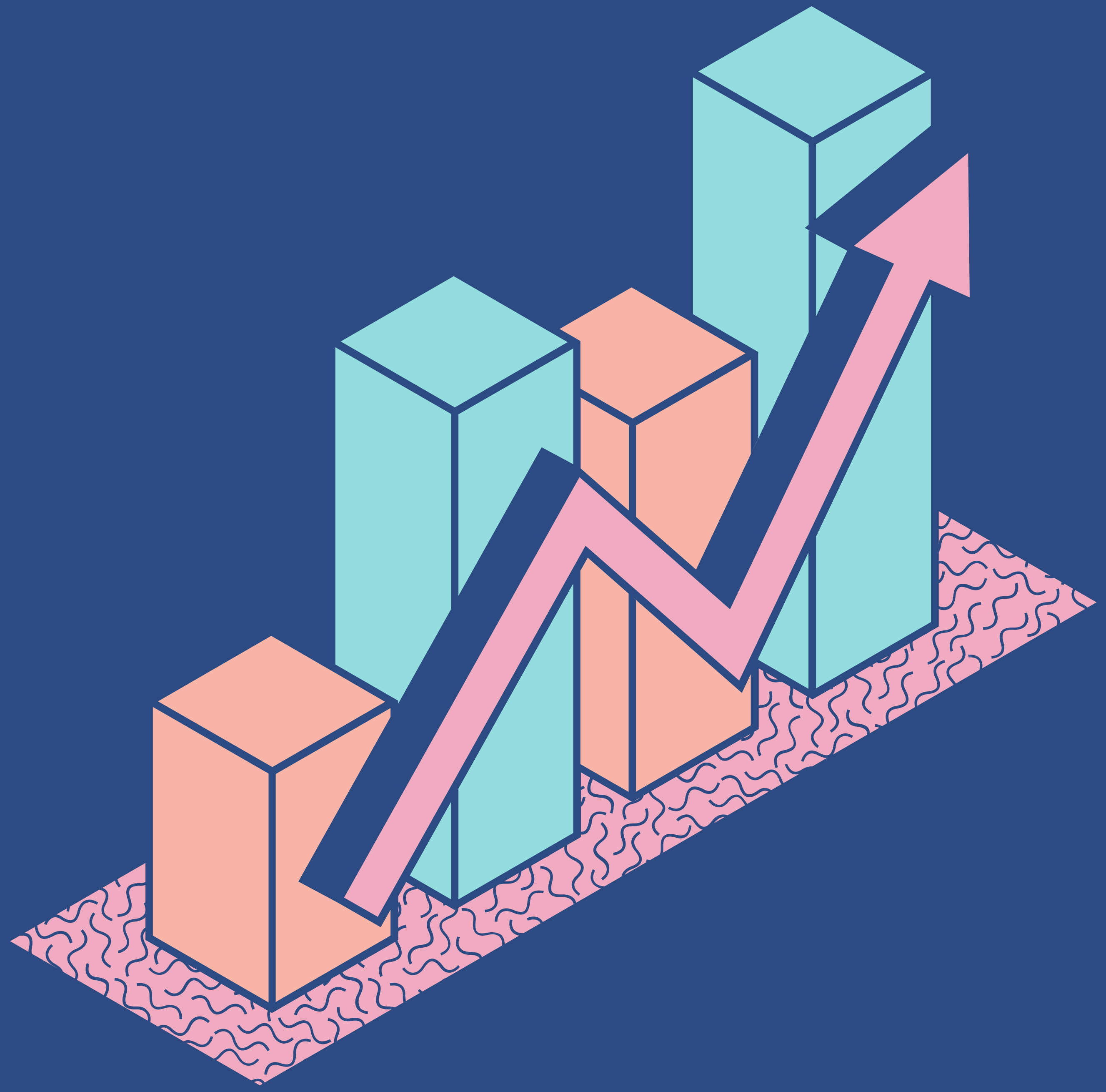
- Dense works fine for small images; it breaks down for larger images because of the huge number of parameters it requires.
  - For example, a 100 × 100 image has 10,000 pixels, and if the first layer has just 1,000 neurons (which already severely restricts the amount of information transmitted to the next layer), this means a total of 10 million connections. And that's just the first layer. CNNs solve this problem using partially connected layers and weight sharing.

- All neurons in a feature map in a CNN share the same parameters which dramatically reduces the number of parameters in the model.
- Once the CNN has learned to recognize a pattern in one location, it can recognize it in any other location.
- In contrast, once a regular DNN has learned to recognize a pattern in one location, it can recognize it only in that particular location.

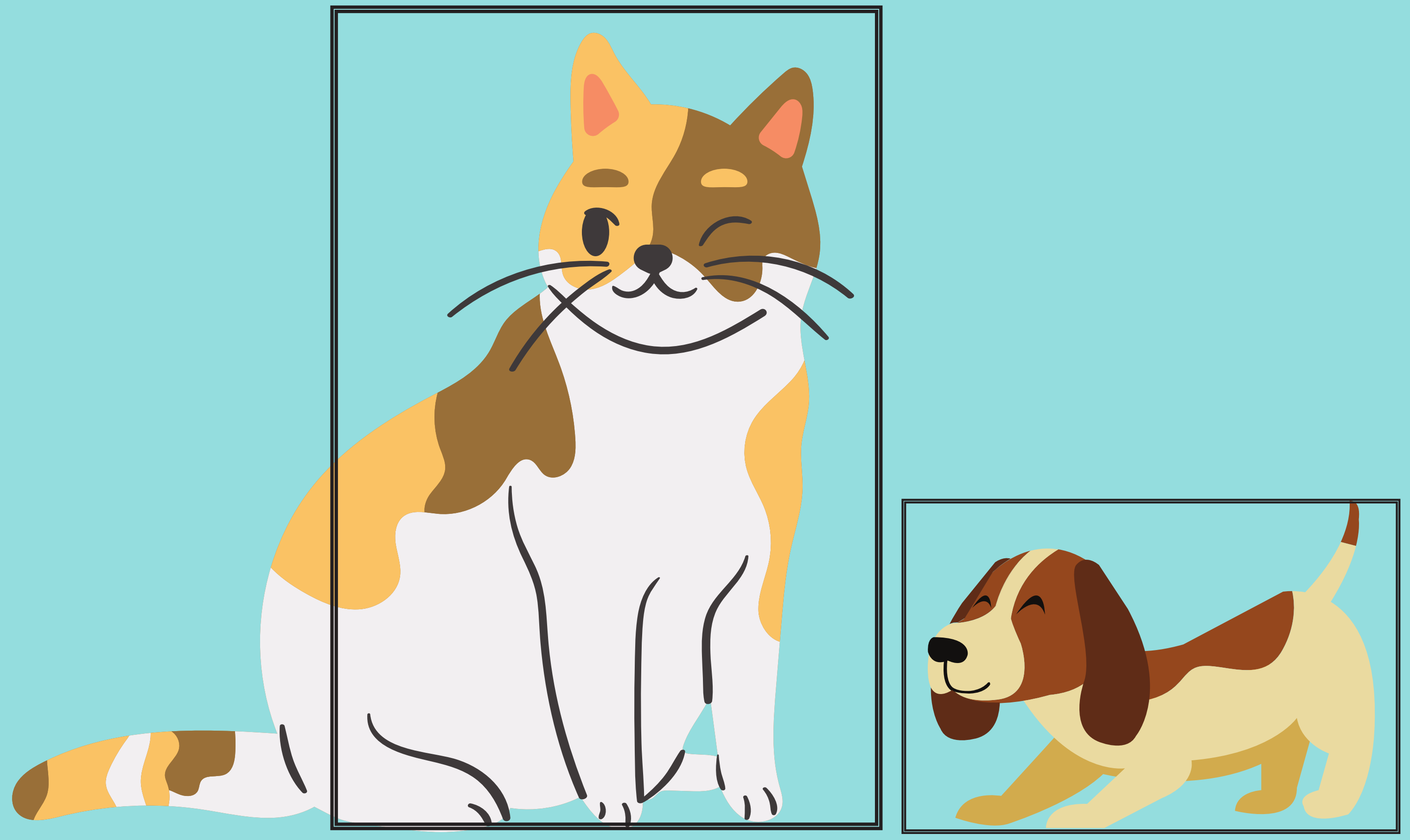


# Object Detection & Image Segmentation

## Last Chapter



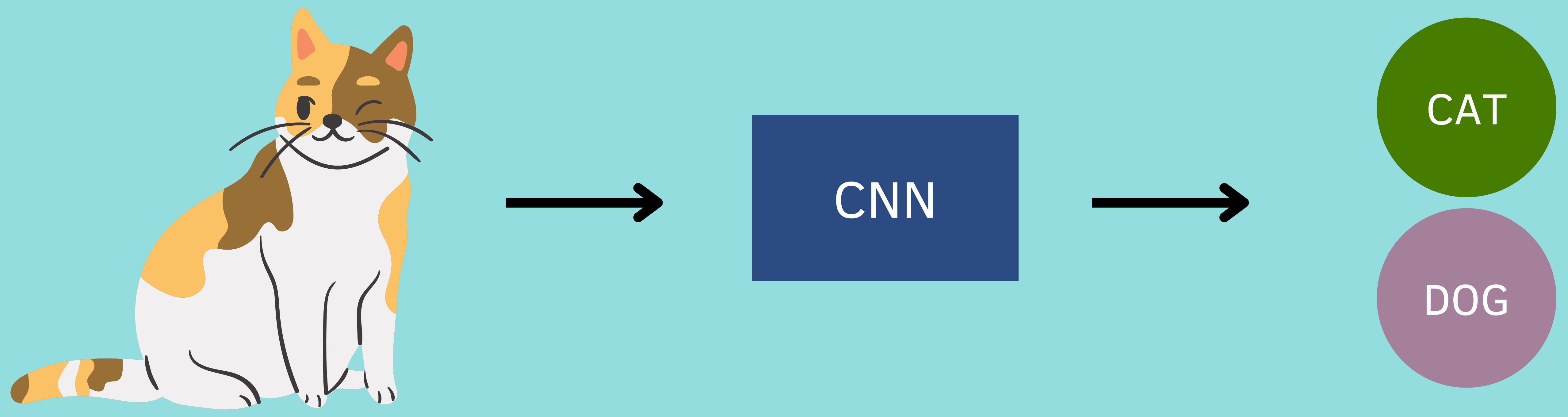
# Localization vs Detection



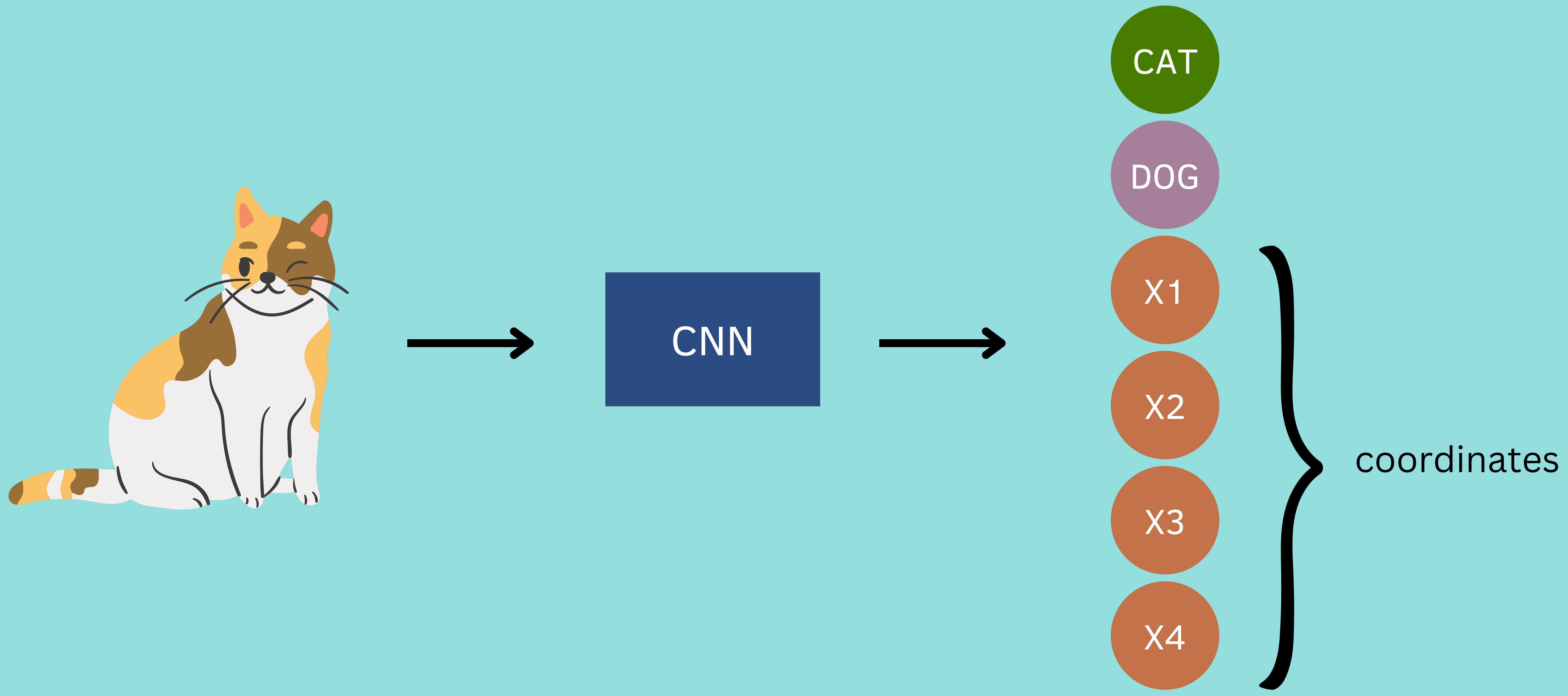
<u>Object Localization</u>: finding what and where a single object is in an image

<u>Object detection</u>: finding what and where multiple objects are in an image

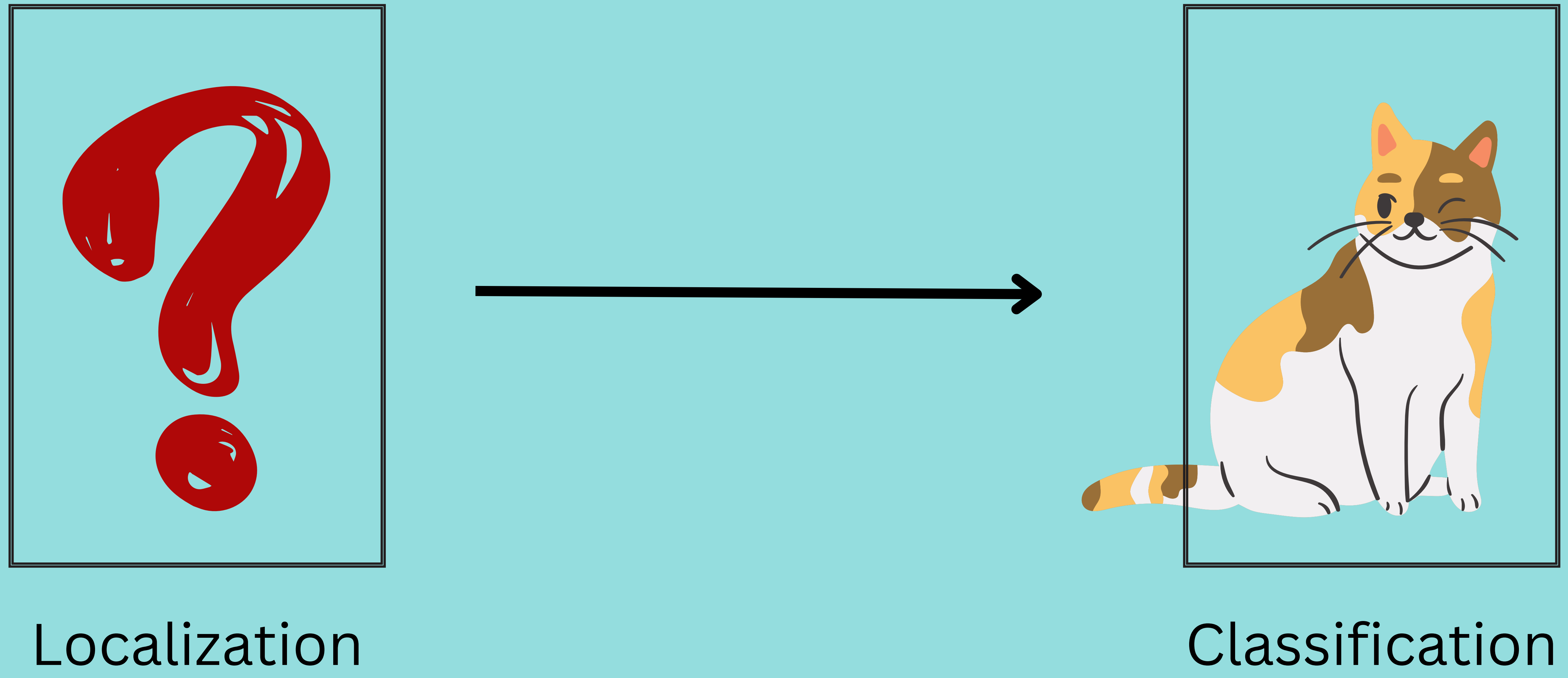
# Classification



# Localization



# Detection

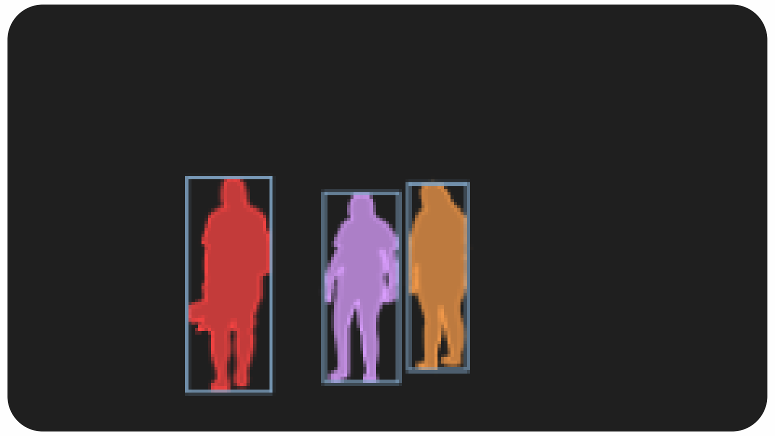


# Segmentation

**Method of dividing a digital image into subgroups called image segments, reducing the complexity of the image and enabling further processing or analysis of each image segment.**

No segmentation

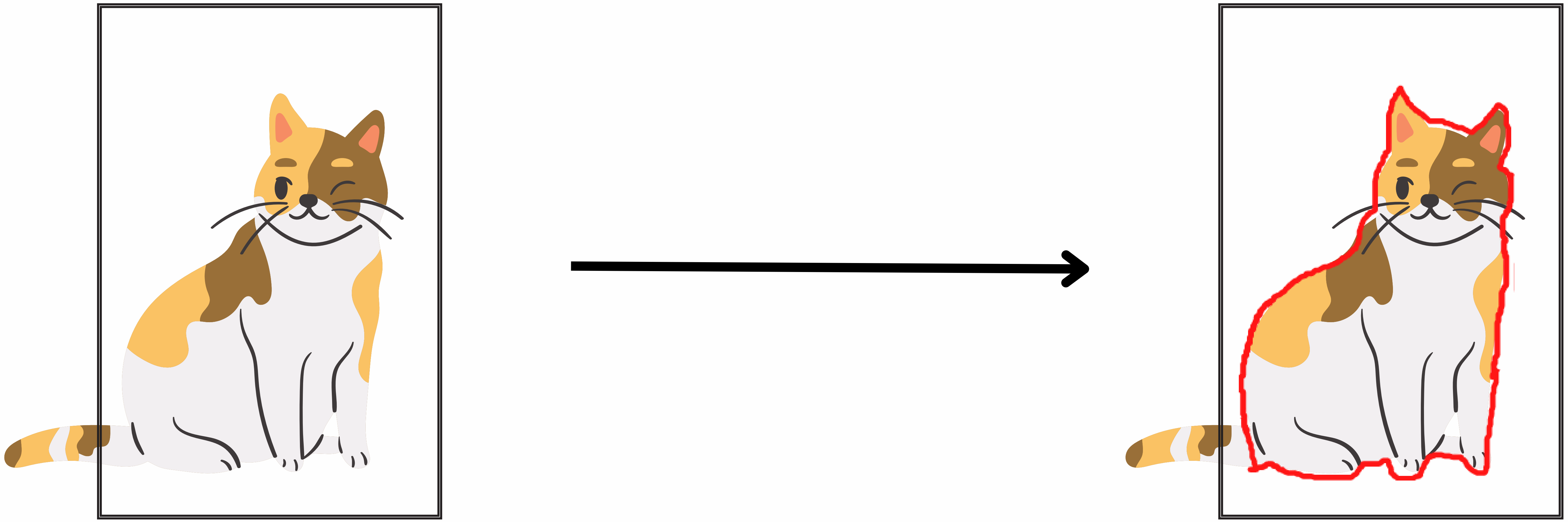
Semantic Seg.



Instance Seg.

Panoptic Seg.

# Instance Segmentation



Object Detection

Semantic Segmentation

# Evaluation Metrics

**Precision**

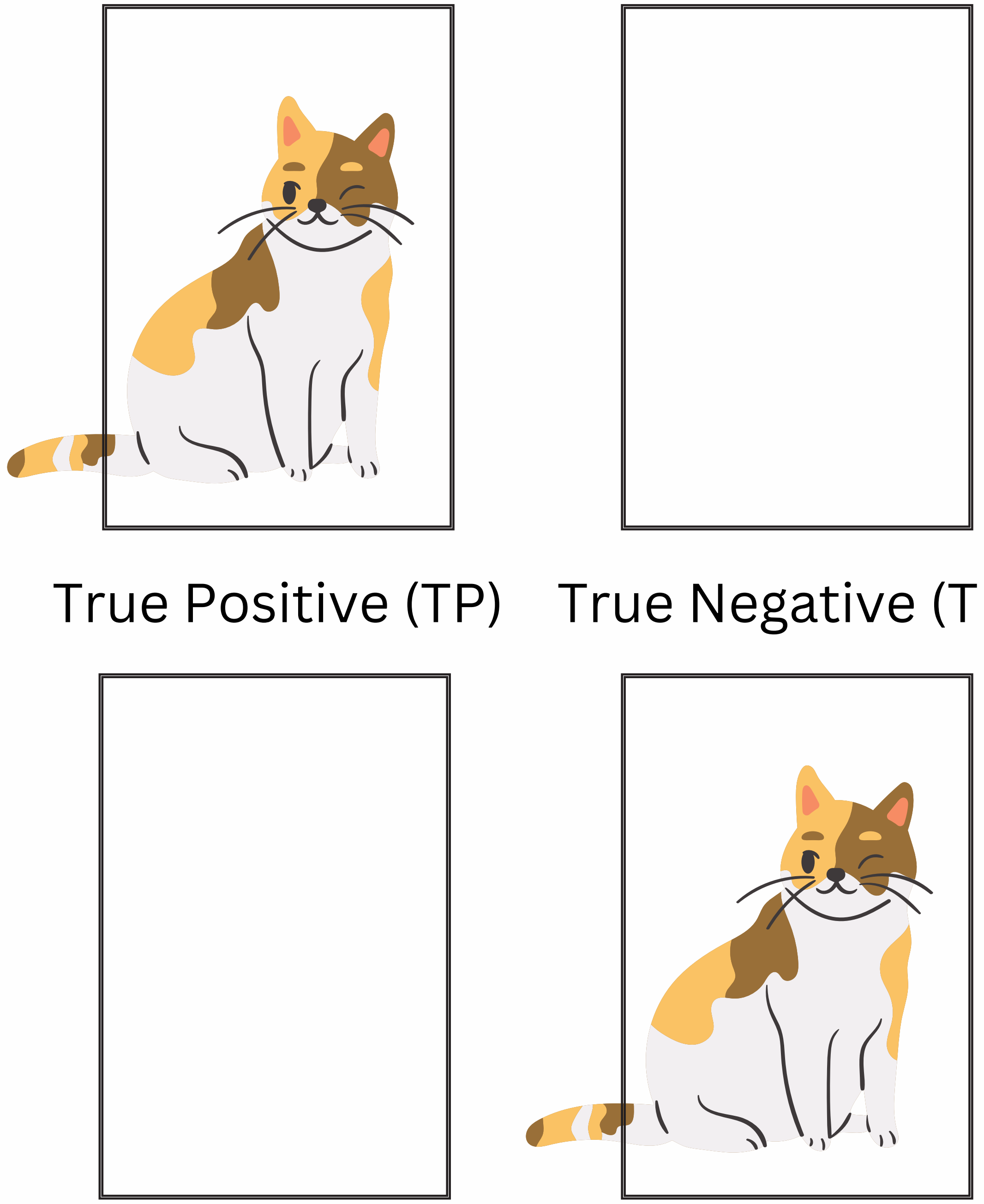
$$\frac{TP}{TP + FP}$$

**Recall**

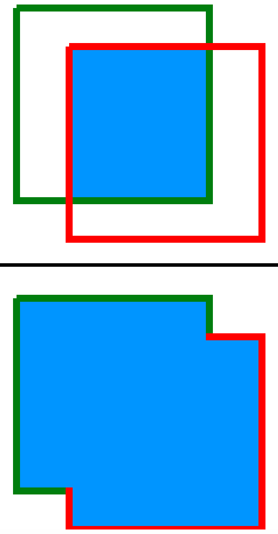
$$\frac{TP}{TP + FN}$$

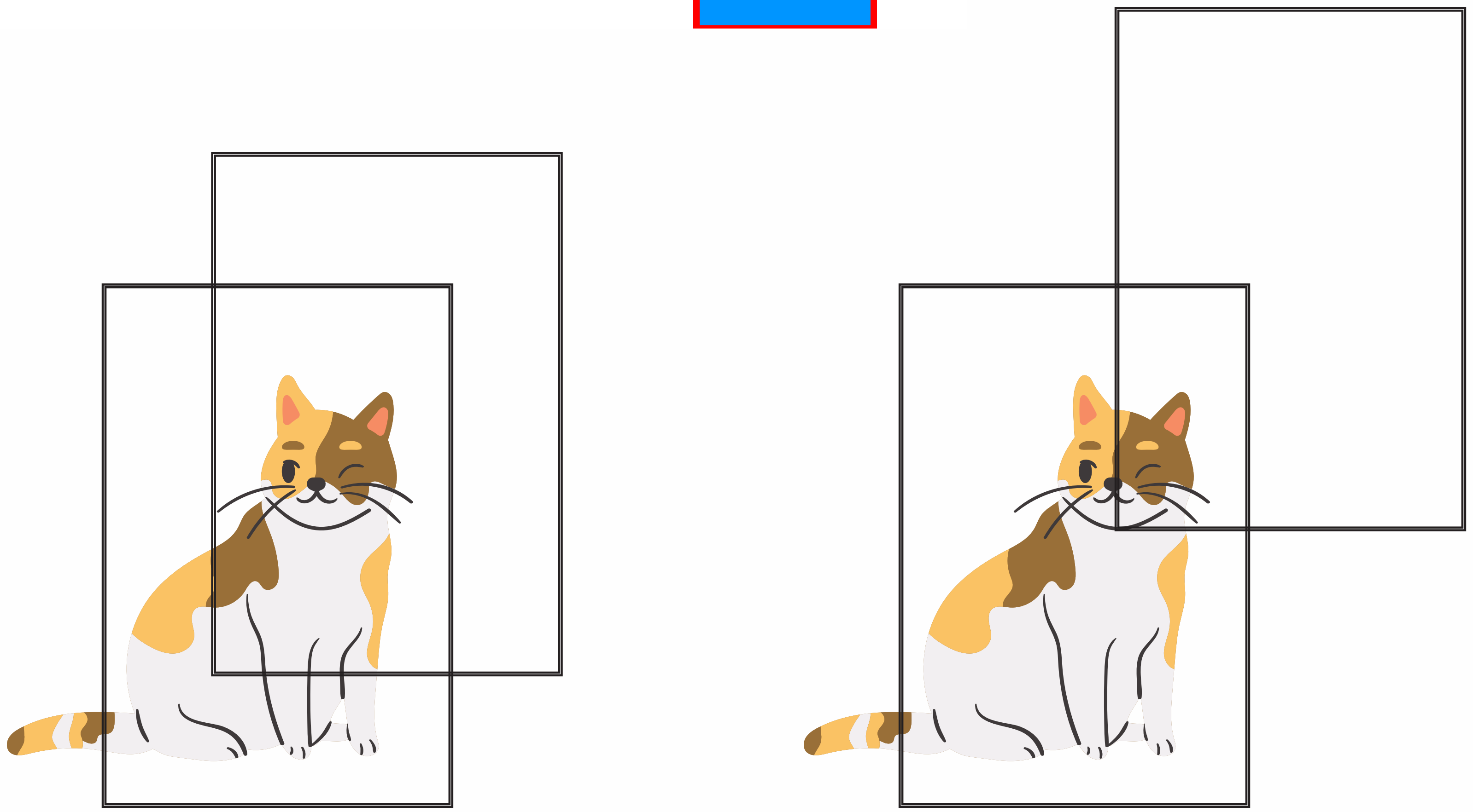
**mAP**

$$1/N \sum^{N} \int P(r)\, dr$$




$$IOU = \frac{\text{area of overlap}}{\text{area of union}} =$$



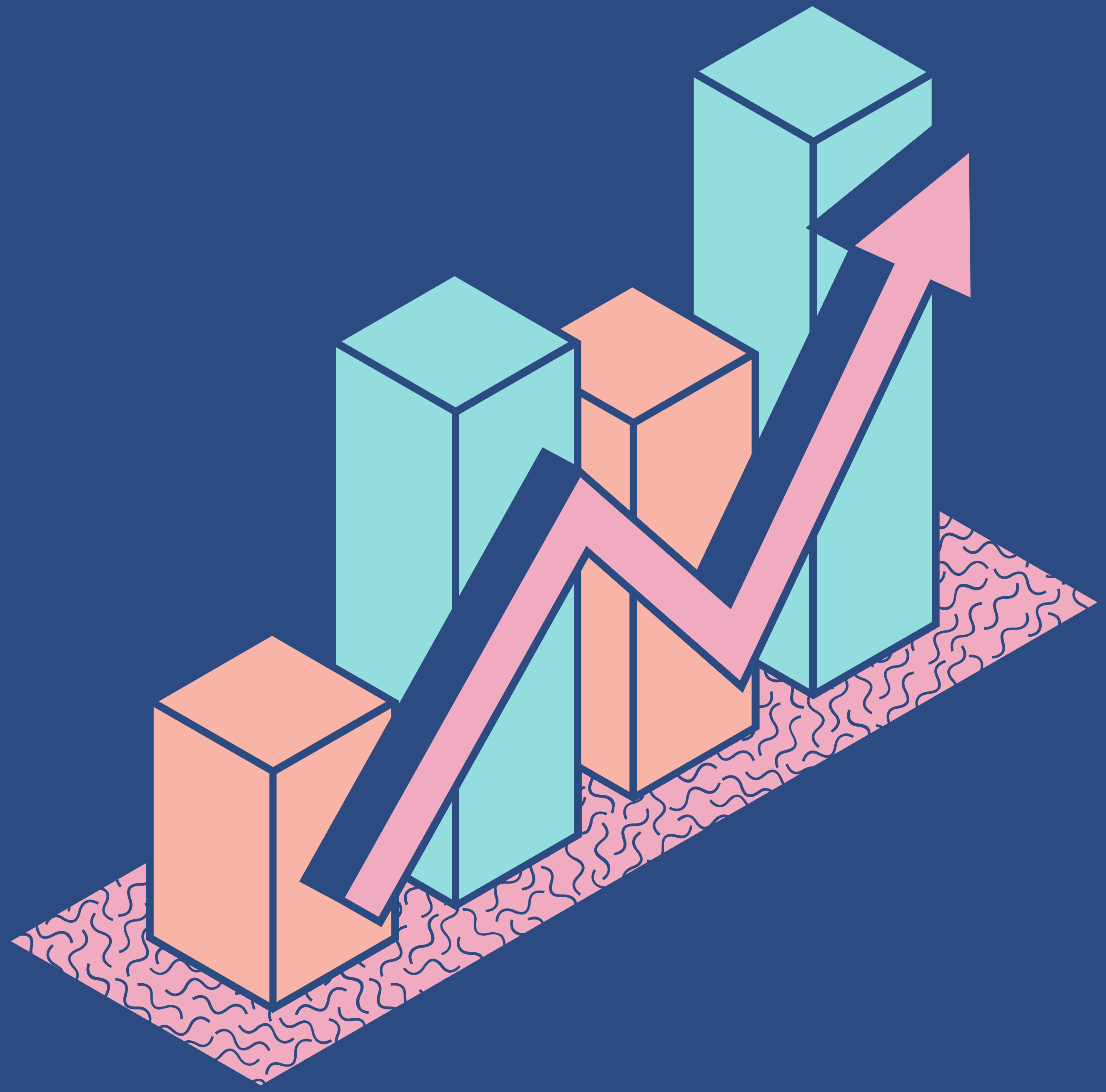


IoU = 0.50

IoU = 0.25

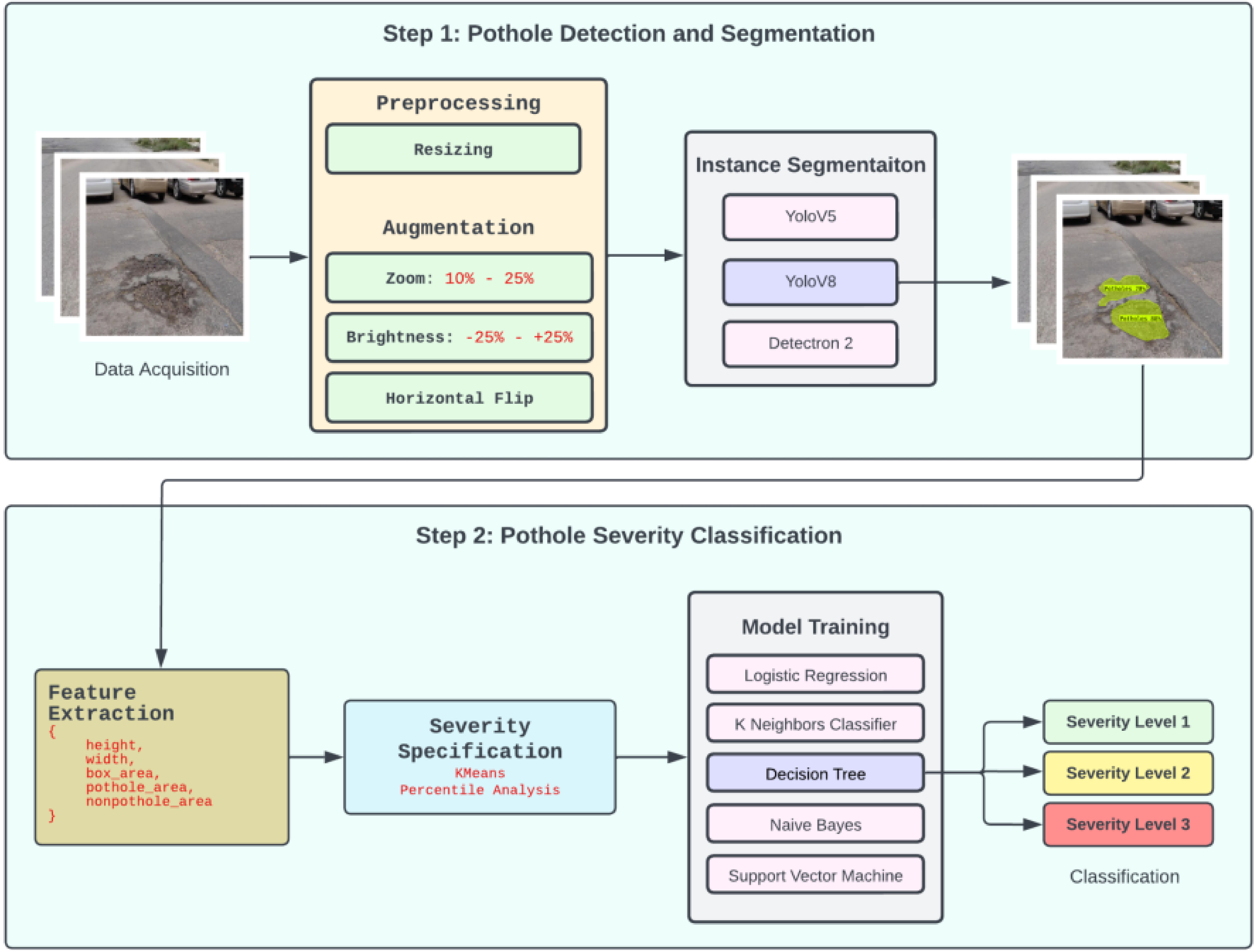
# Pothole Severity Classification

Last Last Chapter



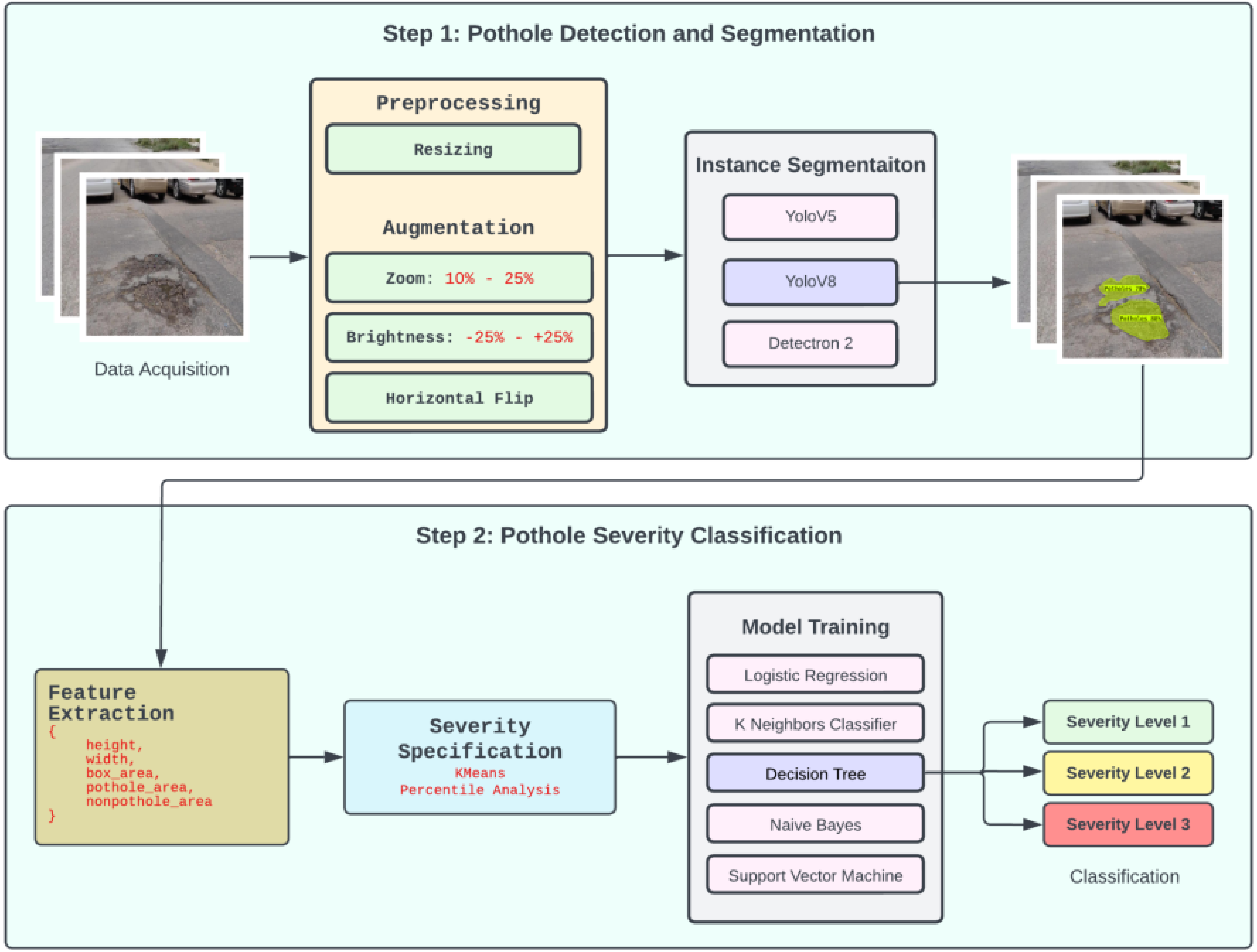
# DATASET PREPARATION

- Dataset contains 768 images of potholes.
- Images are resized to 640x640.
- Training set was augmented; resulted in 1400 images



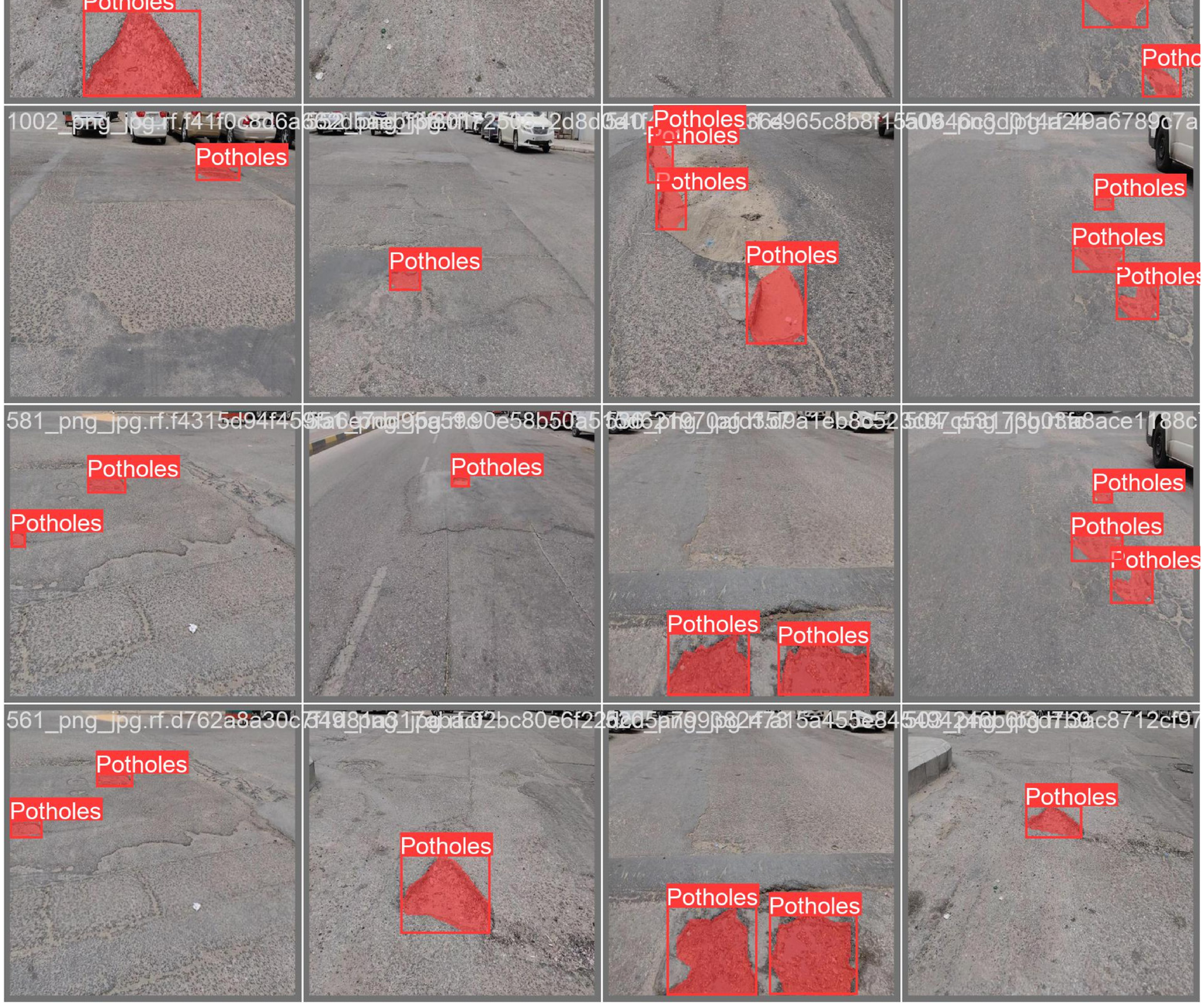
# INSTANCE SEGMENTATION MODEL

- 3 models were trained; YoloV8 performed the best.
- YoloV8 results:
  - mAP:50-95 (B) = 0.734
  - mAP:50-95 (S) = 0.705



# INSTANCE SEGMENTATION MODEL

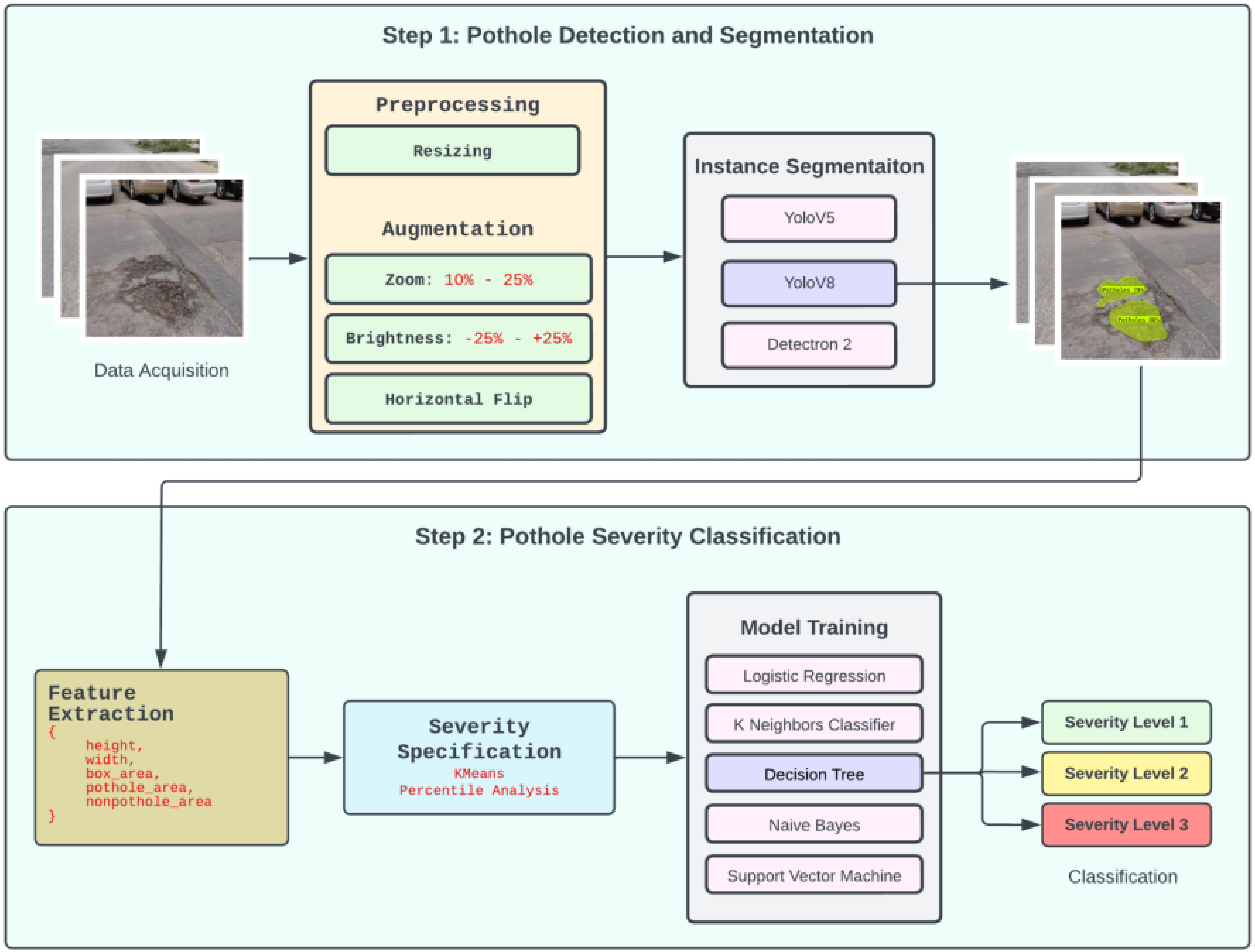
- 3 models were trained; YoloV8 performed the best.
- YoloV8 results:
  - mAP:50-95 (B) = 0.734
  - mAP:50-95 (S) = 0.705



# CLASSIFICATION MODEL

- 5 models were trained
  - Logistic Regression
  - KNN
  - RBF SVM
  - Naive Bayes
  - Decision Tree (best)

- Decision Tree results:
  - Accuracy = 0.989



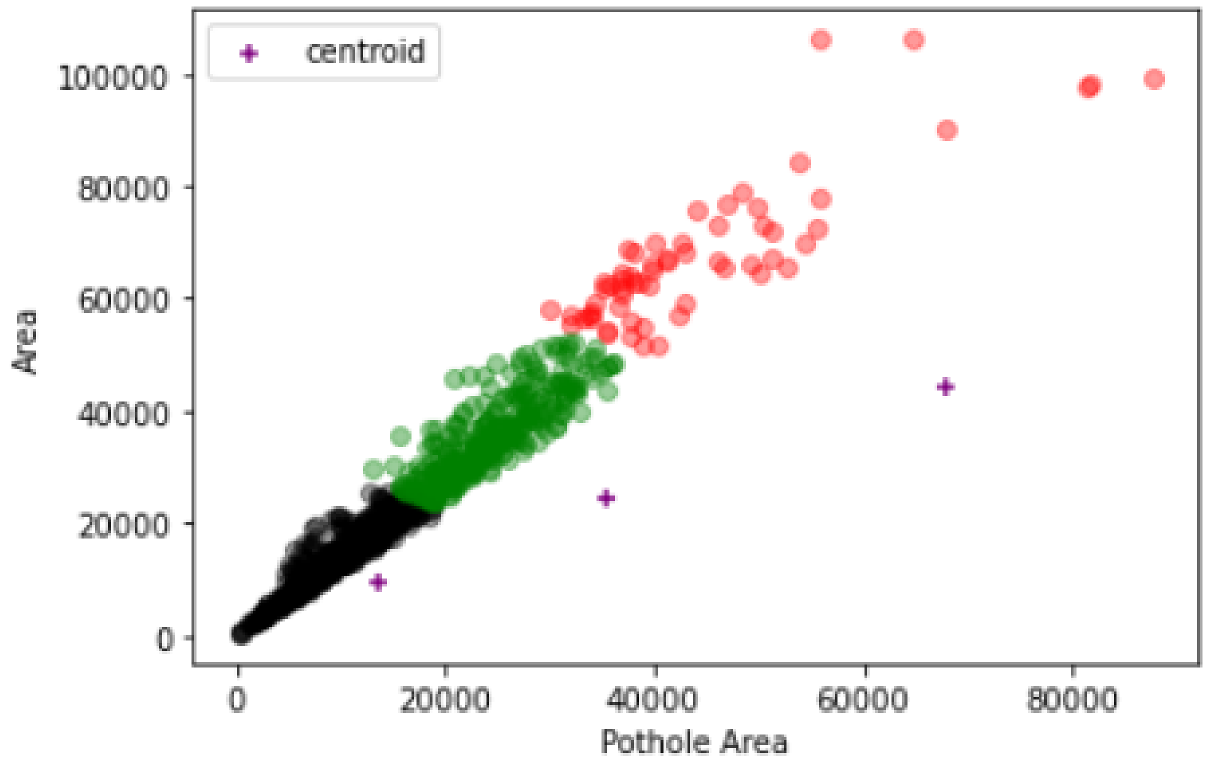
# CLASSIFICATION MODEL

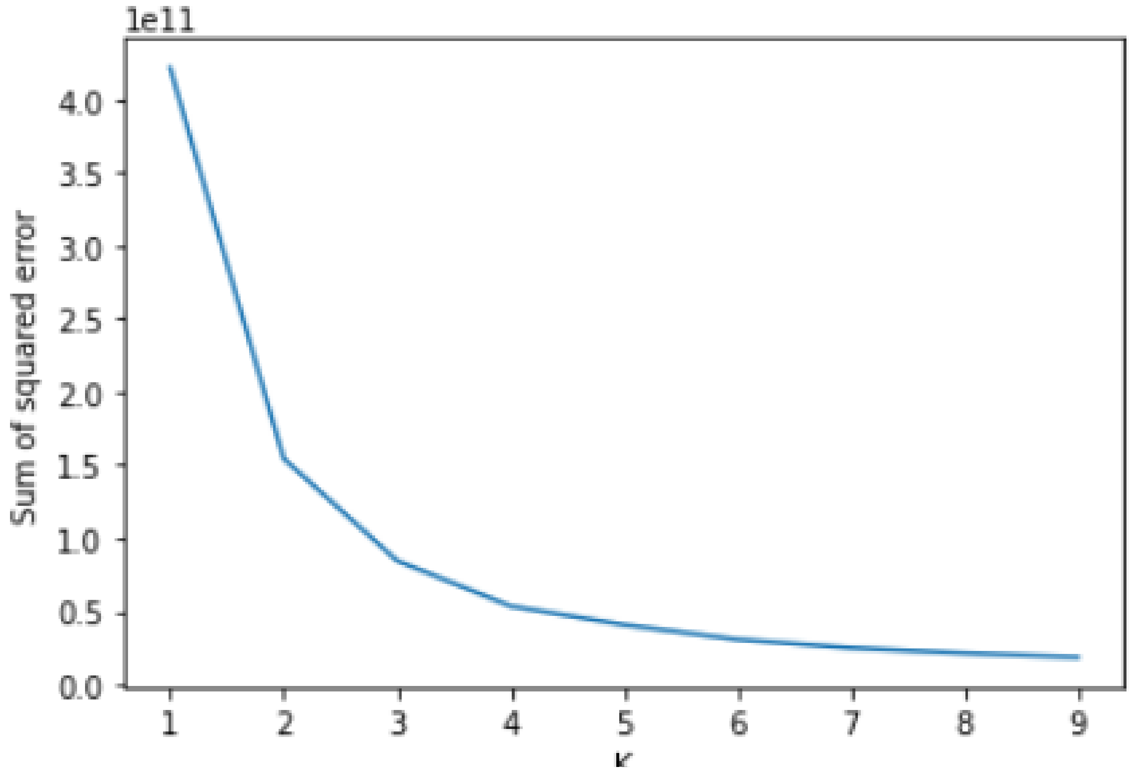
- 5 models were trained
  - Logistic Regression
  - KNN
  - RBF SVM
  - Naive Bayes
  - Decision Tree (best)
- Decision Tree results:
  - Accuracy = 0.989

**Stats Table**

| | height | width | area | pothole_area | nonpothole_area |
|---|---|---|---|---|---|
| mean | 141.246237 | 160.008602 | 24136.586022 | 16398.926882 | 7737.659140 |
| std | 61.173830 | 60.393898 | 16866.241902 | 11359.076938 | 6335.443707 |
| min | 16.000000 | 26.000000 | 416.000000 | 297.000000 | 119.000000 |
| 25% | 93.000000 | 114.000000 | 11819.500000 | 7876.750000 | 3174.500000 |
| 50% | 140.500000 | 152.000000 | 20167.500000 | 13939.000000 | 6038.500000 |
| 75% | 185.000000 | 196.000000 | 33099.000000 | 22620.000000 | 10167.500000 |
| max | 339.000000 | 368.000000 | 106107.00000 | 87658.000000 | 50293.000000 |

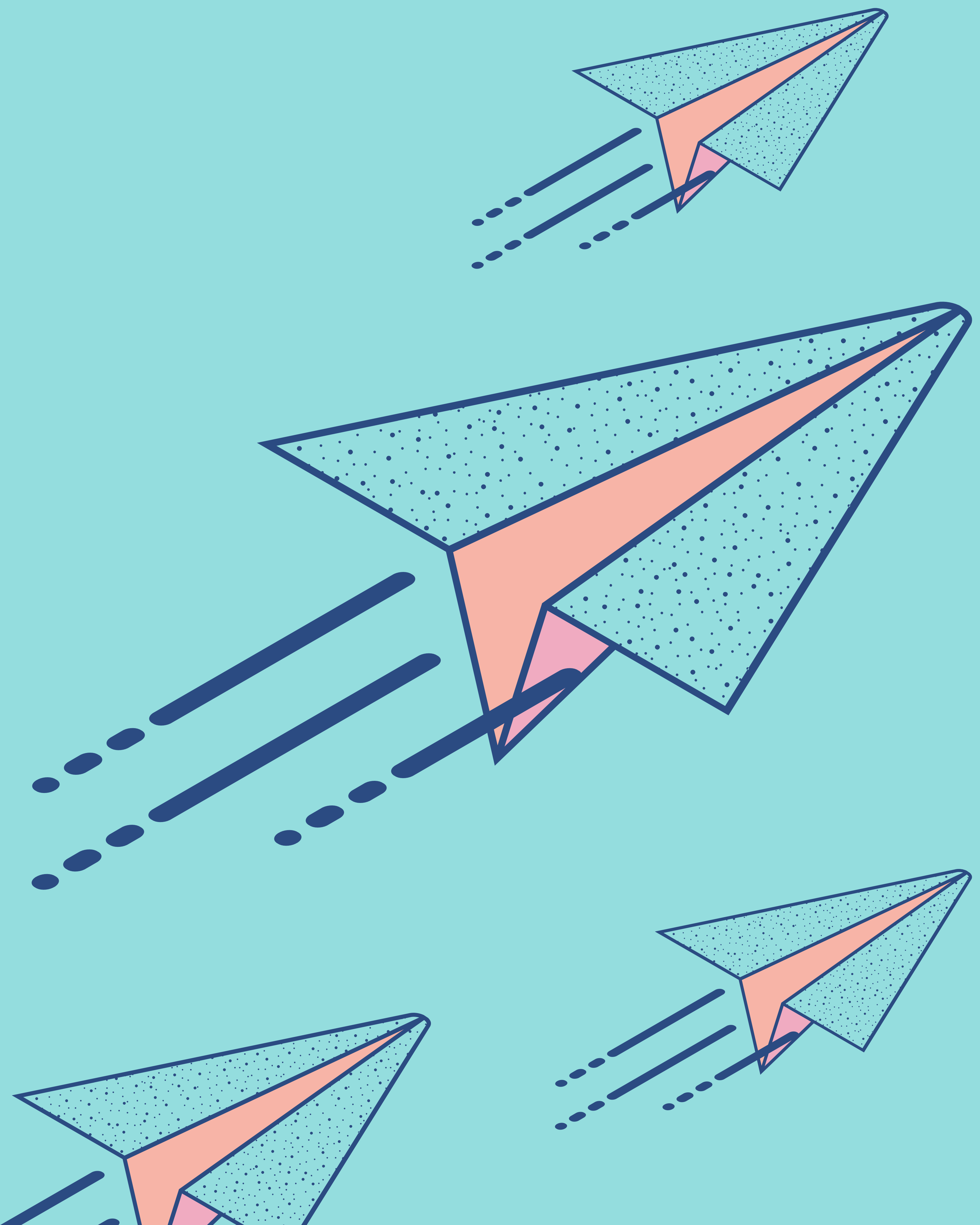
**Plots**



# Do you have any questions?

Ask away!



# Robotics Society Upcoming Events

**Smart Home Workshop**

March 2nd Thursday

**Web Programming Workshop**

March 12th Sunday

**Robotics Workshop**

March 19th Sunday

**3 Day Bootcamp ML vs DL**

TBA

# Goodbye!

You can go now <3