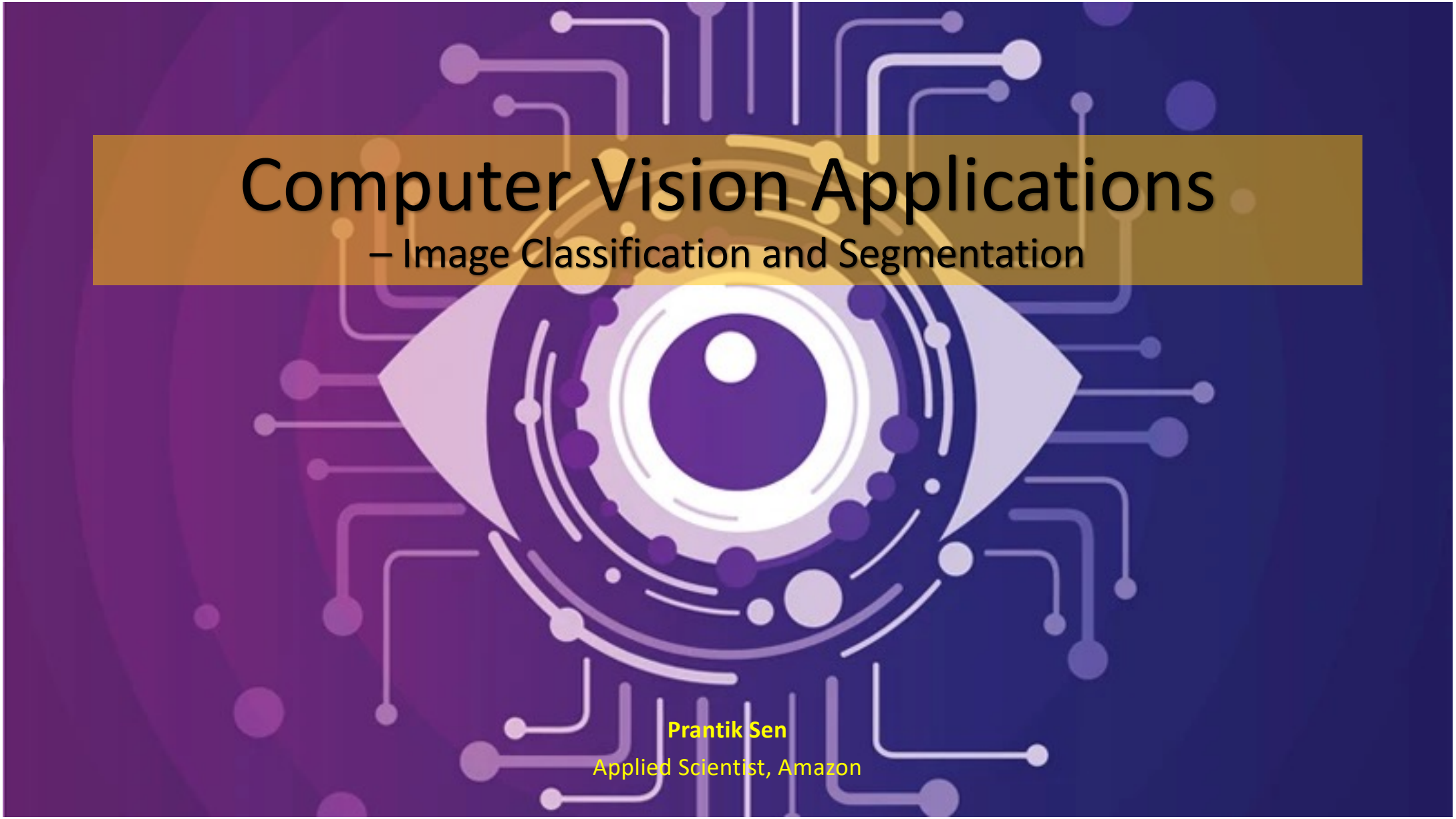


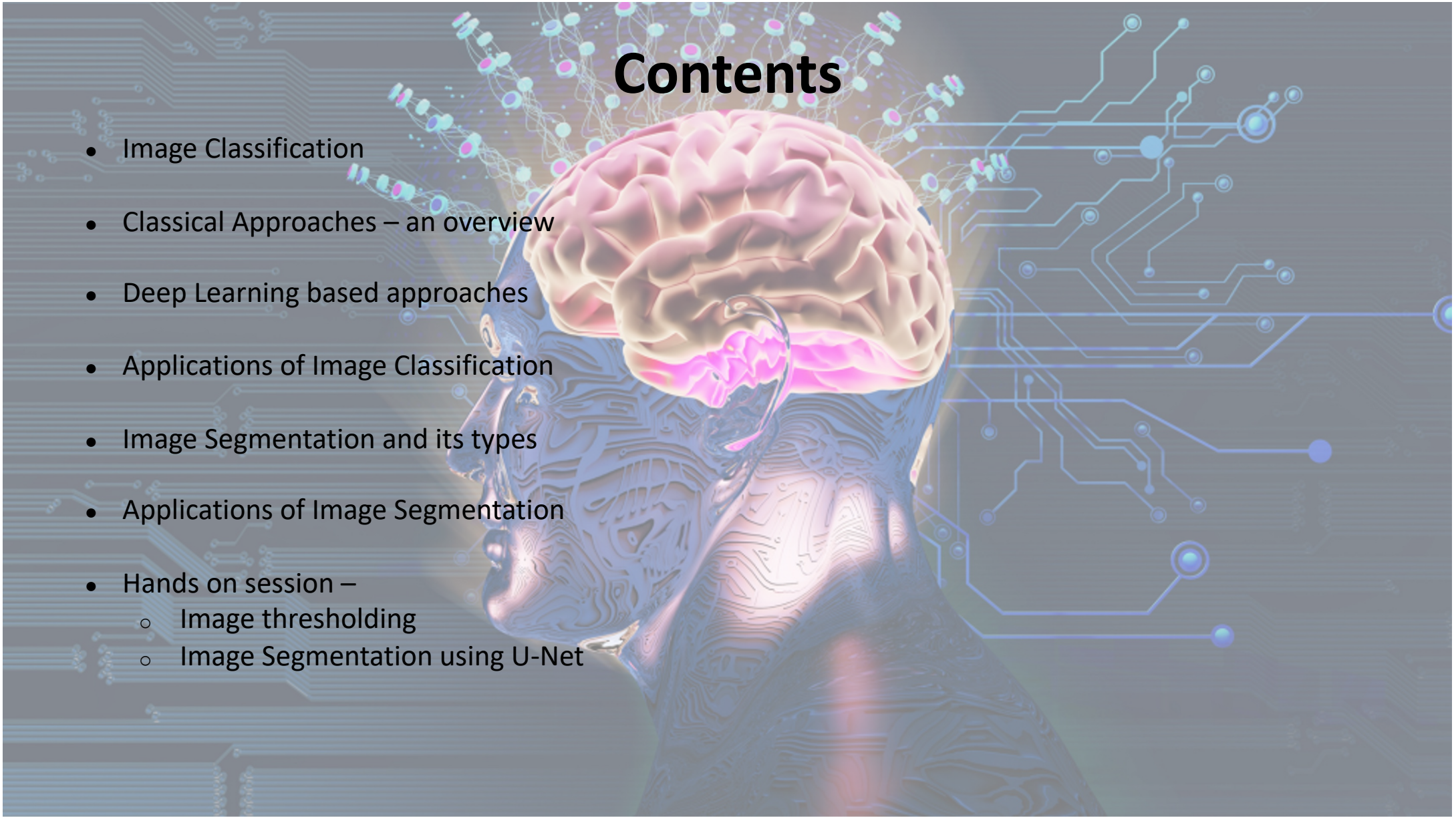
Computer Vision Applications

– Image Classification and Segmentation

Prantik Sen

Applied Scientist, Amazon

Contents



- Image Classification
- Classical Approaches – an overview
- Deep Learning based approaches
- Applications of Image Classification
- Image Segmentation and its types
- Applications of Image Segmentation
- Hands on session –
  - Image thresholding
  - Image Segmentation using U-Net

Image Classification

Given an image, associate it to one or more classes from a discrete set of class labels

Image Classification

Given an image, associate it to one or more classes from a discrete set of class labels

Class Labels – {Dog, Cat, Elephant, Tiger, ….}

Image Classification

Given an image, associate it to one or more classes from a discrete set of class labels

Class Labels – {Dog, Cat, Elephant, Tiger, ….}

Cat

Image Classification

Given an image, associate it to one or more classes from a discrete set of class labels

Class Labels – {Dog, Cat, Elephant, Tiger, ….}

Mathematically, a classification system is nothing but a function ***f***, that transforms an image ***I*** to one or more class labels ***y***

Image Classification

Given an image, associate it to one or more classes from a discrete set of class labels

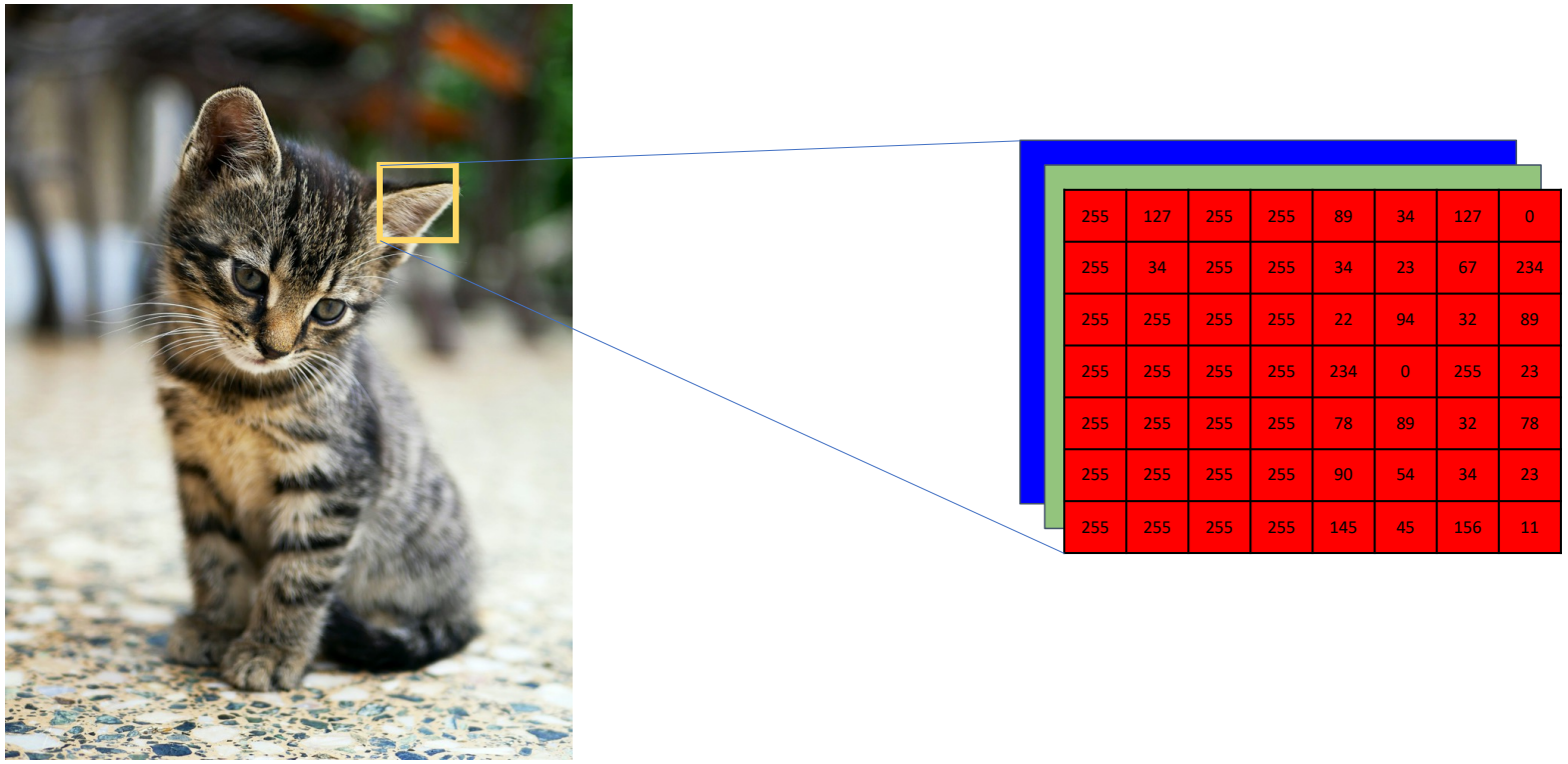
Class Labels – {Dog, Cat, Elephant, Tiger, ….}

Mathematically, a classification system is nothing but a function ***f***, that transforms an image ***I*** to one or more class labels ***y***

$$y = f(I)$$

Image Classification

Given an image, associate it to one or more classes from a discrete set of class labels



Why is Image Classification difficult ?

Viewpoint variation

Why is Image Classification difficult ?

Viewpoint variation

Illumination

Why is Image Classification difficult ?

Viewpoint variation

Illumination

Intraclass Variation

Why is Image Classification difficult ?

Viewpoint variation

Illumination

Intraclass Variation

Background Clutter

Why is Image Classification difficult ?

Viewpoint variation

Illumination

Occlusion

Intraclass Variation

Background Clutter

Why is Image Classification difficult ?

Can we come up with a set of handcrafted features
to represent a cat ?

Why is Image Classification difficult ?

Can we come up with a set of handcrafted features to represent a cat ?

Answer: To an extent

Why is Image Classification difficult ?

Can we come up with a set of handcrafted features
to represent a cat ?

Answer: To an extent

Handcrafted features in an image refer to **manually designed image descriptors** that capture important visual patterns—like edges, textures, corners, colours or shapes

Why is Image Classification difficult ?

Can we come up with a set of handcrafted features to represent a cat ?

Common Handcrafted Feature Types:

1.Edges

1. Detected using **Sobel**, **Canny**, or **Prewitt** filters.

2.Corners and Key-points

1. Detected using **Harris Corner Detector**, **FAST**, or **Shi-Tomasi**.

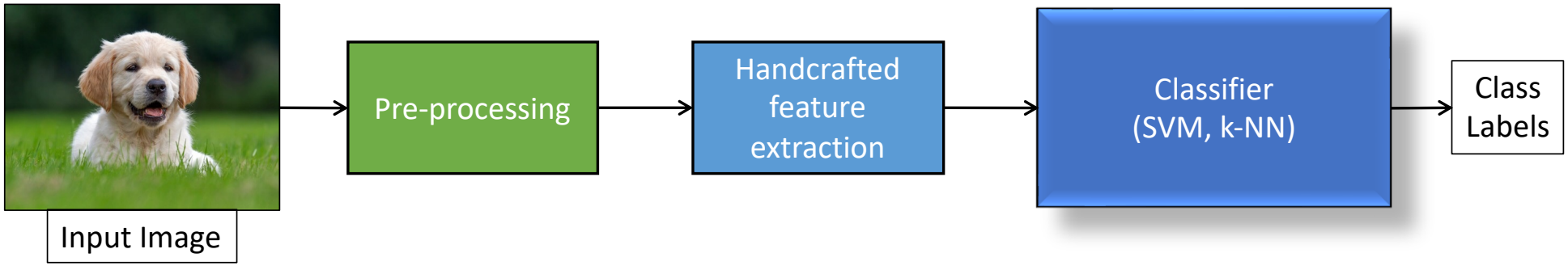
3.Texture

1. Described using **Local Binary Patterns (LBP)** or **Gabor filters**.

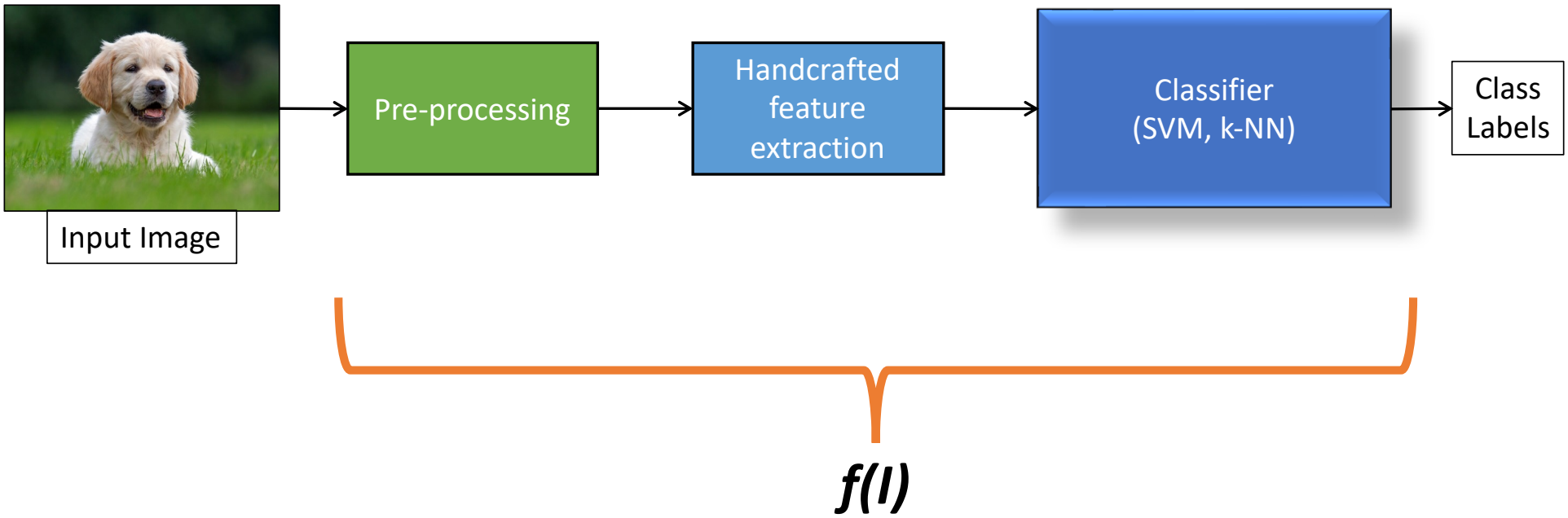
4.Shape Descriptors

1. **Histogram of Oriented Gradients (HOG)**: Captures edge directionality.
2. **SIFT (Scale-Invariant Feature Transform)**: Detects scale- and rotation-invariant features.

Classical Image Classification Pipeline

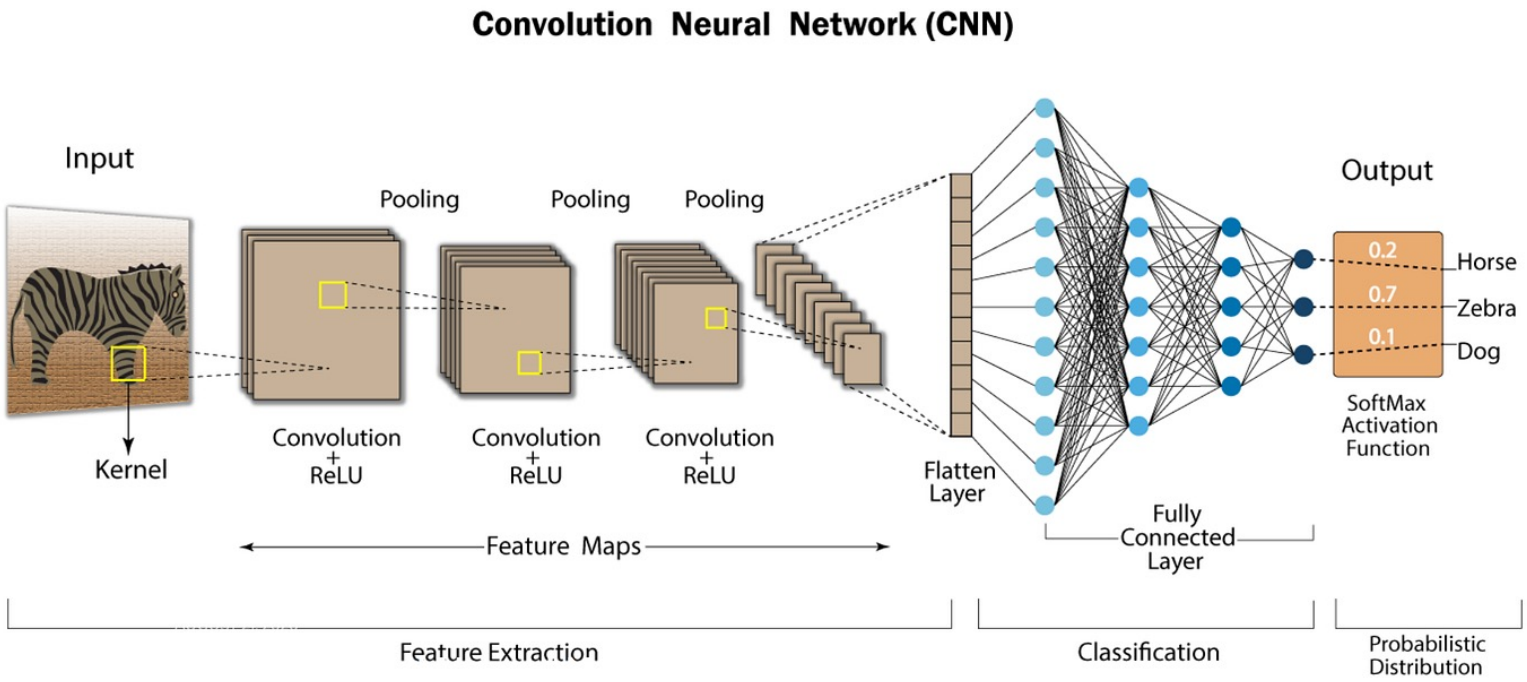


Classical Image Classification Pipeline

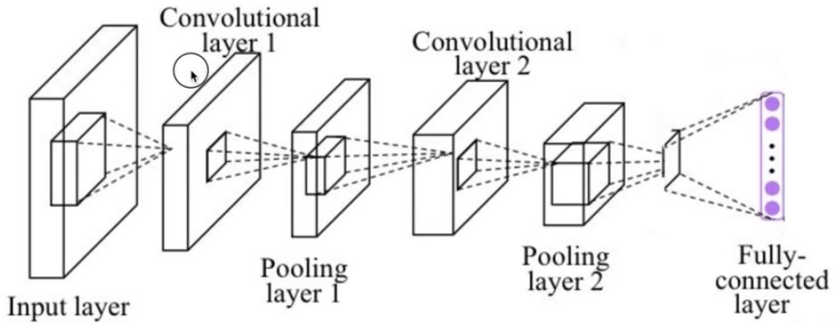


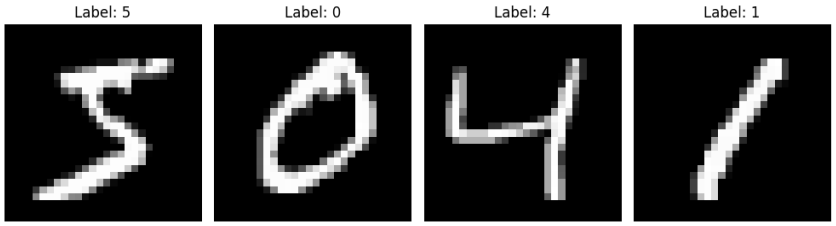
Data Driven Approach

What if I combine all the three blocks below into a single one ?

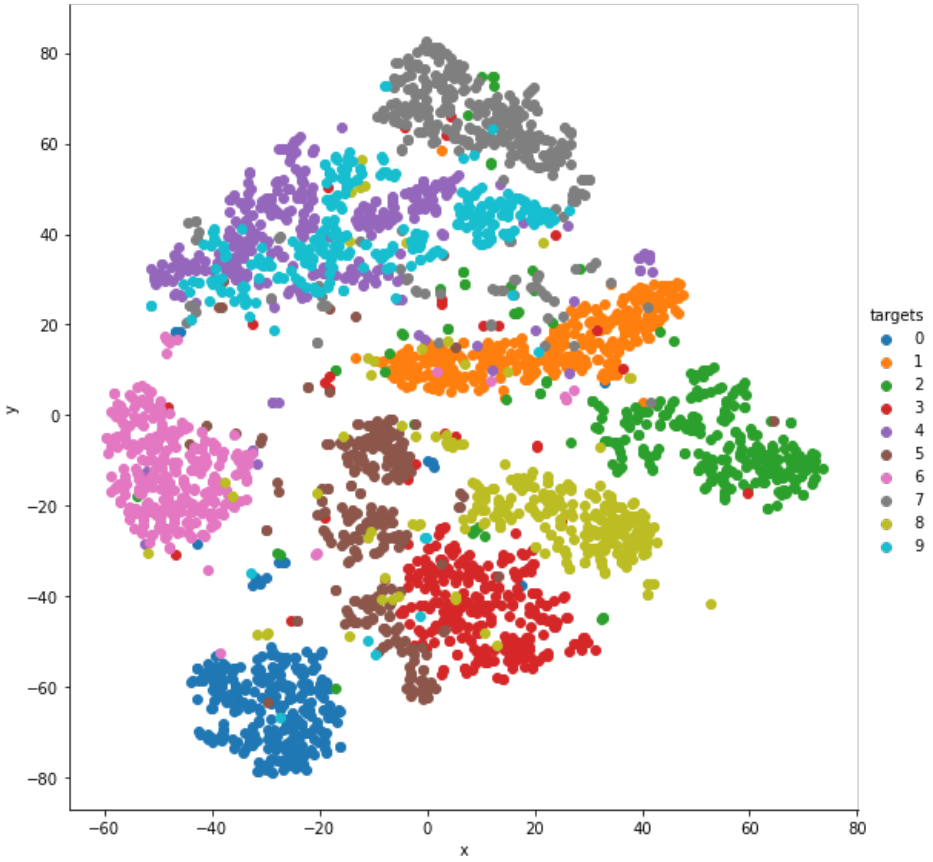


Data Driven Approach





What your Image Classifier is actually doing ?

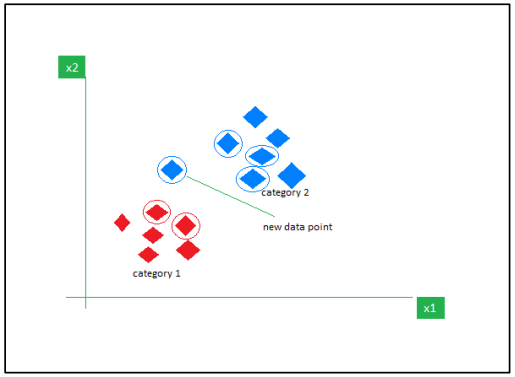


t-SNE Visualization

Evolution of Image Classification

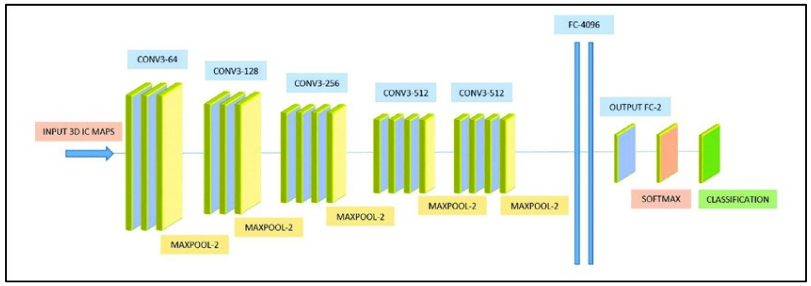
Pre-2012

1. Hand crafted features
2. Classifiers like k-NN, SVM, Random Forests etc.



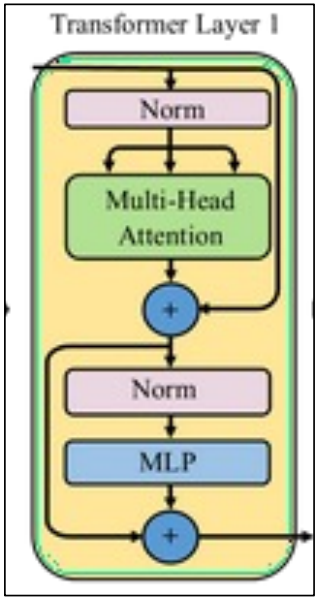
2012-2020

ConvNets



2020-present

Transformers



Applications - Healthcare

(a) Normal

(b) Bacterial Pneumonia

(c) Viral Pneumonia

(d) COVID-19 Pneumonia

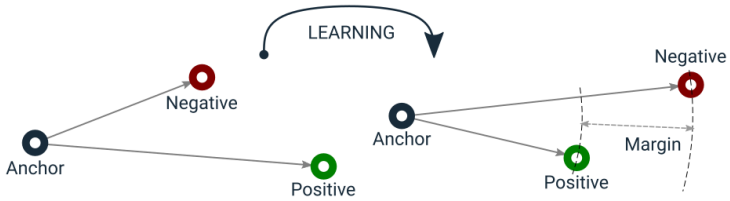
Applications - Retail and E-commerce

Image classification can be used for product categorization, helping retailers organize their inventories and provide better search results to customers.

Applications – Security and Surveillance

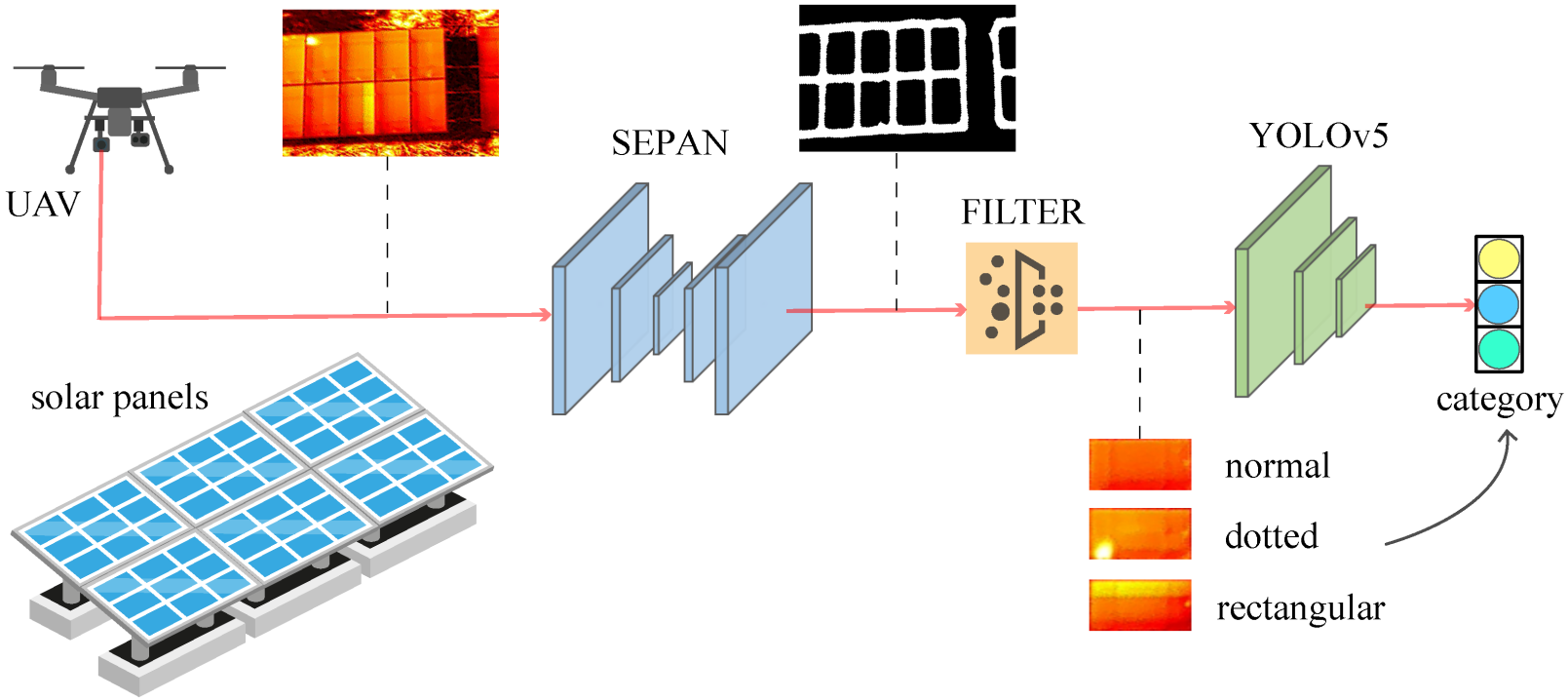
FaceNet - https://arxiv.org/pdf/1503.03832

Applications – Security and Surveillance



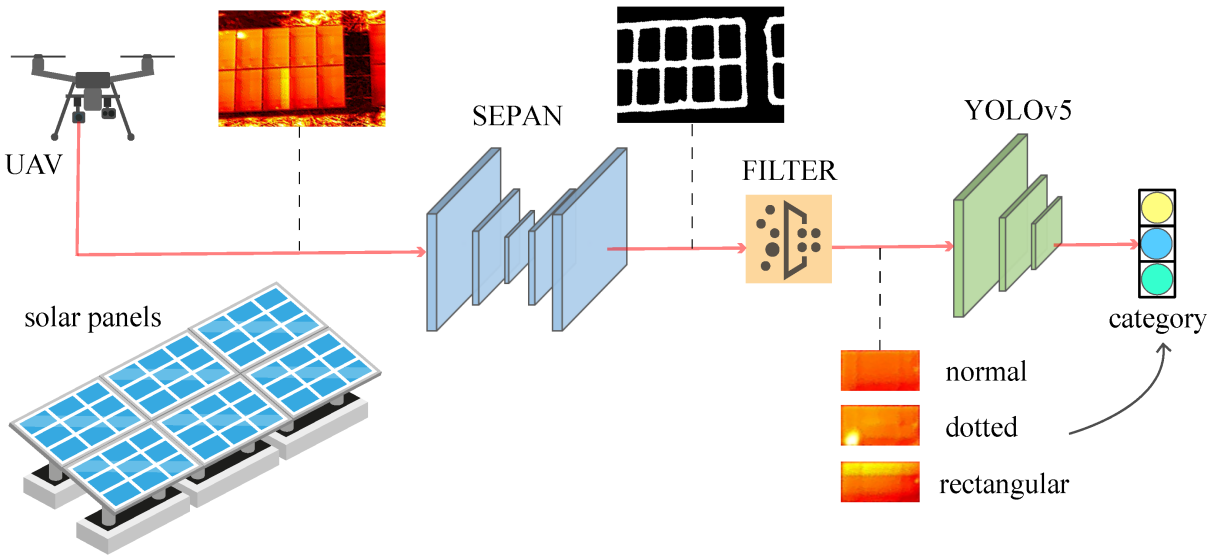
https://arxiv.org/pdf/1503.03832

Applications – Fault Detection



Ling, H.; Liu, M.; Fang, Y. Deep Edge-Based Fault Detection for Solar Panels. *Sensors* **2024**, *24*, 5348. https://doi.org/10.3390/s24165348

Applications – Fault Detection

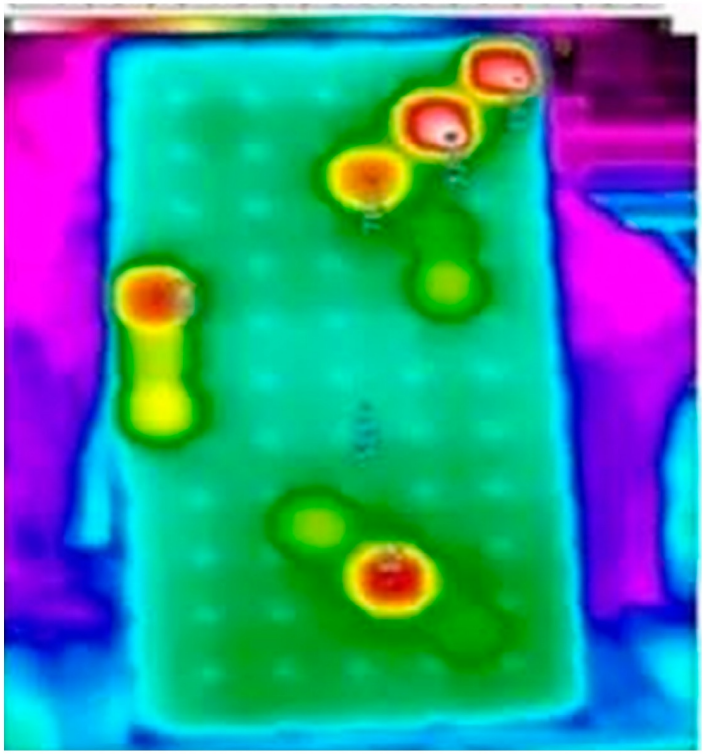


Ling, H.; Liu, M.; Fang, Y. Deep Edge-Based Fault Detection for Solar Panels. *Sensors* **2024**, *24*, 5348. https://doi.org/10.3390/s24165348

IR images are often colorized to enhance visual interpretation. Different palettes are used, with warmer colors (like red, orange, yellow) representing warmer temperatures, and cooler colors (like blue, green) representing cooler temperatures.

Applications – Fault Detection

(a)



(b)

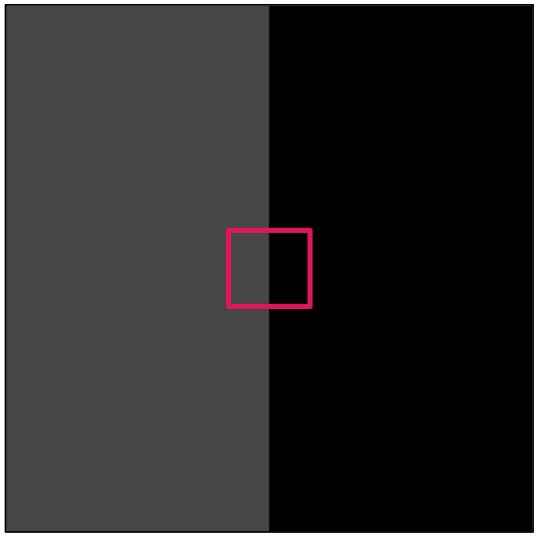
Applications - Fraud Detection

https://tech.olx.com/fighting-fraud-with-triplet-loss-86e5f79c7a3e

Applications – Self driving Car

Credits: auro.ai

What is Convolution ?



70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

How can you detect the edge in the image above ?

What is Convolution ?

+1	0	-1
+1	0	-1
+1	0	-1

kernel

70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

Original Image

Output Image

What is Convolution ?

+1	0	-1
+1	0	-1
+1	0	-1

kernel

70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

Original Image

0					

Output Image

(+1 x 70) + (0 x 70) + (-1 x 70) + (+1 x 70) + (0 x 70) + (-1 x 70) + (+1 x 70) + (0 x 70) + (-1 x 70)

= 70 + 0 - 70 + 70 + 0 - 70 + 70 + 0 - 70

= 0

What is Convolution ?

+1	0	-1
+1	0	-1
+1	0	-1

kernel

70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

Original Image

0	0				

Output Image

What is Convolution ?

+1	0	-1
+1	0	-1
+1	0	-1

kernel

70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

Original Image

0	0	210			

Output Image

What is Convolution ?

+1	0	-1
+1	0	-1
+1	0	-1

kernel

70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

Original Image

0	0	210	210		

Output Image

What is Convolution ?

+1	0	-1
+1	0	-1
+1	0	-1

kernel

70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

Original Image

0	0	210	210	0	

Output Image

What is Convolution ?

+1	0	-1
+1	0	-1
+1	0	-1

kernel

70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

Original Image

0	0	210	210	0	0

Output Image

What is Convolution ?

+1	0	-1
+1	0	-1
+1	0	-1

kernel

70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

Original Image

0	0	210	210	0	0
0					

Output Image

What is Convolution ?

+1	0	-1
+1	0	-1
+1	0	-1

kernel

Prewitt filter

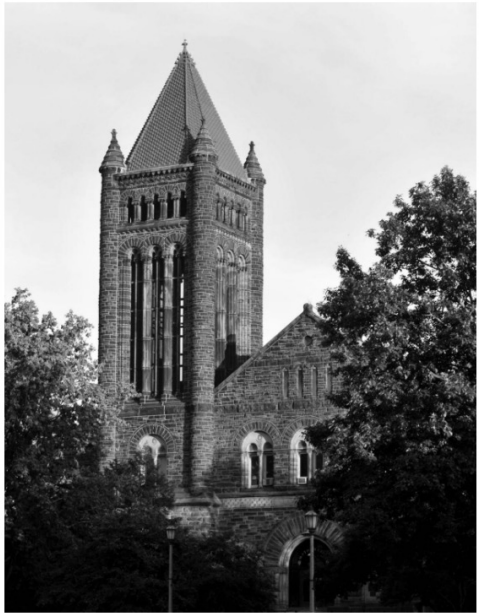
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0
70	70	70	70	0	0	0	0

Original Image

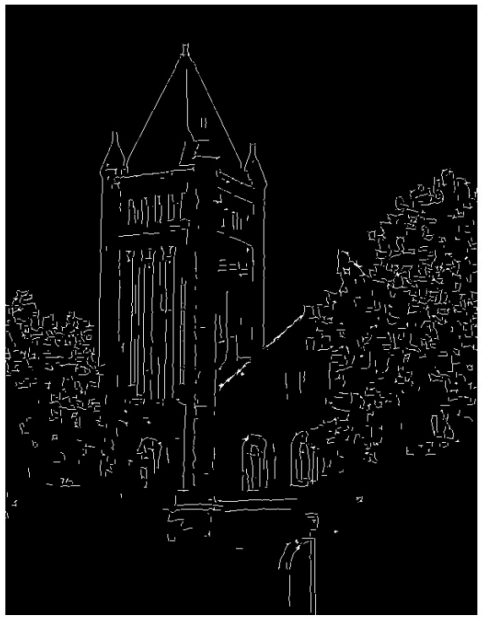
0	0	210	210	0	0
0	0	210	210	0	0
0	0	210	210	0	0
0	0	210	210	0	0
0	0	210	210	0	0
0	0	210	210	0	0

Output Image

Edge Detection



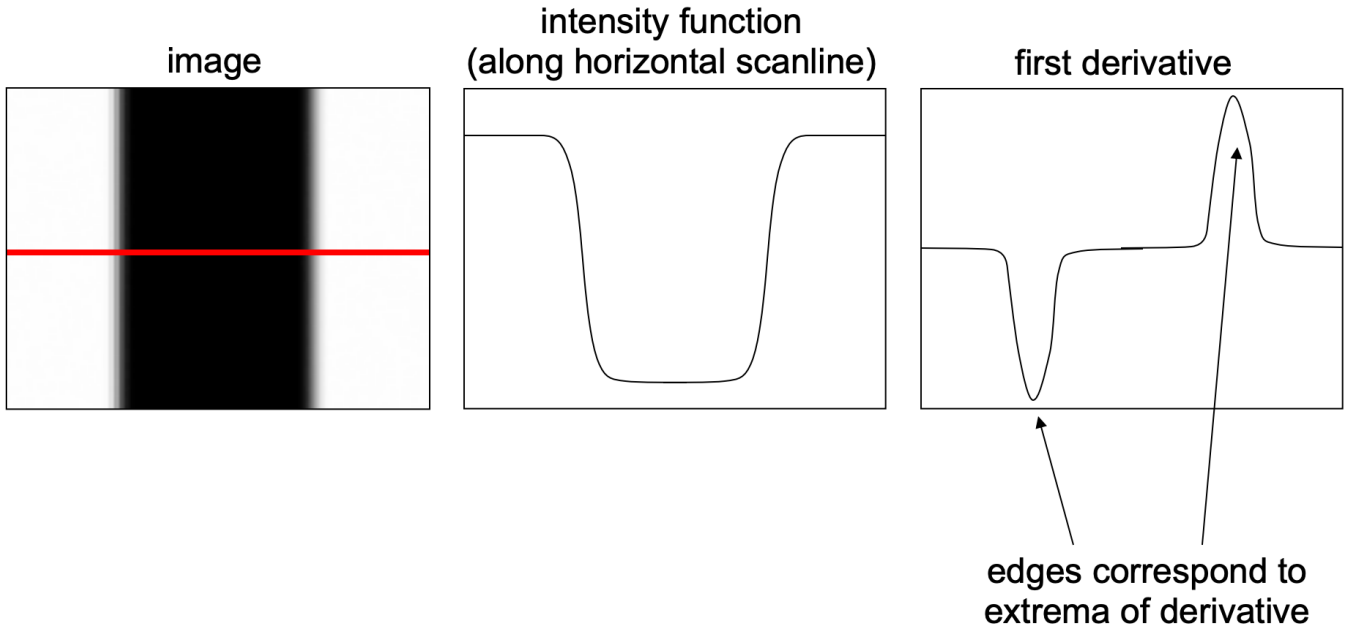
Ideal: artist's line drawing



Reality

Edge Detection

- An edge is a place of rapid change in the image intensity function



Edge Detection

Partial derivatives of an image

Which shows changes with respect to x?

Image Segmentation

Image Segmentation - Objective

Group similar pixels into meaningful regions

Image Segmentation - Objective

What do you see here ?

Image Segmentation - Objective

How can a computer understand this image ?

Image Segmentation - Objective

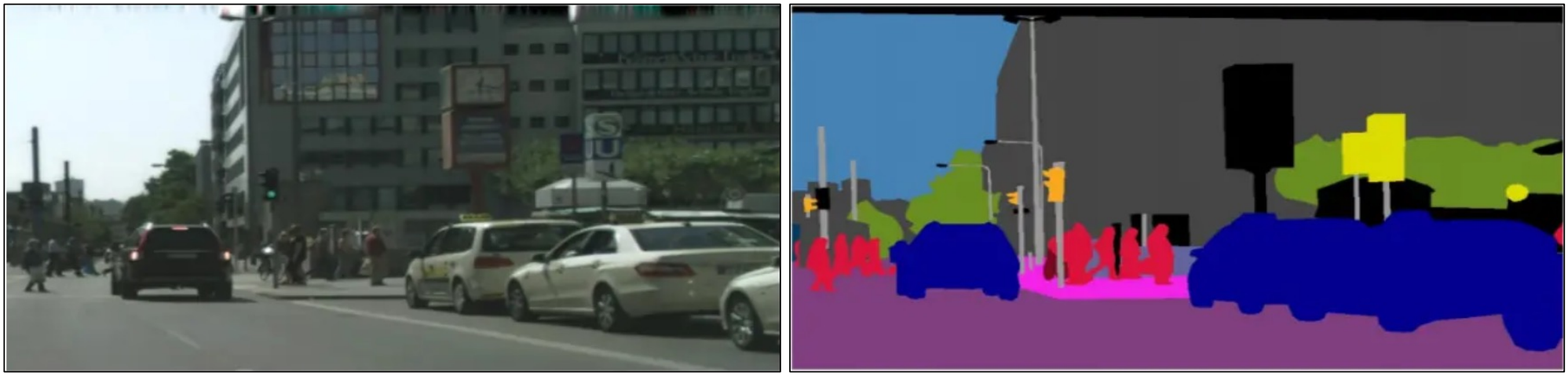
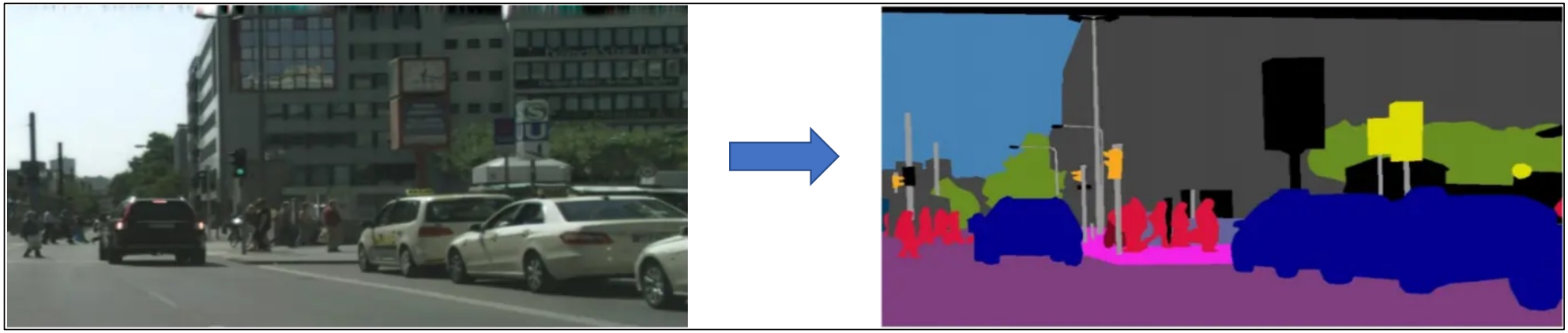


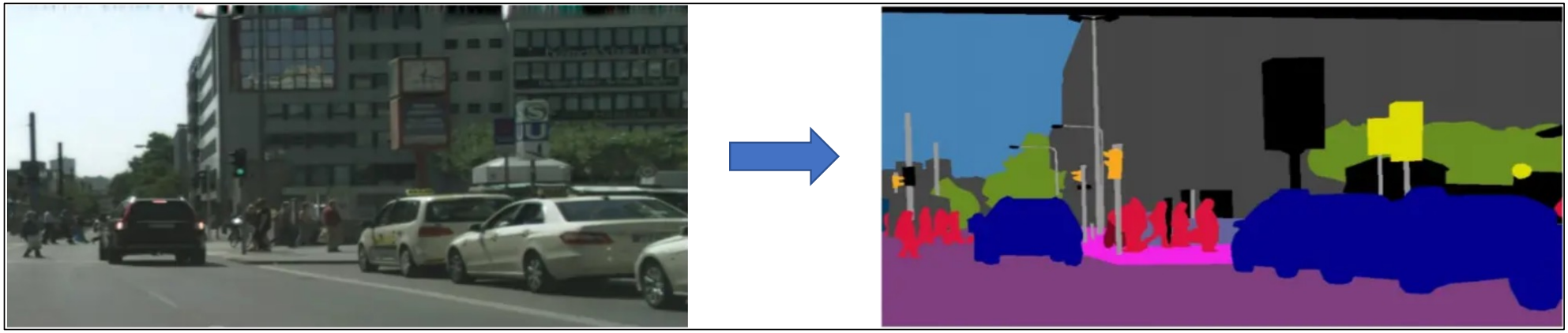
Image Segmentation: group similar pixels into meaningful regions

Types of Image Segmentation



Semantic Segmentation: What category is each pixel?

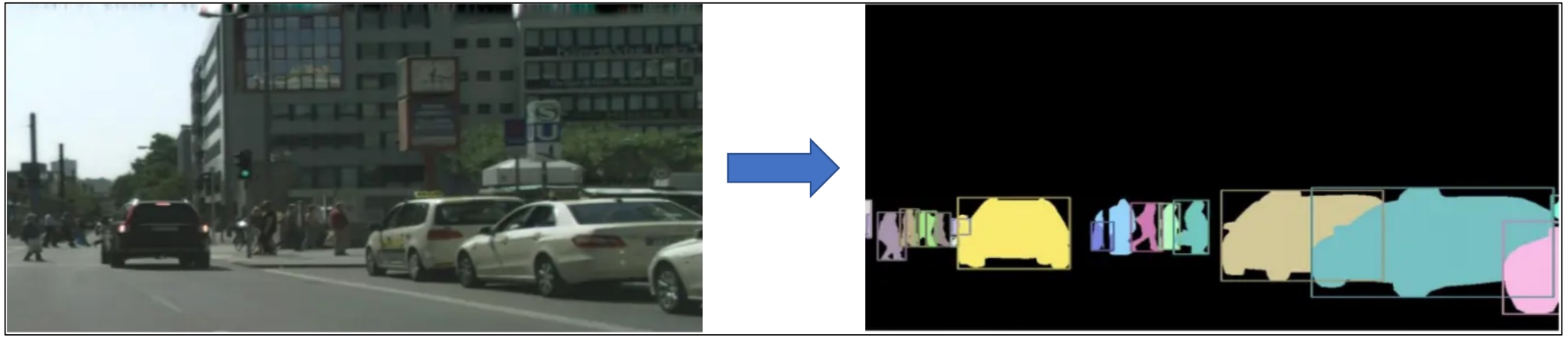
Types of Image Segmentation



Semantic Segmentation: What category is each pixel?

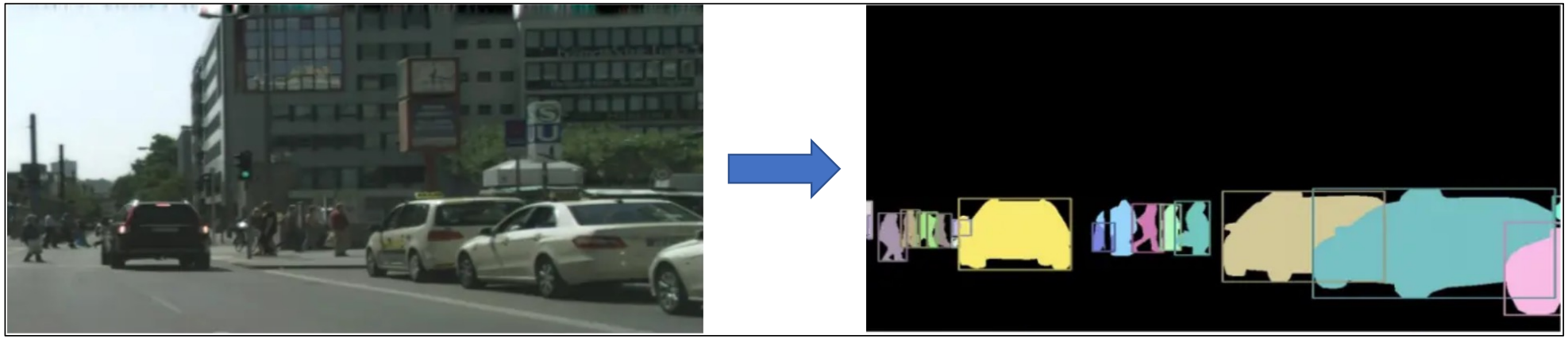
The algorithm only cares about the category and not individual instances

Types of Image Segmentation



Instance Segmentation: Which specific object does each pixel belong to?

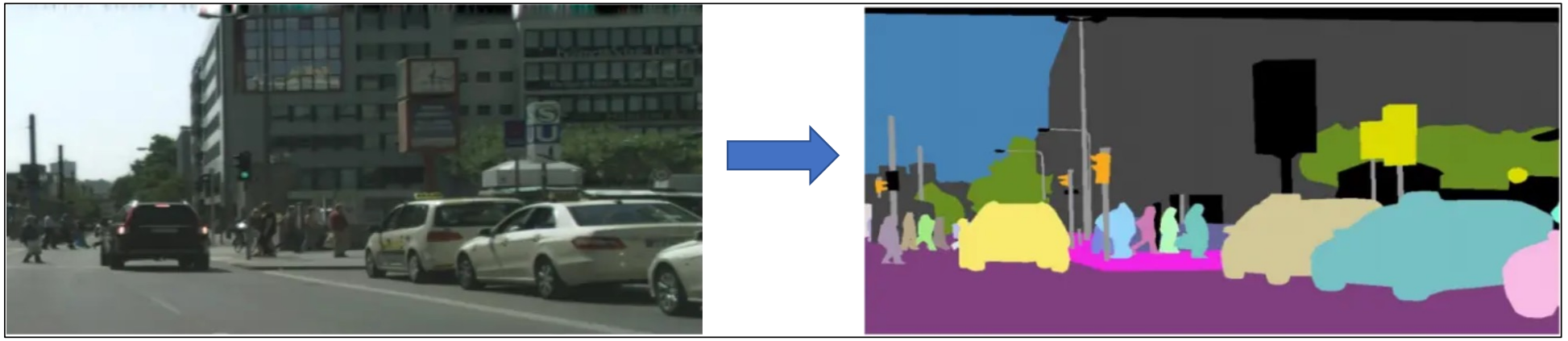
Types of Image Segmentation



Instance Segmentation: Which specific object does each pixel belong to?

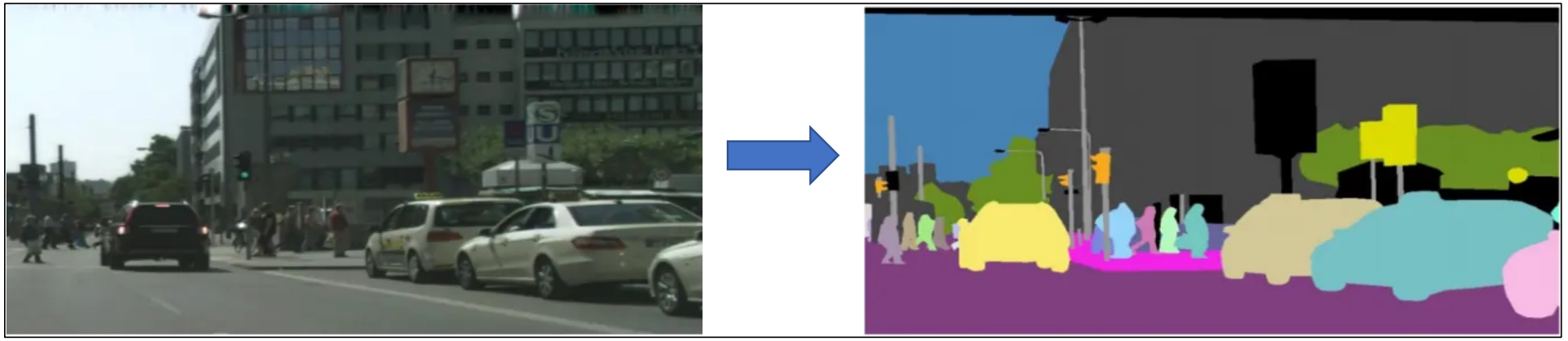
Now each car has its own colour! Car #1 is yellow, Car #2 is light blue, Car #3 is pink, etc.

Types of Image Segmentation



Panoptic Segmentation: complete scene understanding - What + Which

Types of Image Segmentation



Panoptic Segmentation: complete scene understanding - What + Which

This is what self-driving cars need - they must know there's a road (semantic) AND identify each specific car to avoid collision (instance)

Evolution of Image Segmentation

1970 - 80s

- Manual segmentation
- Basic thresholding
- Edge detection

1990s

- Region growing algorithms
- Watershed segmentation
- Active contours

2000s

- Graph cuts
- GrabCut
- Random walker

2024 -

- Foundation Models

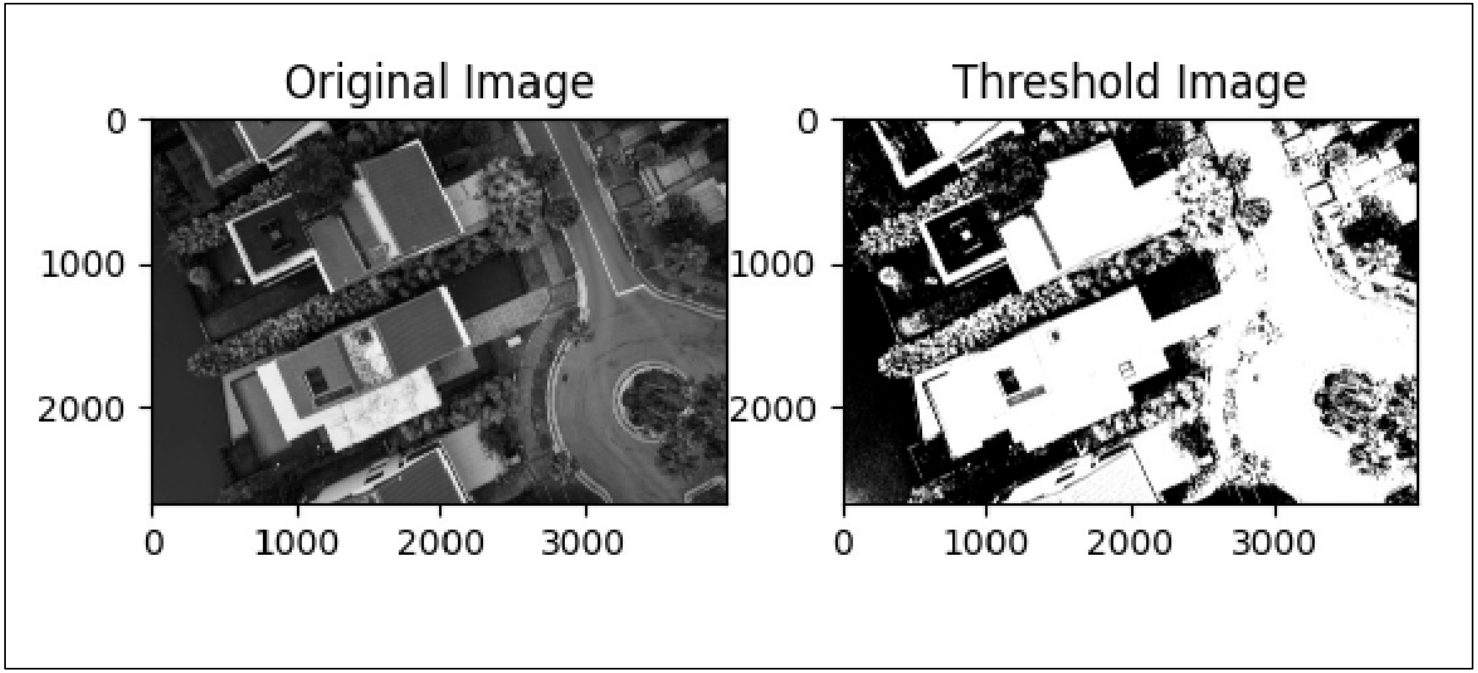
2020s

- Transformer-based methods
- Vision Transformer

2010s

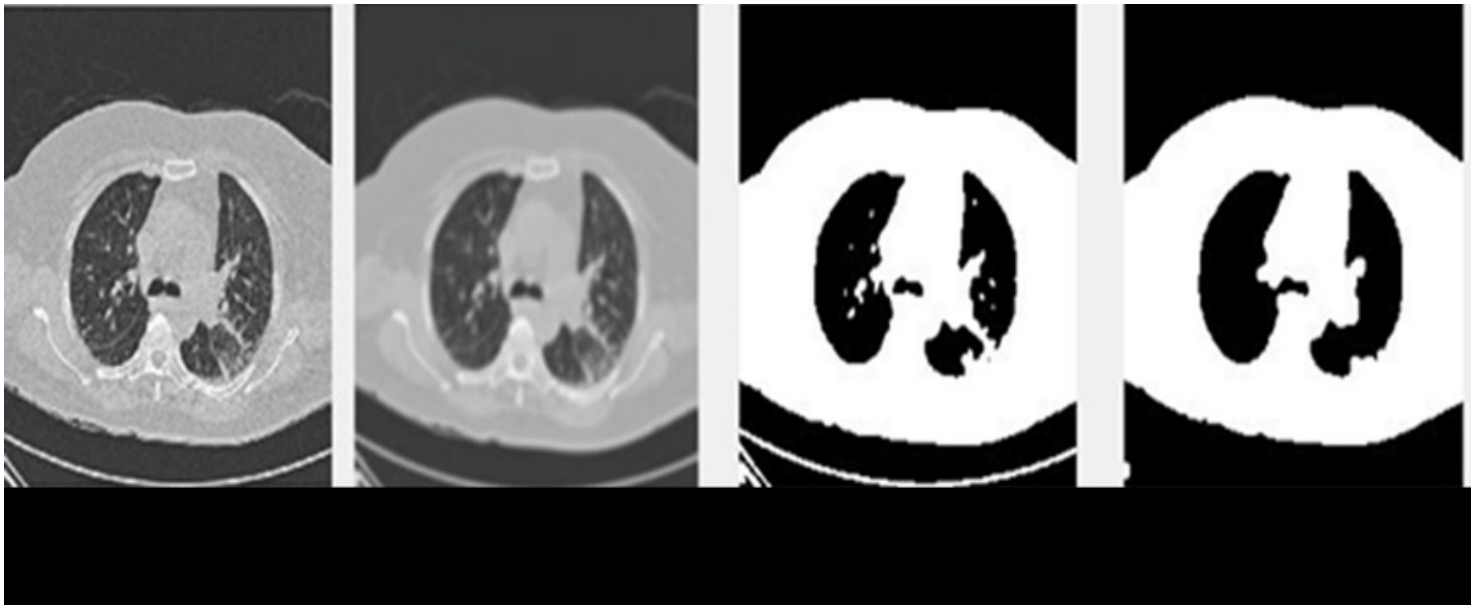
- Machine learning approaches
- Deep learning emergence

Image Thresholding



- **Simple rule:** “If pixel is brighter than threshold value, it is important”
- **Binary decision:** Above threshold = object, below = background

Image Thresholding



$$b(x,y) = \begin{cases} 1 & \text{if } I(x,y) > T \\ 0 & \text{otherwise} \end{cases}$$

Image Thresholding

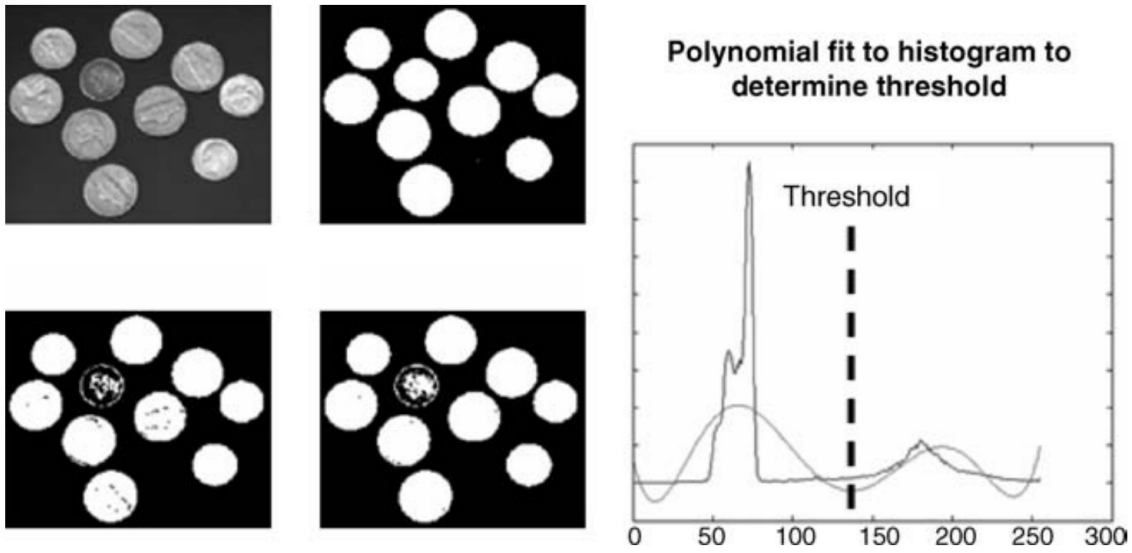
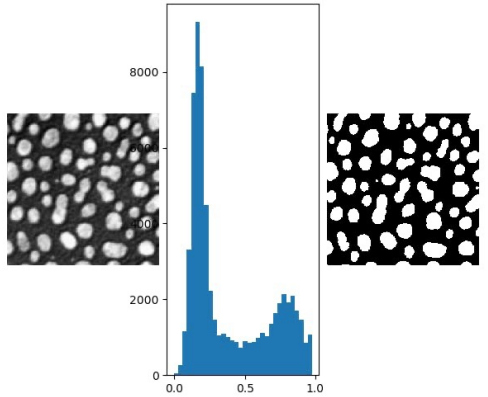
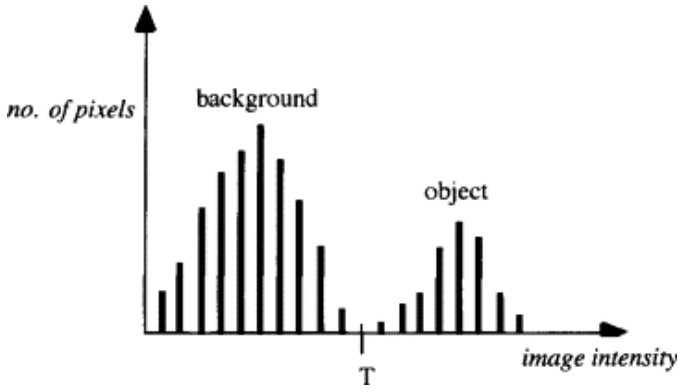


Figure 10.1 Top left: original image. Top right: result after manual selection of threshold. Bottom left: result of automatic threshold selection by polynomial fitting to image histogram. Bottom right: result of automatic threshold selection using Otsu's method. The image on the right shows the histogram and the result of fitting a sixth-order polynomial

How to fix threshold ?

- **Use Domain Knowledge: "**Bones are bright on X-rays" => threshold separates bone from tissue
- **Look at the Histogram:** Find the "valley" between two
- **Let Algorithms Decide (Otsu's Method):** Computer finds optimal threshold automatically
- **Consider Local vs Global:** One threshold for whole image vs different thresholds for different regions

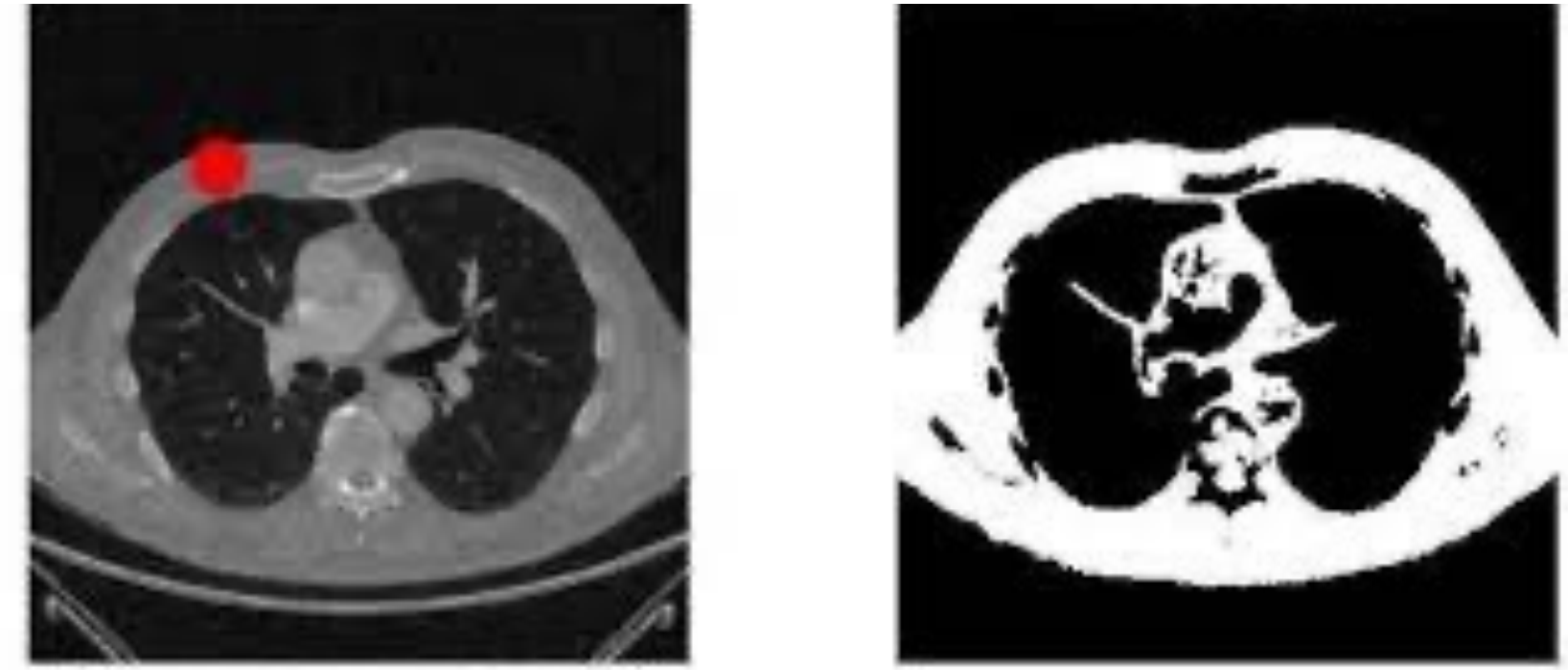




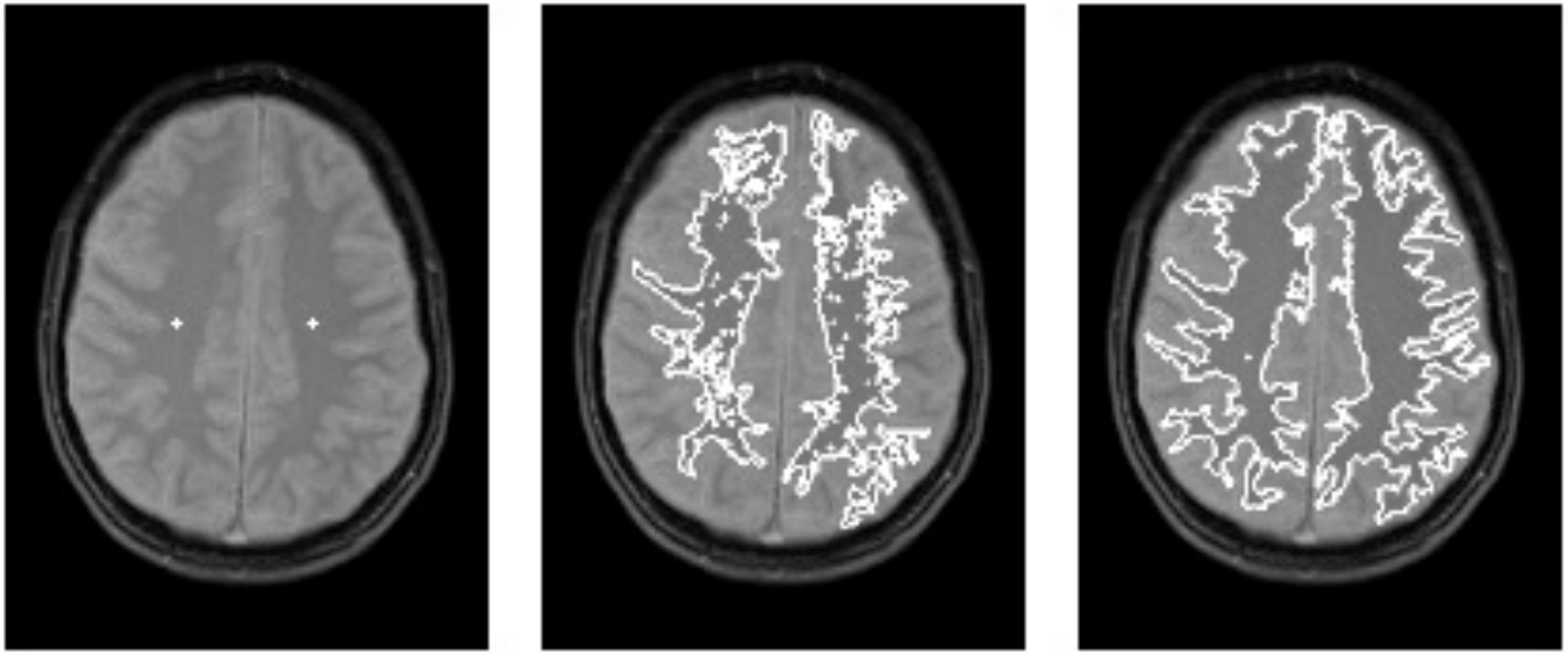
Thresholding Limitations

- **Poor Contrast Images:** When foreground and background have similar intensities
- **Noise and Artifacts:** Salt-and-pepper noise creates false positives/negatives
- **Uneven Lighting:** Same object appears bright on one side, dark on another
- **Overlapping Intensity Ranges:** Different tissues with same pixel values
- **Complex Scenes:** Multiple objects with different optimal thresholds

Region Growing



Region Growing

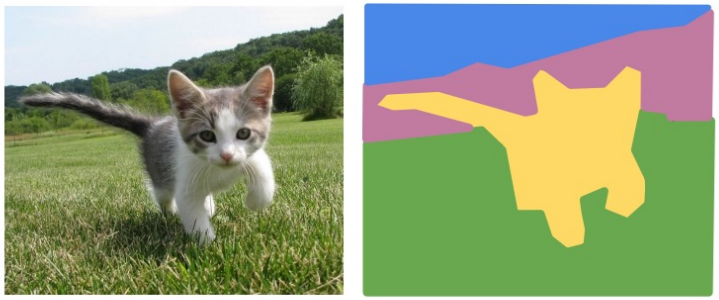


Region Growing

- **Start with Seed Points:**
  - User clicks on pixels of interest (or algorithm picks them)
  - Seeds act as “starting points” for each region

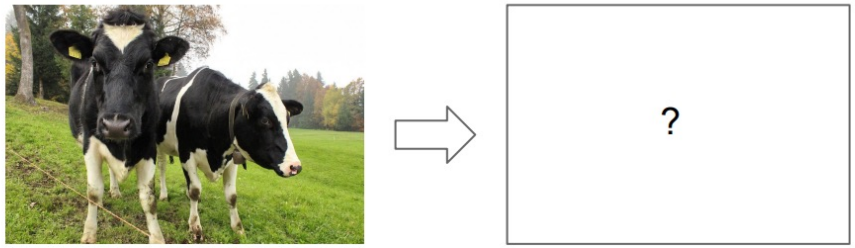
- **Grow Based on Similarity**
  - Check neighbouring pixels: "Are you similar to the seed?”
  - If yes, add to region; if no, stop growing
  - Similarity = intensity, colour, texture, or other features

Image Segmentation using Deep Learning



GRASS, **CAT**,
TREE, **SKY**, **...**

Paired training data: for each training image,
each pixel is labeled with a semantic category.



At test time, classify each pixel of a new image.

Image Segmentation using Deep Learning

Semantic Segmentation Idea: Sliding Window

Image Segmentation using Deep Learning

Semantic Segmentation Idea: Sliding Window

Full image

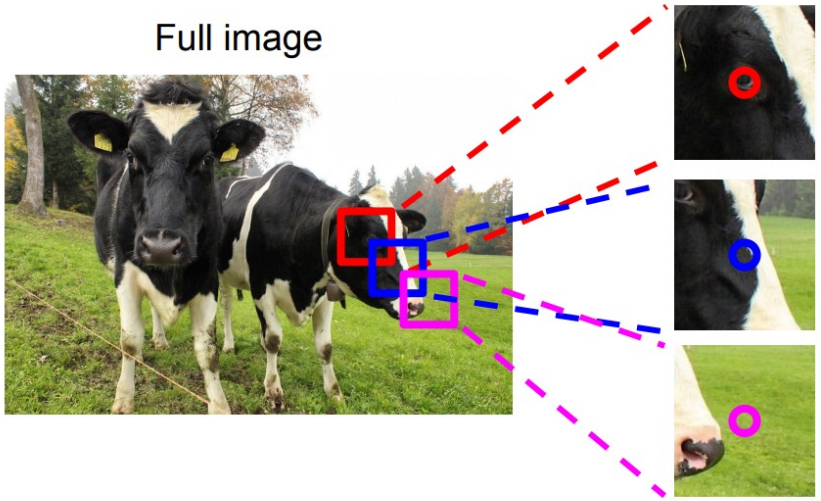
?

Impossible to classify without context

Q: how do we include context?

Image Segmentation using Deep Learning

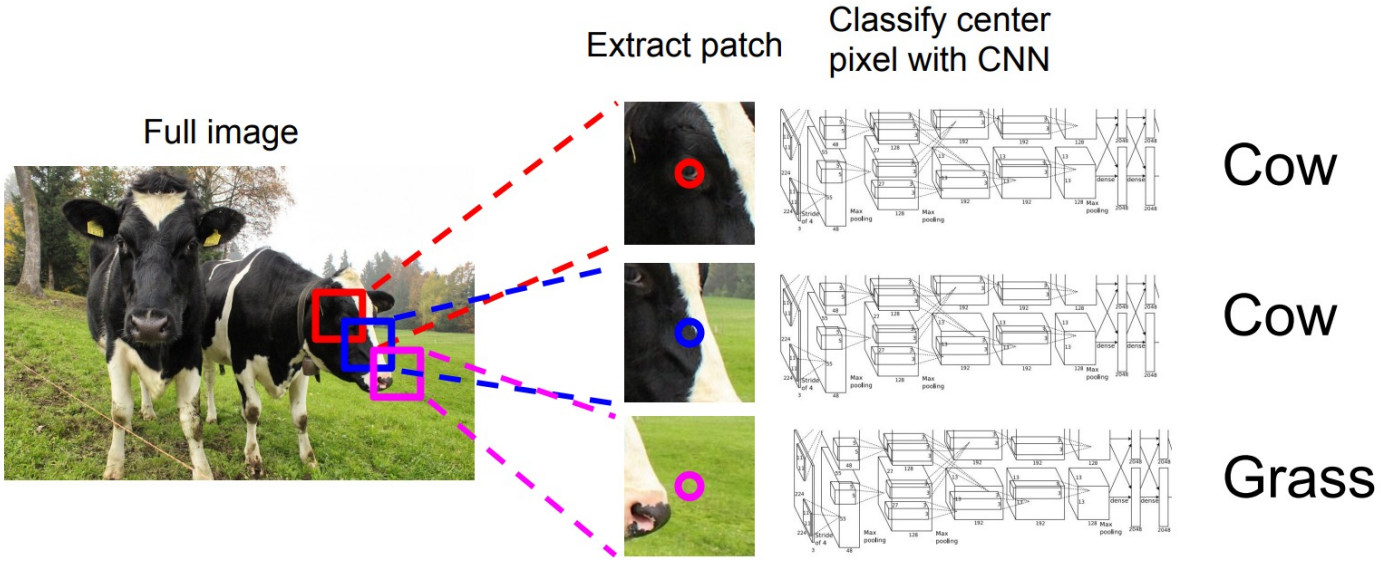
Semantic Segmentation Idea: Sliding Window



Q: how do we model this?

Image Segmentation using Deep Learning

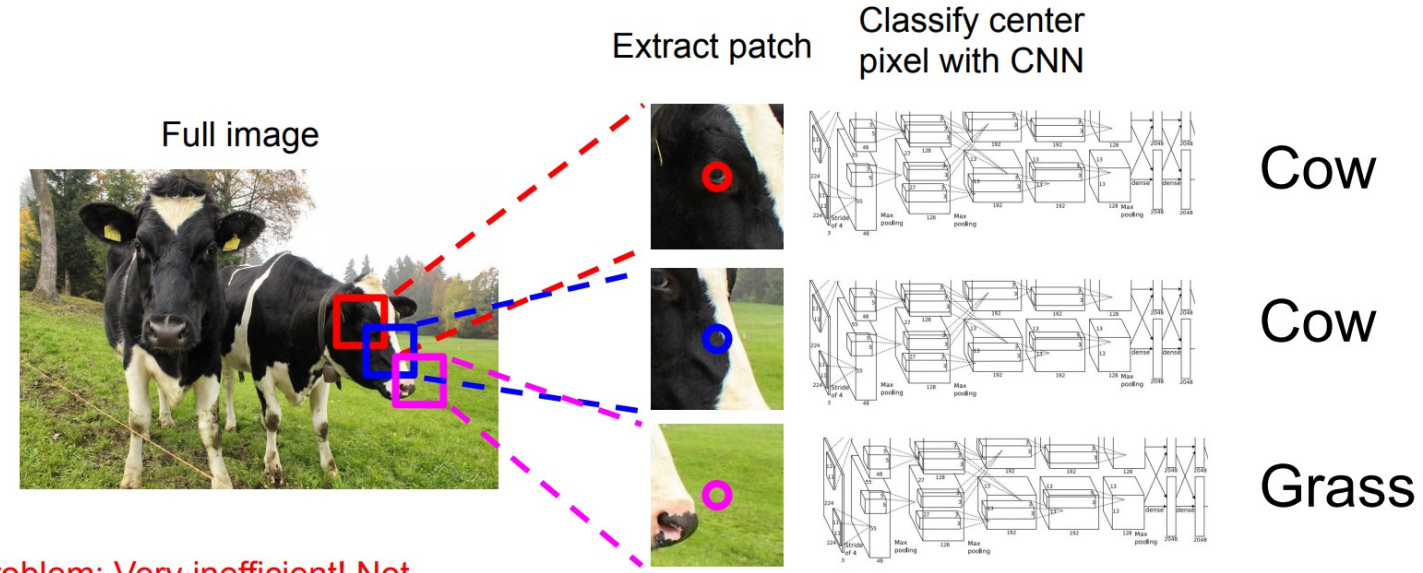
Semantic Segmentation Idea: Sliding Window



Farabet et al, “Learning Hierarchical Features for Scene Labeling,” TPAMI 2013
Pinheiro and Collobert, “Recurrent Convolutional Neural Networks for Scene Labeling”, ICML 2014

Image Segmentation using Deep Learning

Semantic Segmentation Idea: Sliding Window



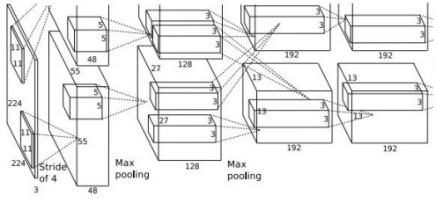
Problem: Very inefficient! Not reusing shared features between overlapping patches

Farabet et al, "Learning Hierarchical Features for Scene Labeling," TPAMI 2013
Pinheiro and Collobert, "Recurrent Convolutional Neural Networks for Scene Labeling", ICML 2014

Image Segmentation using Deep Learning

Semantic Segmentation Idea: Convolution

Full image



An intuitive idea: encode the entire image with conv net, and do semantic segmentation on top.

Problem: classification architectures often reduce feature spatial sizes to go deeper, but semantic segmentation requires the output size to be the same as input size.

Image Segmentation using Deep Learning

Semantic Segmentation Idea: Fully Convolutional

Design a network with only convolutional layers without downsampling operators to make predictions for pixels all at once!

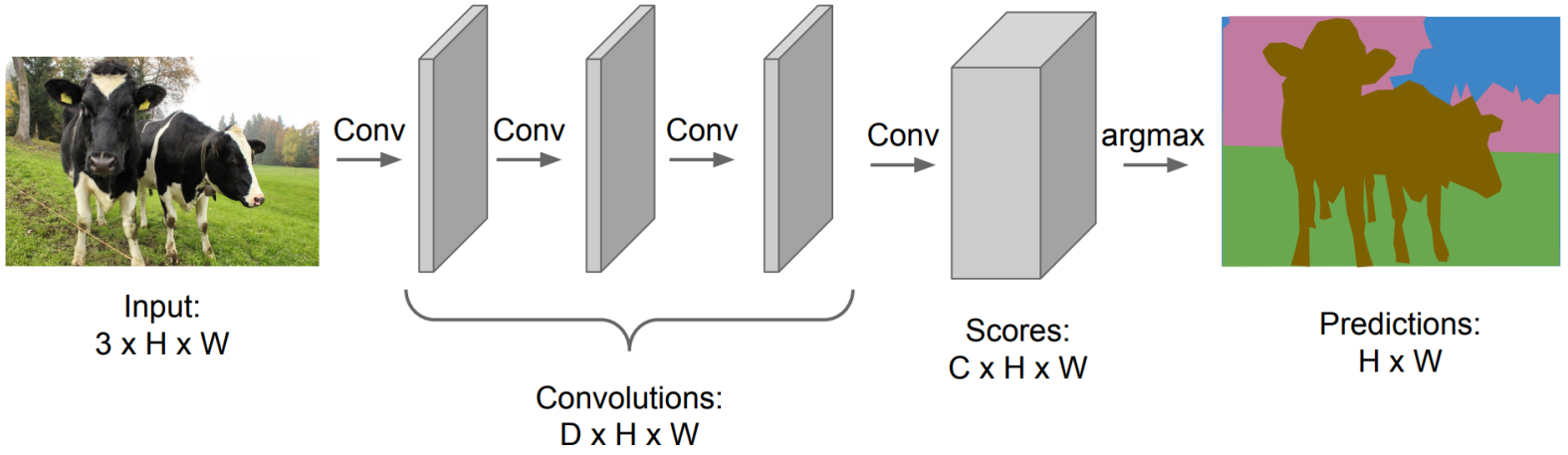


Image Segmentation using Deep Learning

Semantic Segmentation Idea: Fully Convolutional

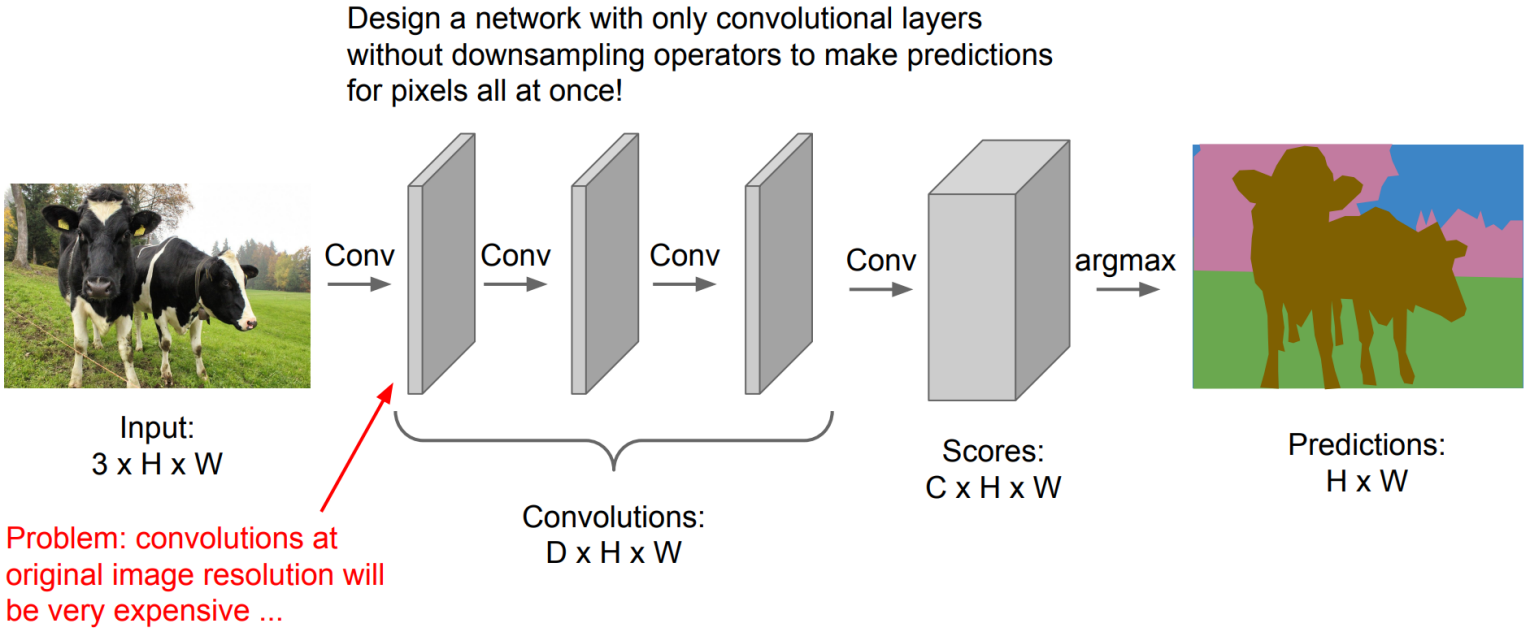
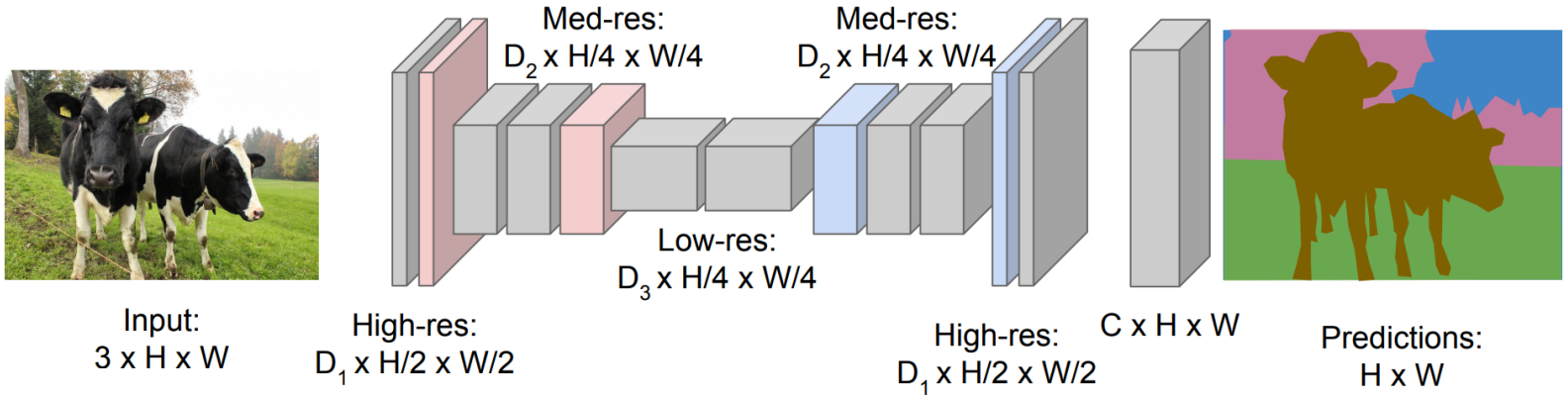


Image Segmentation using Deep Learning

Semantic Segmentation Idea: Fully Convolutional

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!



.ong, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Image Segmentation using Deep Learning

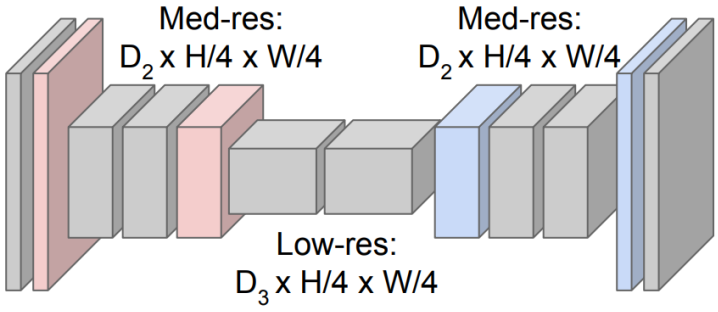
Semantic Segmentation Idea: Fully Convolutional

Downsampling:
Pooling, strided convolution

Design network as a bunch of convolutional layers, with **downsampling** and **upsampling** inside the network!

Upsampling:
???

Input:
3 x H x W



High-res:
D_1 x H/2 x W/2

High-res:
D_1 x H/2 x W/2

C x H x W

Predictions:
H x W

Long, Shelhamer, and Darrell, "Fully Convolutional Networks for Semantic Segmentation", CVPR 2015
Noh et al, "Learning Deconvolution Network for Semantic Segmentation", ICCV 2015

Image Segmentation using Deep Learning

In-Network upsampling: “Unpooling”

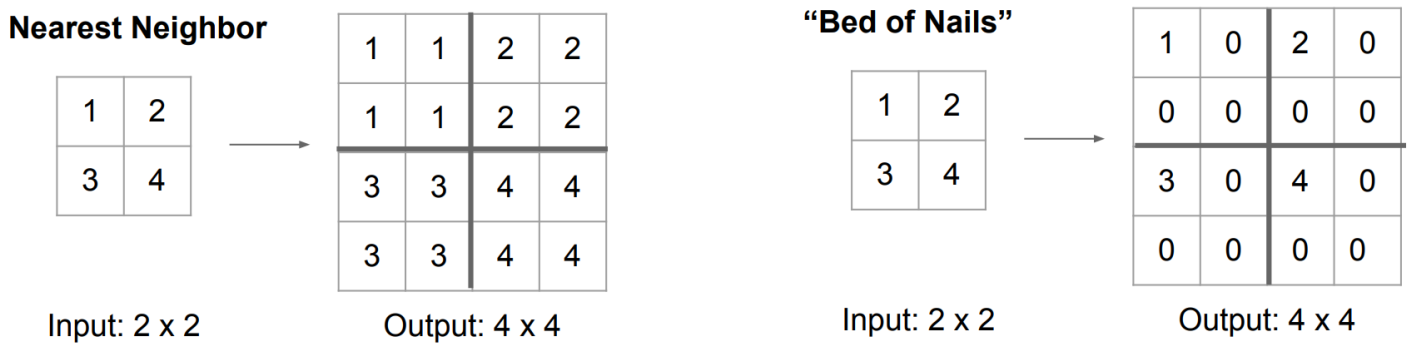


Image Segmentation using Deep Learning

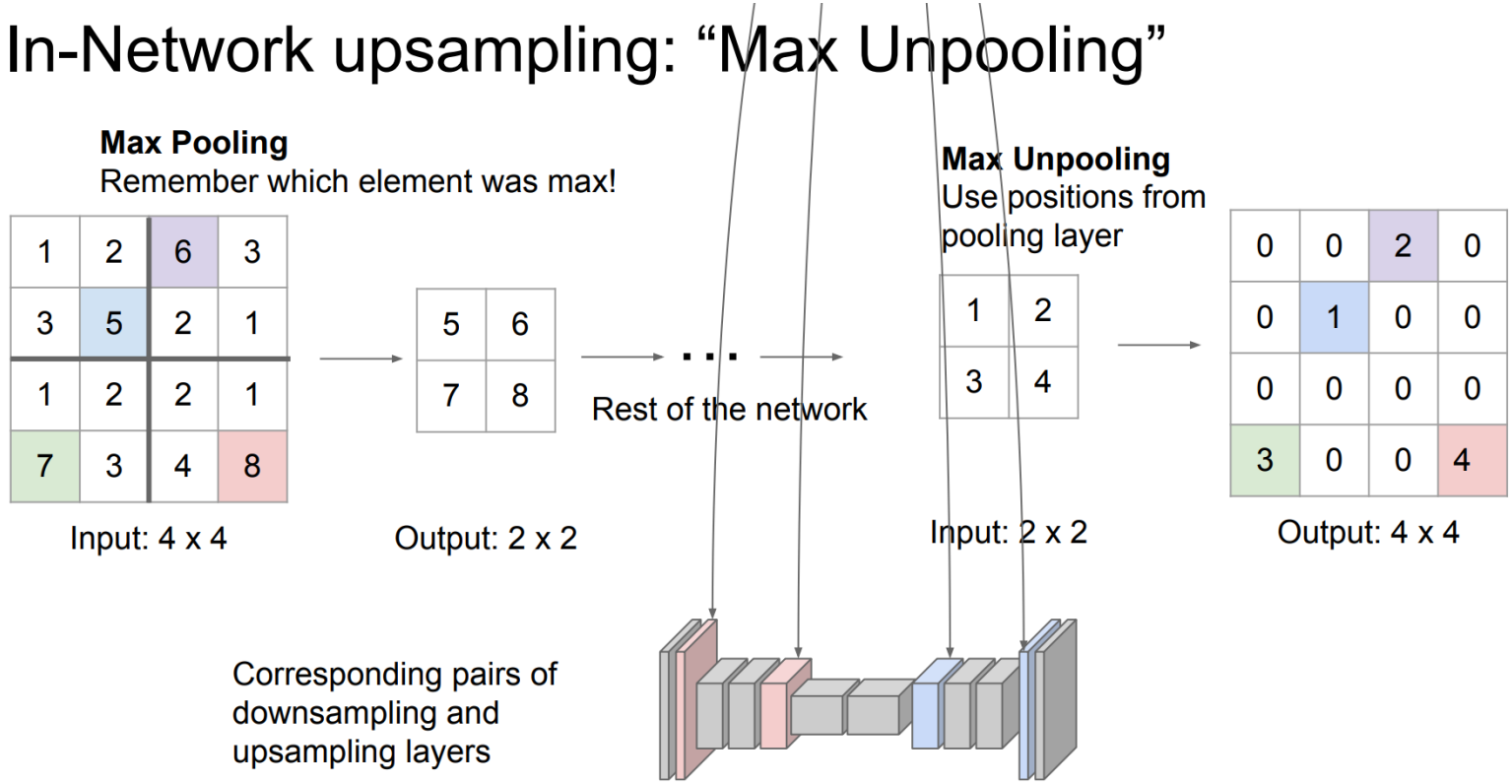
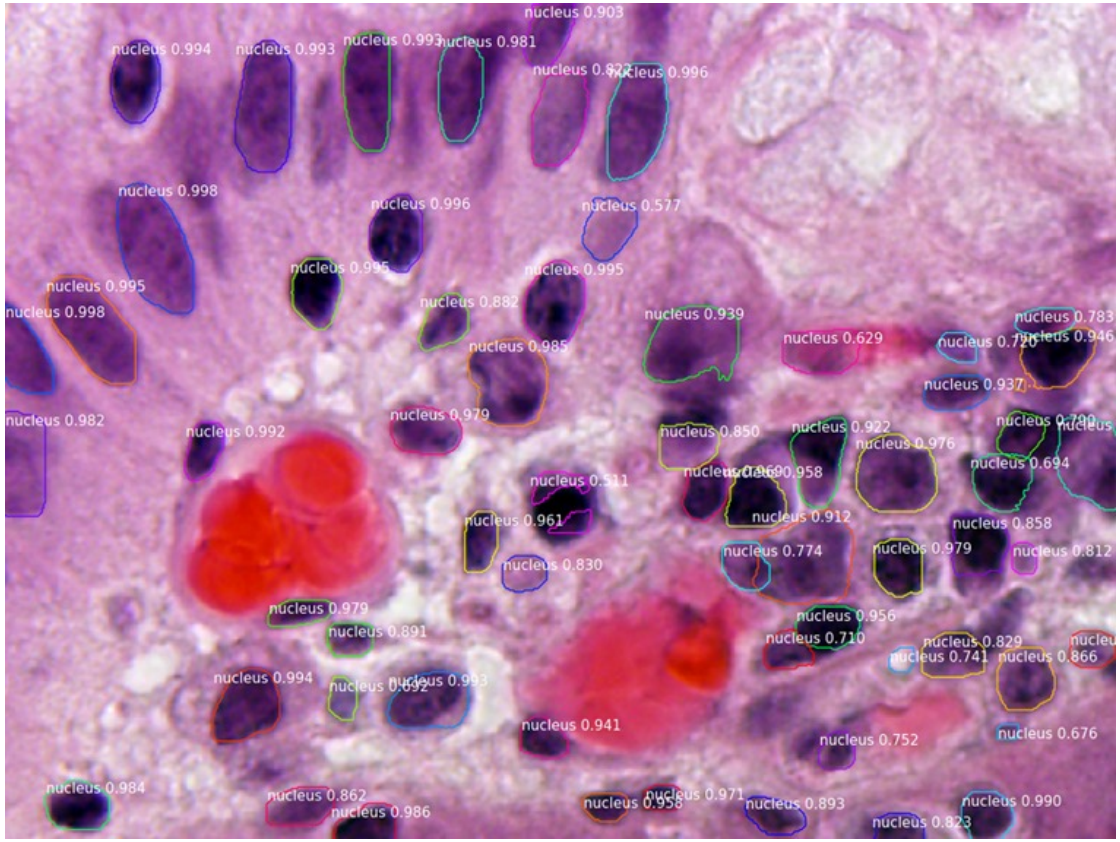
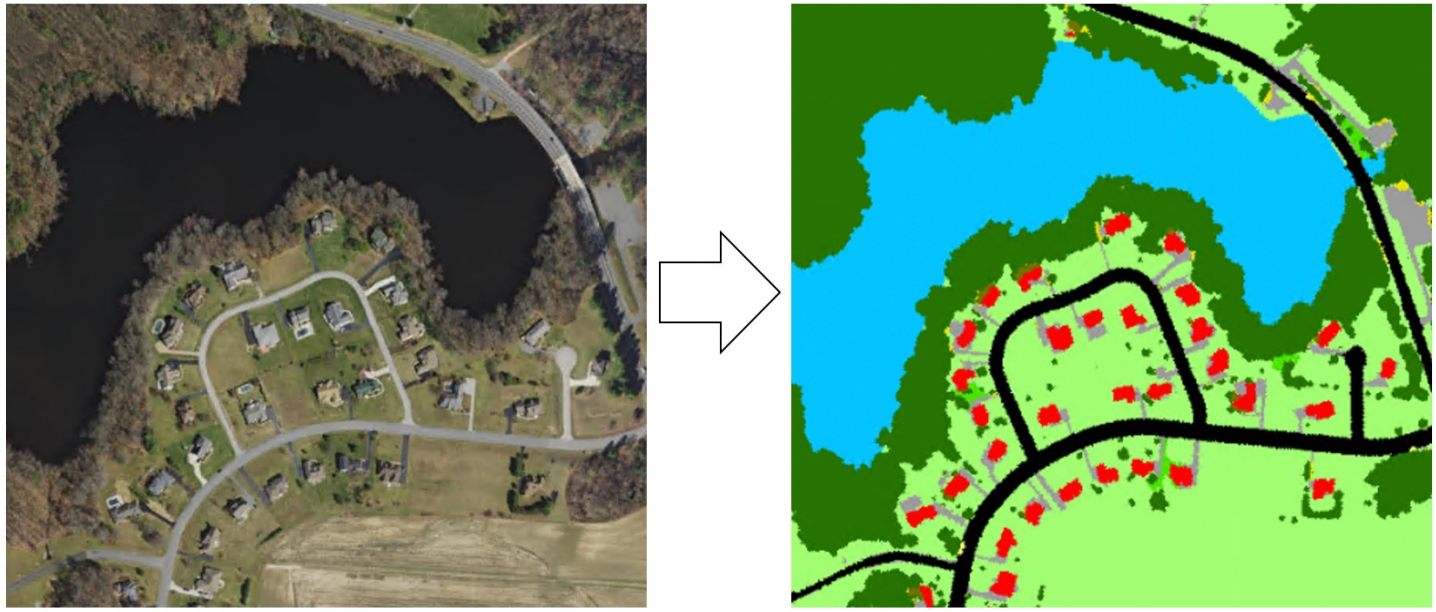


Image Segmentation using Deep Learning



[Segmenting Nuclei in Microscopy Images](). **Built for the** [2018 Data Science Bowl]()

Image Segmentation using Deep Learning

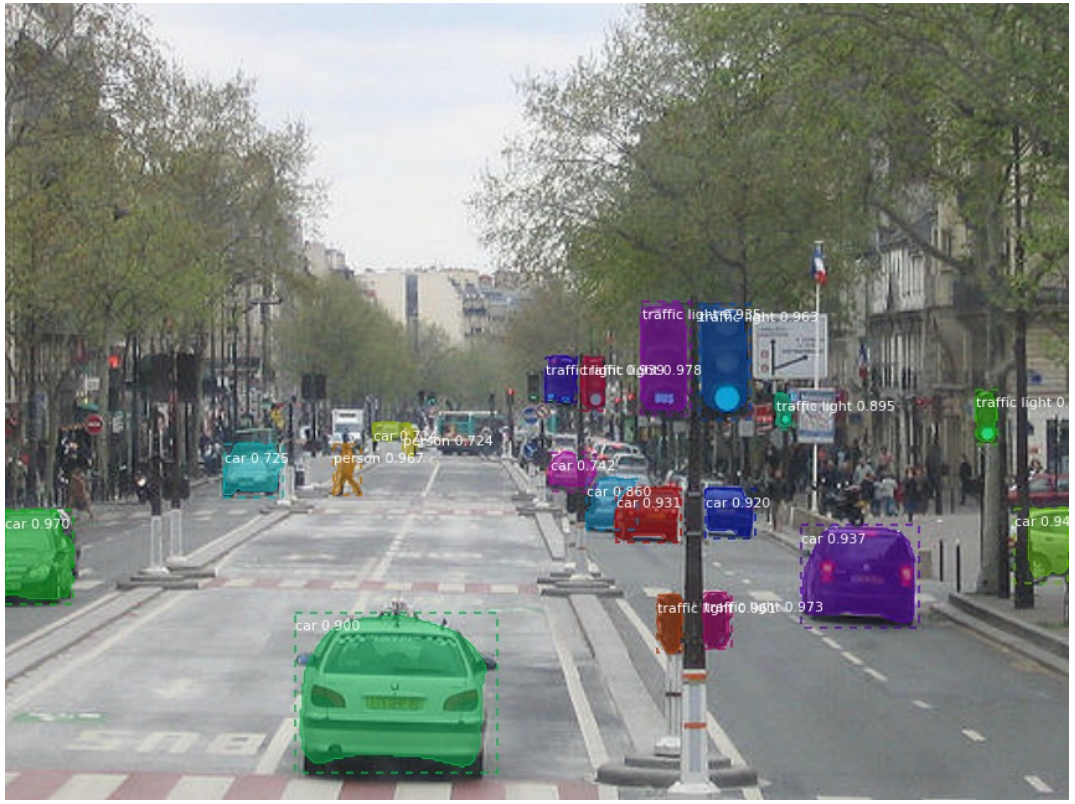


https://developers.arcgis.com/python/latest/guide/how-unet-works/

Image Segmentation using Deep Learning

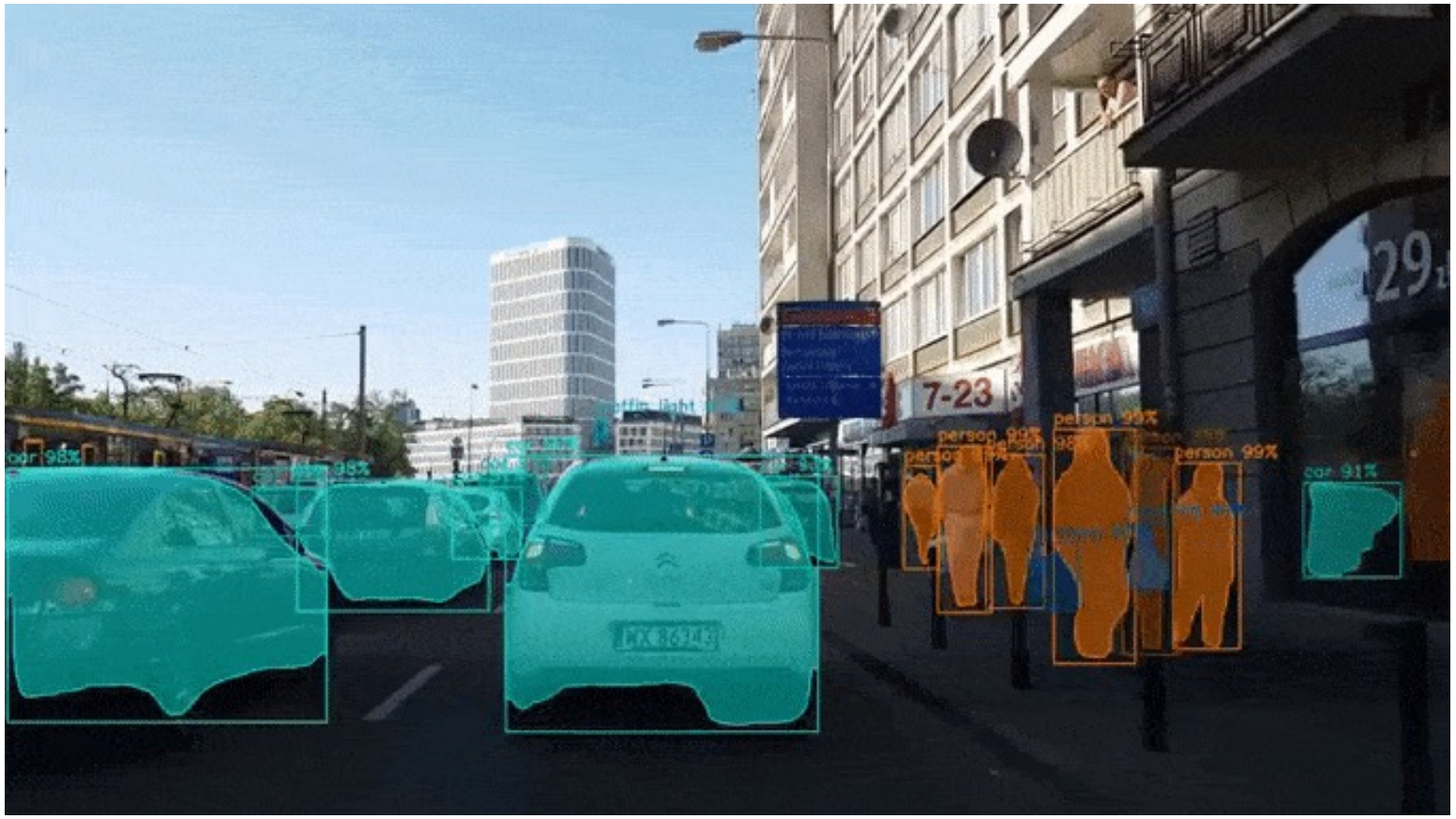
Images to OSM: Improve OpenStreetMap by adding baseball, soccer, tennis, football, and basketball fields

Image Segmentation using Deep Learning



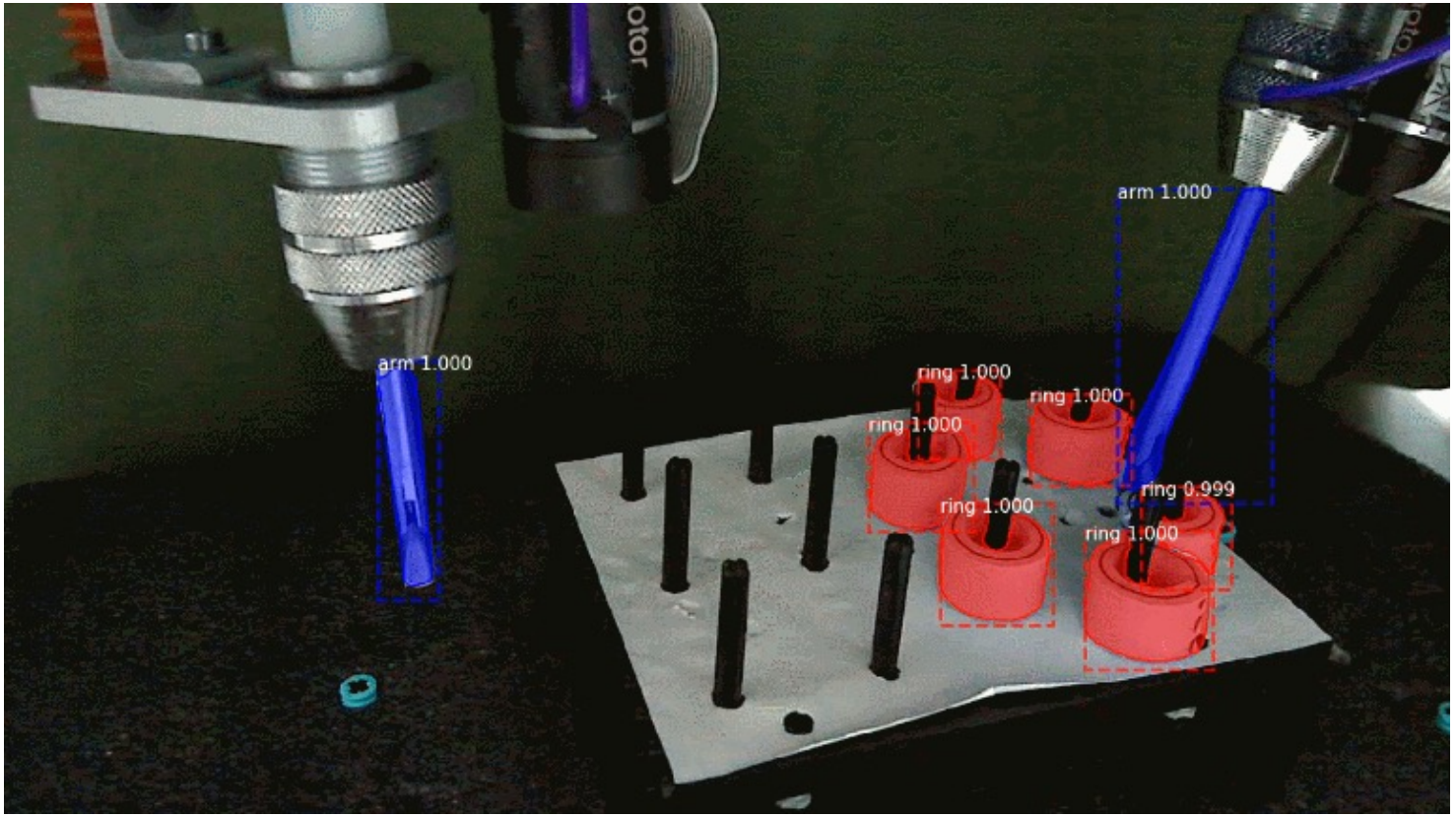
https://github.com/matterport/Mask_RCNN

Image Segmentation using Deep Learning



https://github.com/matterport/Mask_RCNN

Image Segmentation using Deep Learning



Detection and Segmentation for Surgery Robots by the NUS Control & Mechatronics Lab


Thank You