Computer Vision

Introduction

Computer Vision in general

- Image processing - input is image and output is also an image
- Computer vision - Image/video as inputs and output is interpretation
  - Analysis and understanding of single or multiple images
  - Use single or multiple cameras and apply pre-processing
    then apply pattern recognition or AI algorithms for decision making
  - Automatically recognize images and describe them accurately and efficiently
  - Objective is to see objects like humans and possibly even better

Ex: Computer Vision and Human Vision

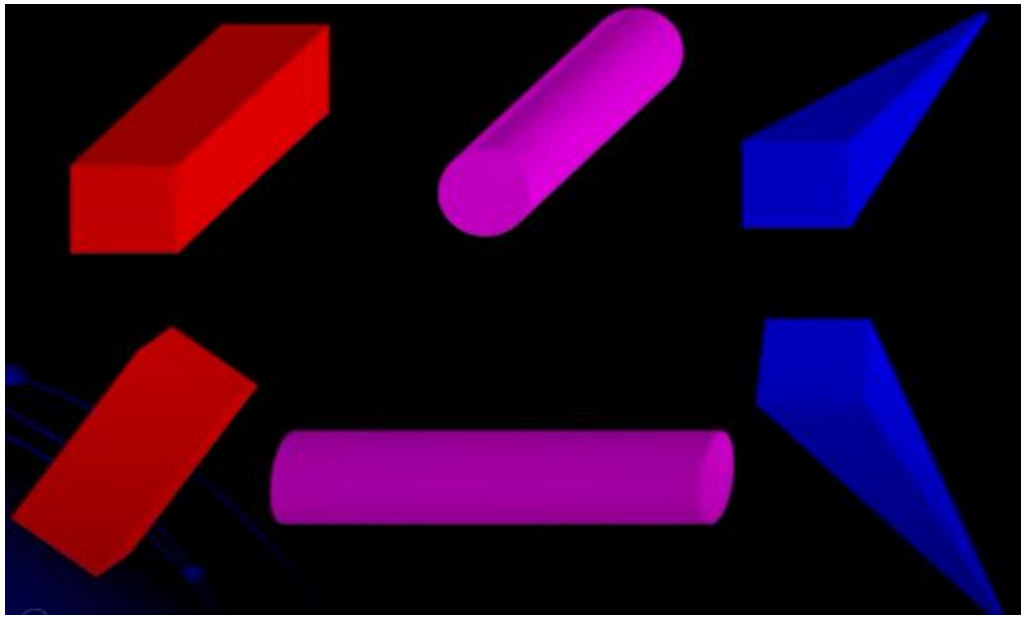
- Human can
  - observe in a few seconds
  - process and take intelligent decisions
  - perform tasks effortlessly and effectively
- For computer
  - not fast and not easy

Computer vision enable computers to see the world in the same way humans do

Features of Human Vision (stereo vision, shape)

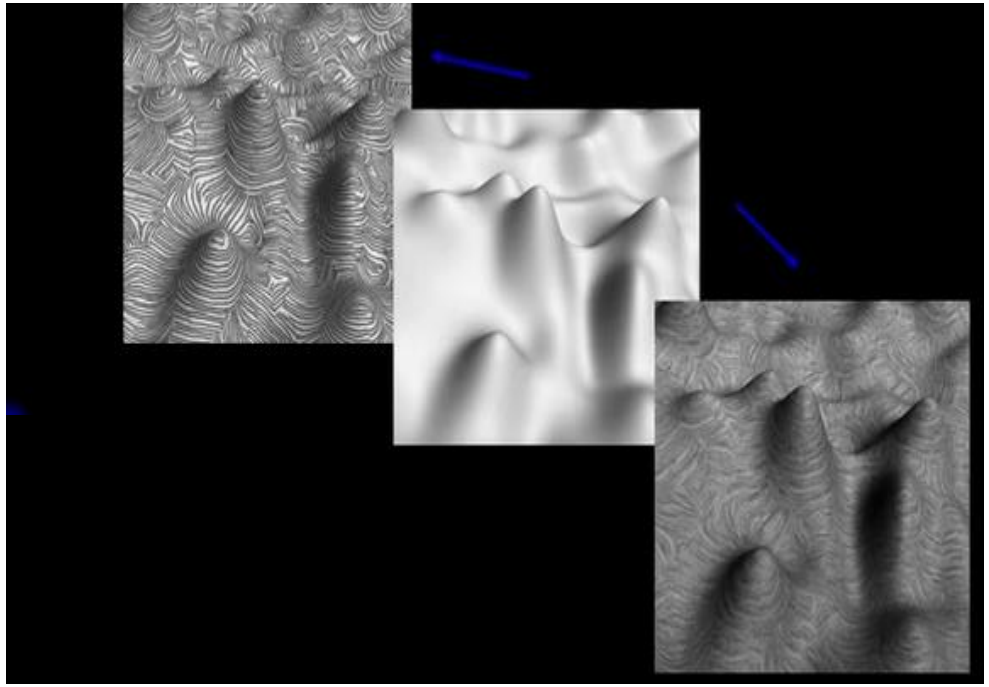
- Two eyes measure different distances from object
- Using distance, eyes can perceive depth
- Stereo Vision using two cameras is used for depth calculation

Shape similarity with object at different depths



Features of Human Vision (texture, color)

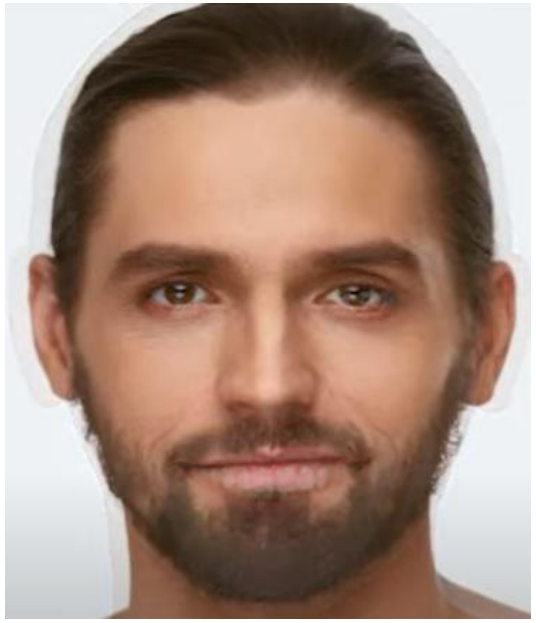
Different texture patterns can have different shape visualization



Identify objects Based on color

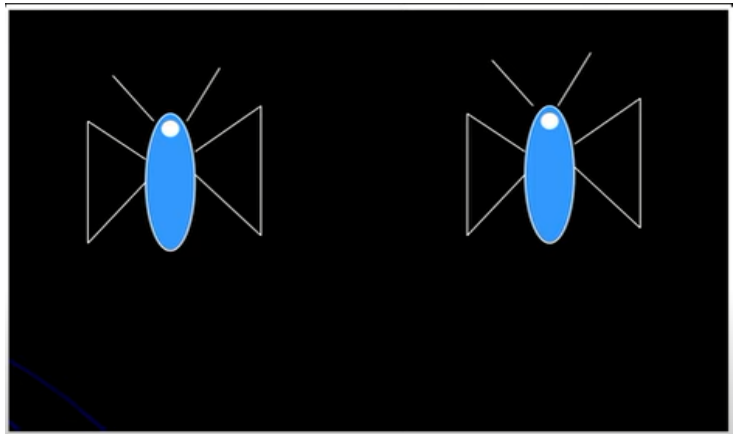
Features of Human Vision (object recognition)

Recognize a person in a photograph clicked many years ago

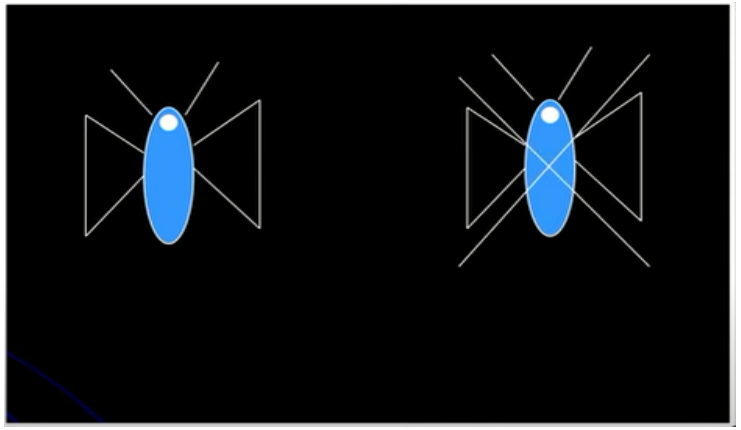


- Human can recognize image with different illumination, view point, expressions etc
- Human can store several images in our brain for future recognition

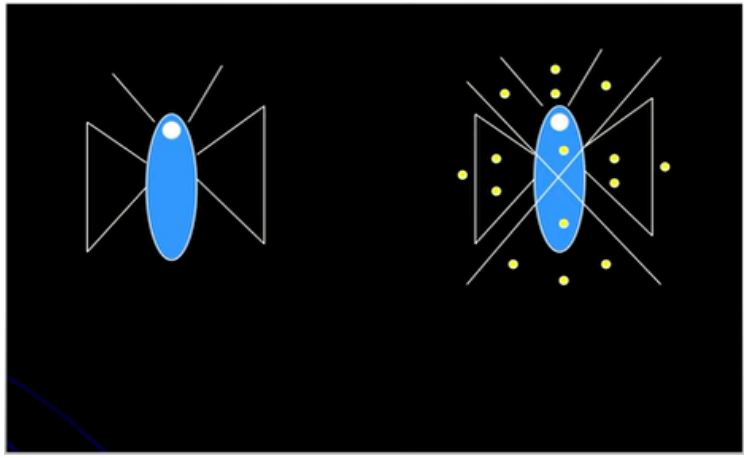
Features of Human Vision (object identification)



Human and computer vision identify



- Human can identify
- Challenge for computer vision



- Human can identify
- Challenge for computer vision

Features of Human Vision (object identification)

- Human vision can infer the context and key information

- Computer vision is more difficult task then human vision

Human Vision

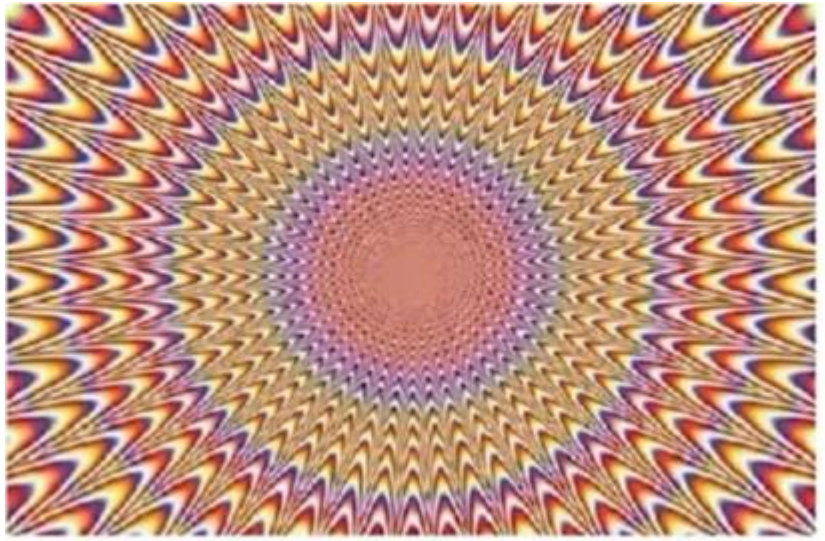
- Human vision can provide
  - Depth perception
  - Relative position/ occlusion
  - Shading of objects
  - Sharpness of edges of objects
  - Size and shape of objects
  - Structure of object
- Limitations of Human Vision
  - limited memory
  - limited to visible spectrum
  - illusion
  - …

Limitations of Human Vision

- Human vision has limited memory
- Human vision is able to establish the context
- Computer vision have complete observations within a short time and can interpret the scene

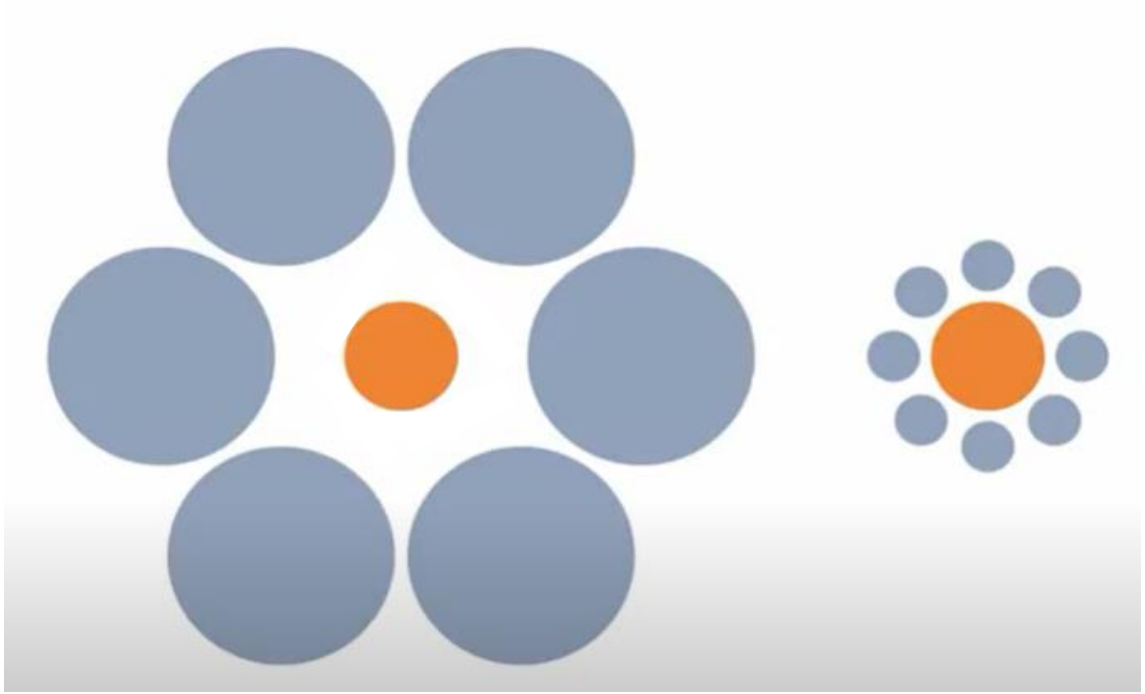
Limitations of Human Vision

Difference in distance



Difference in perception

Limitations of Human Vision



- Sizes of orange circle appear to be different
- Human does interpolation of objects

Circles have the same size

Computer Vision vs Image Processing

- Image processing
  - Image-to-image transformation
  - Uses low level processing
  - Typical image processing operations include
    - image compression
    - image restoration
    - image enhancement
- Computer vision
    - Works on images which already pre-processed to improve image quality
    - Extracts image descriptors
    - Uses higher level techniques (object recognition)
    - Provides semantic understanding of image or video (qualitative or quantitative)
- Objective of computer vision is to see objects like humans and possibly even better

Real-world computer vision applications

- Self-driving cars (allows self-driving cars to safely steer through streets and highways)
- Facial recognition (match images of people's faces to their identities)
- Augmented reality (mix virtual objects with real-world images)
- Medical imaging (scan X-rays, MRIs, and ultrasounds to detect health problems)
- to spot defective products on the assembly line and prevent them from shipping to customers
- Intelligent Video Analytics
- Manufacturing and Construction
- OCR
- Retail
- Banks use it to verify customers' identities before conducting large transactions

Levels of processing for computer Vision

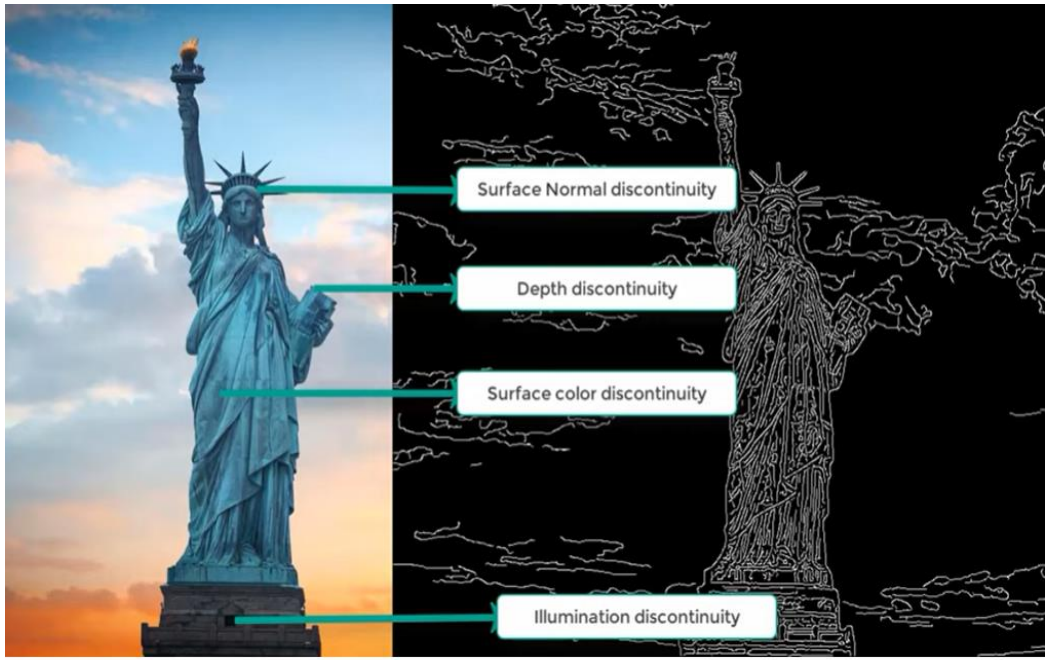
- Low Level Processing
  - Enhance image quality, extract useful information
  - Processed manually or algorithms for specific images
- Mid Level Processing
  - Use low level features to generate symbolic information
- High Level Processing
  - Utilize high-level features like objects and scenes to classify and identify the content or scenario of the image
  - Generate image comprehension

Low Level Processing

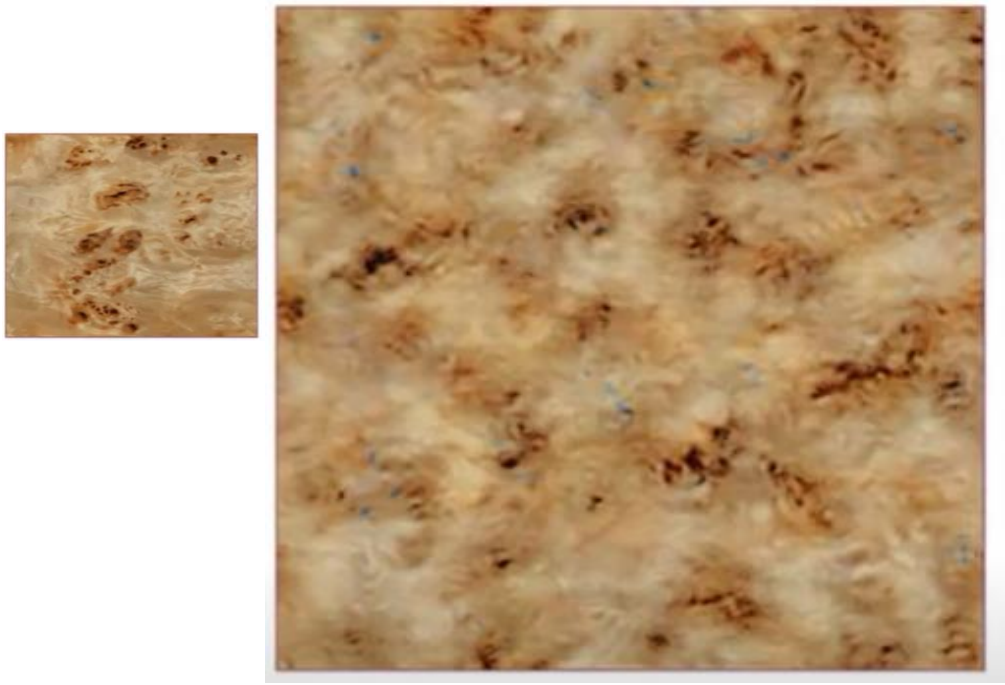
- Image is processed. Both input and output are images
- Analysis usually is not dependent on the type of objects
- Concerned with extracting descriptions/ basic features from images such as colors, textures, and edges
- Steps for low level processing
  - Image preprocessing
    - Use contrast enhancement and noise reduction to improve image quality
  - Feature extraction
    - Use algorithms like Canny edge detection and Sobel filters to identify key features like corners and edges, angles and colors
  - Segmentation
    - Divide image into meaningful segments to analyze individual components

Low Level Processing

- Image enhancement
- Apply edge detection, corner detection, filtering, morphology



Edge Detection



Texture to determine repetitive pattern

Low Level Processing

Extract Low Level Features

Lines

Corners

Salient points

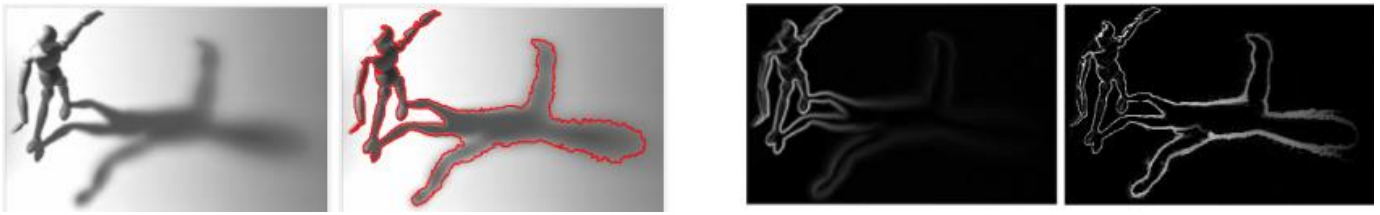
Low Level Processing

- Ex: One is looking at an image of a coffee mug on a desk
- Low level features shows place of the mug edges, specular highlights on the mug surface, identify the colors on the mug
- Features facilitate the representation and analysis of visual input
- This technique applies to any image, not just to the mug
- Unique to a single image or video
- Apply higher level processing to find patterns and correlations for other types of images
- Make predictions or conclusions based on this knowledge

Example1 of Low Level Processing

- Segmentation algorithm to detect low-level structure
- Partition a given image into regions, corresponding to image structures

  regardless of their shapes and sizes
- Draw contours by connecting set of pixels that is surrounded by edges



Example2 of Low Level Processing

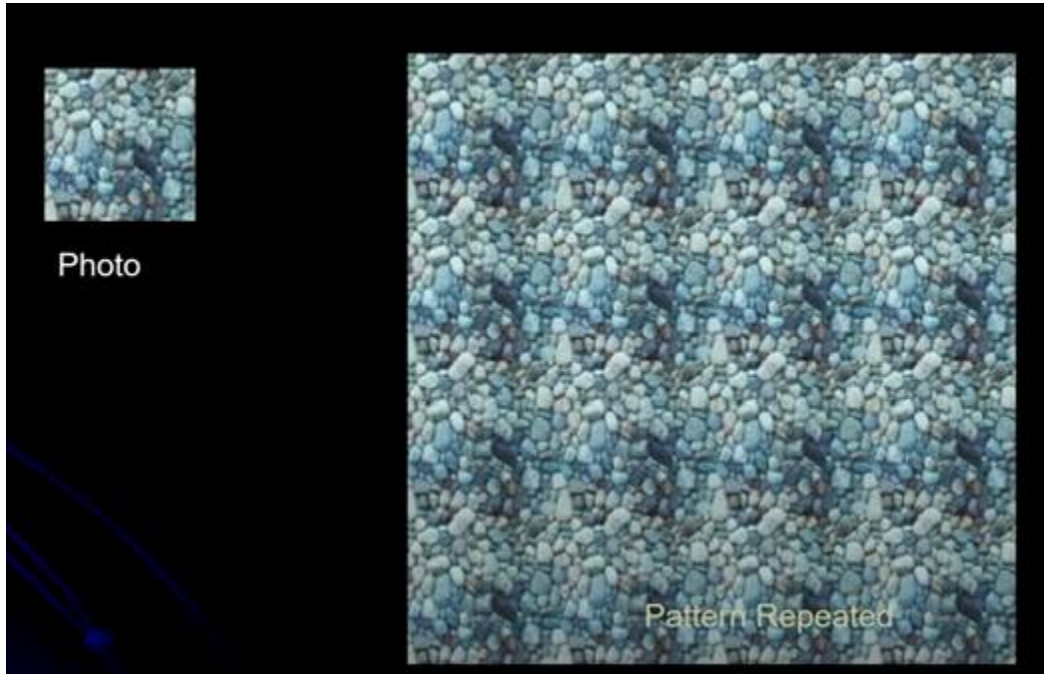
- Generate super-resolution of a single image
- Estimate the number of high resolution pixels from a few number of low resolution pixels available

Example3 of Low Level Processing

- Obtain a noise-free, high resolution image, from a noisy, low resolution image

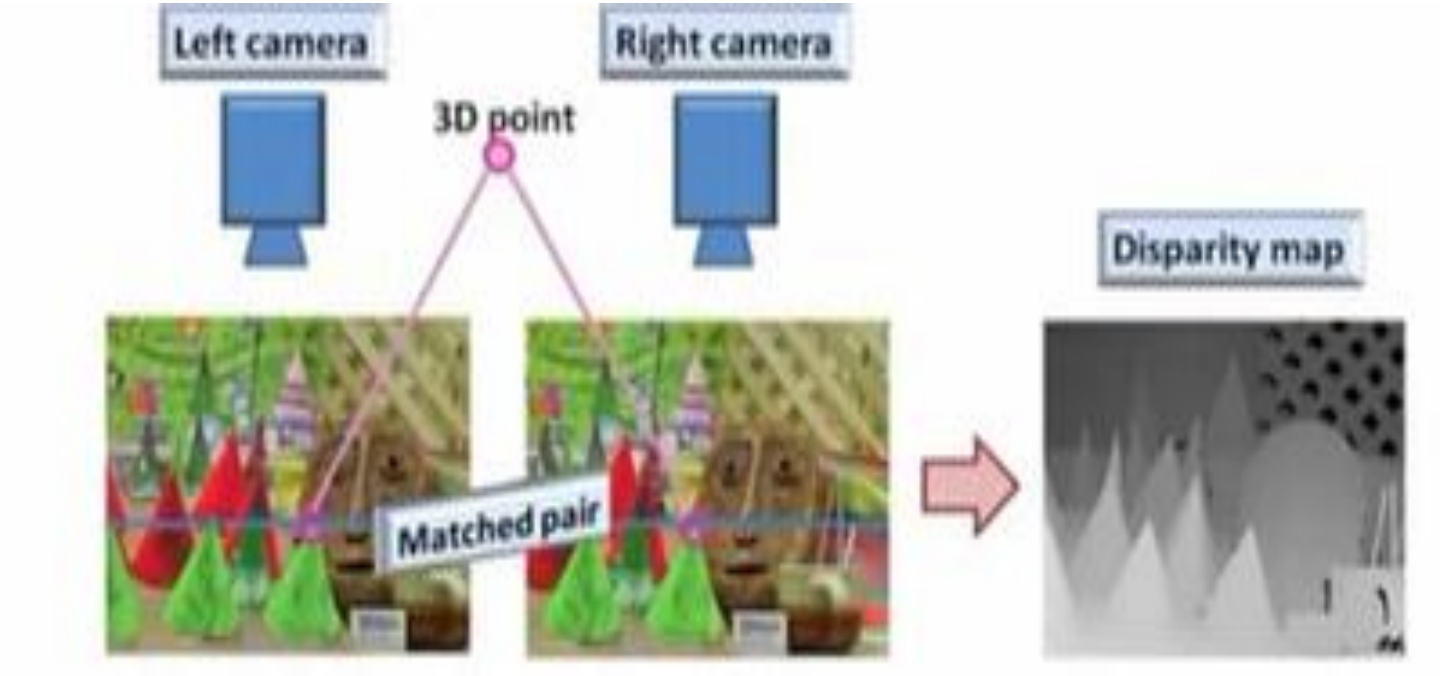
Low Level Processing

Boundary detection



Variation in texture information determines shape of objects

Low Level Processing



Images of left and right cameras are used to determine disparity map which gives depth information

Mid Level Processing

- Builds upon the low-level features to recognize patterns and shapes
- Extracting feature vectors of segmented objects
- Based on feature vectors, classify objects and label them
  - Object Detection
  - Image Classification
  - Assign labels to images based on their content

Mid Level Processing

Uses low level feature for image matching

Image Matching

Mid Level Processing

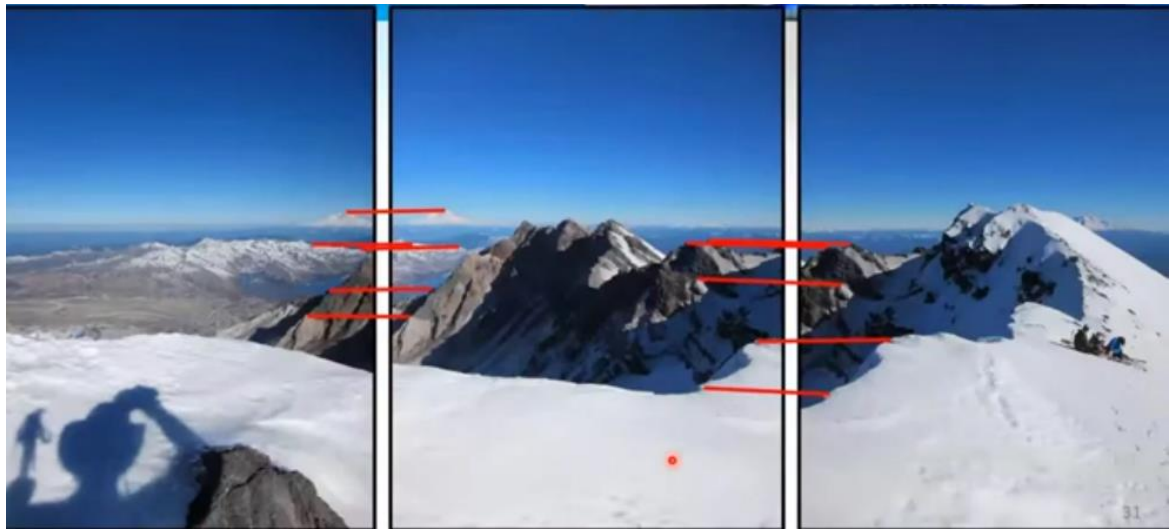


Image Matching

Image Stitching

Panorama Stitching

Mid Level Processing

- Find video sequences that correspond to one scenario
- Keep track of moving object in video

Find correspondence between frames through a sequence of video frames

Track object using background subtraction

Mid Level Processing

Determine optical flow to predict speed and direction of movement of object

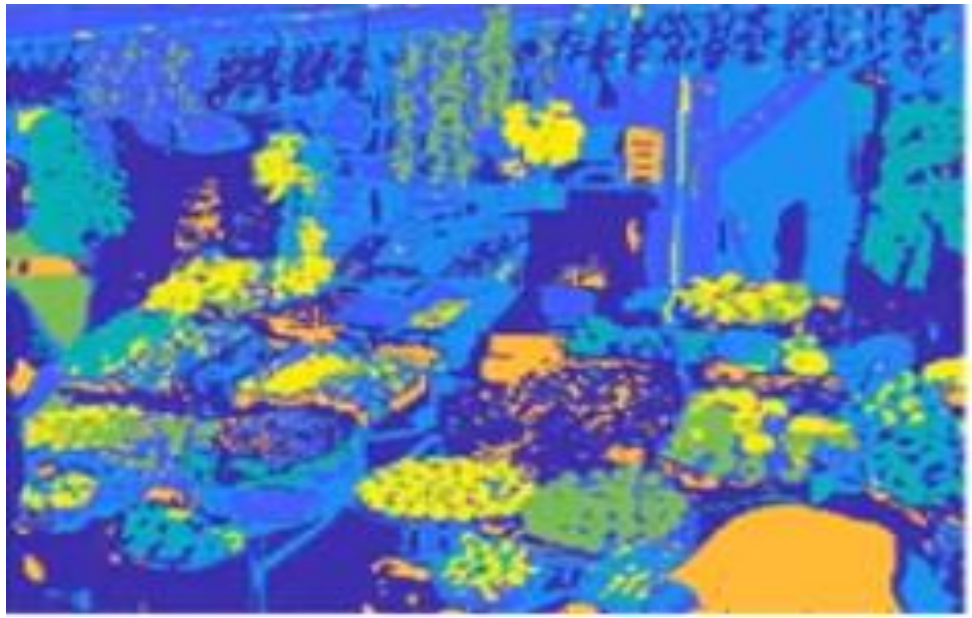
Mid Level Processing

Object tracking - Grouping objects which have similar optical flow

- Detect area of interest and predict where the object will be in next frame

Mid Level Processing

- K-means clustering (k=7) to segment objects

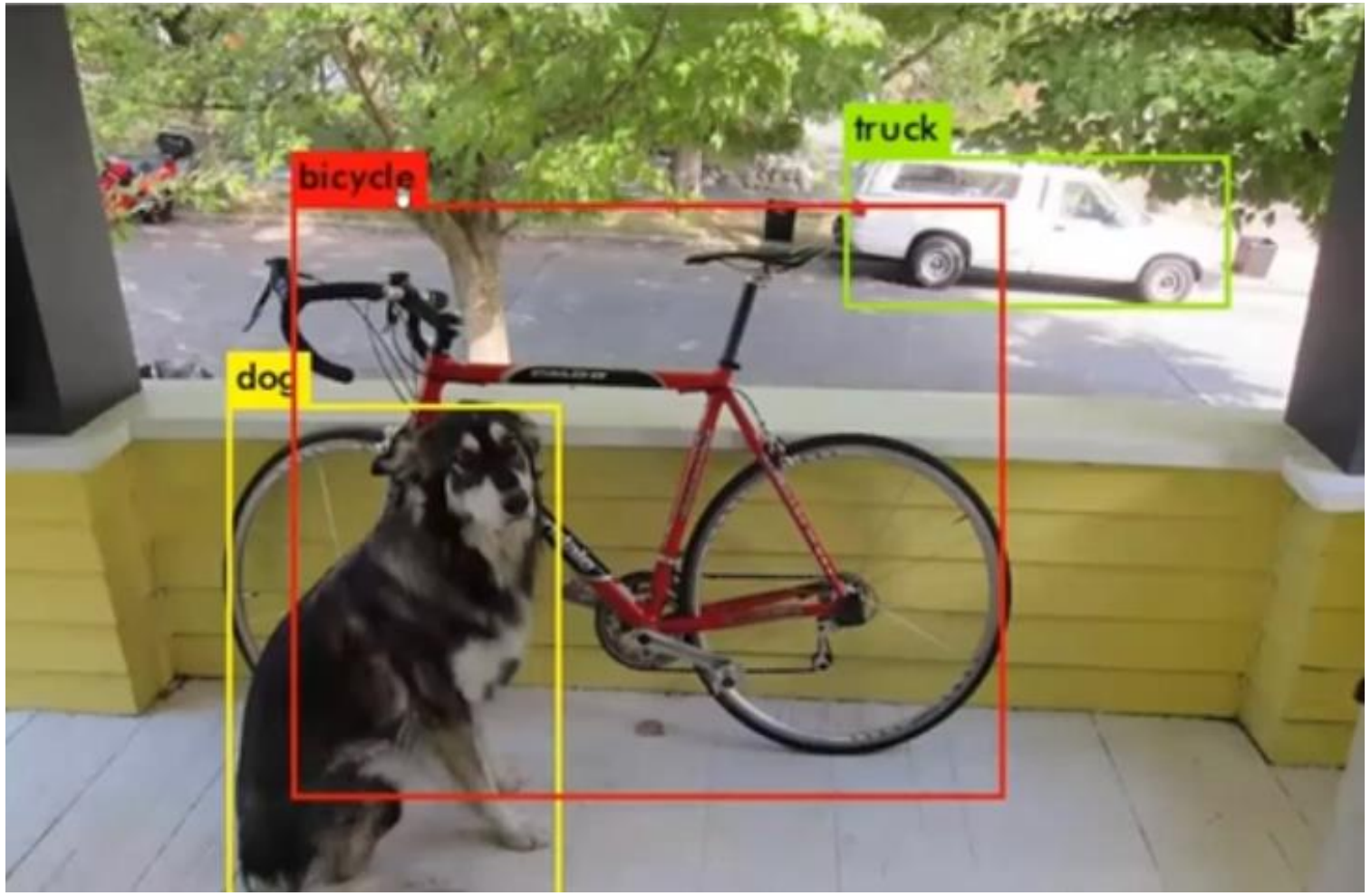


High Level Processing

- Contain more complicated details about image/video
- Built upon low-level feature
- Achieve high levels of accuracy and understanding of visual input by integrating low and high-level information
- Involves more complex tasks such as scene understanding and image captioning where models interpret the context and relationships between objects
- Generally based on Image Understanding
- Ex: Convolutional neural networks (CNNs), Recurrent Neural Networks (RNNs) for learning high-level features
- Ex: Reconstruct, interpret and understand a 3D scene from its 2D images in terms of the properties of the structures present in the scene

High Level Processing

- Identify Objects

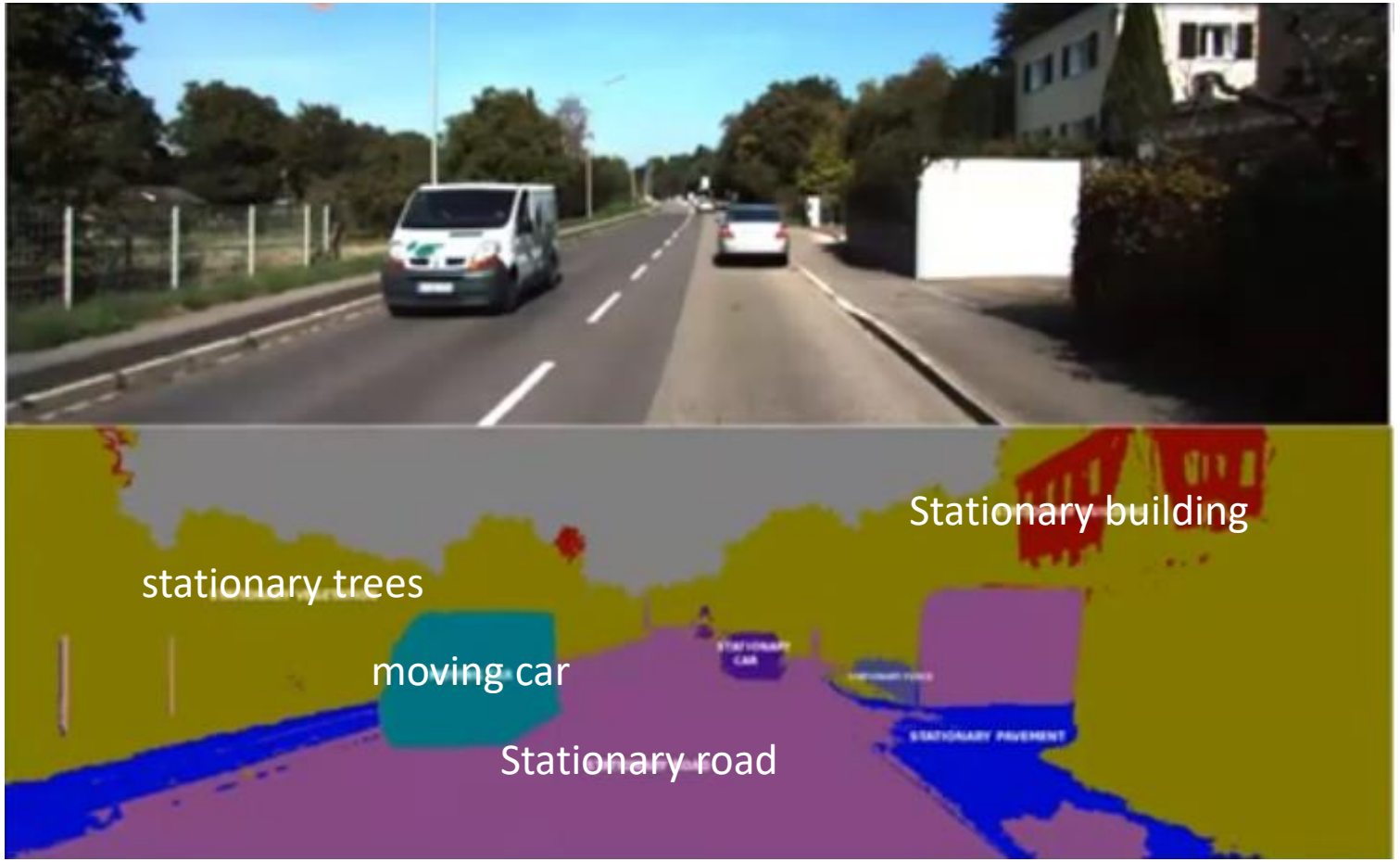


High Level Processing

- Semantic segmentation
  - Classifies pixels based on their semantic meaning

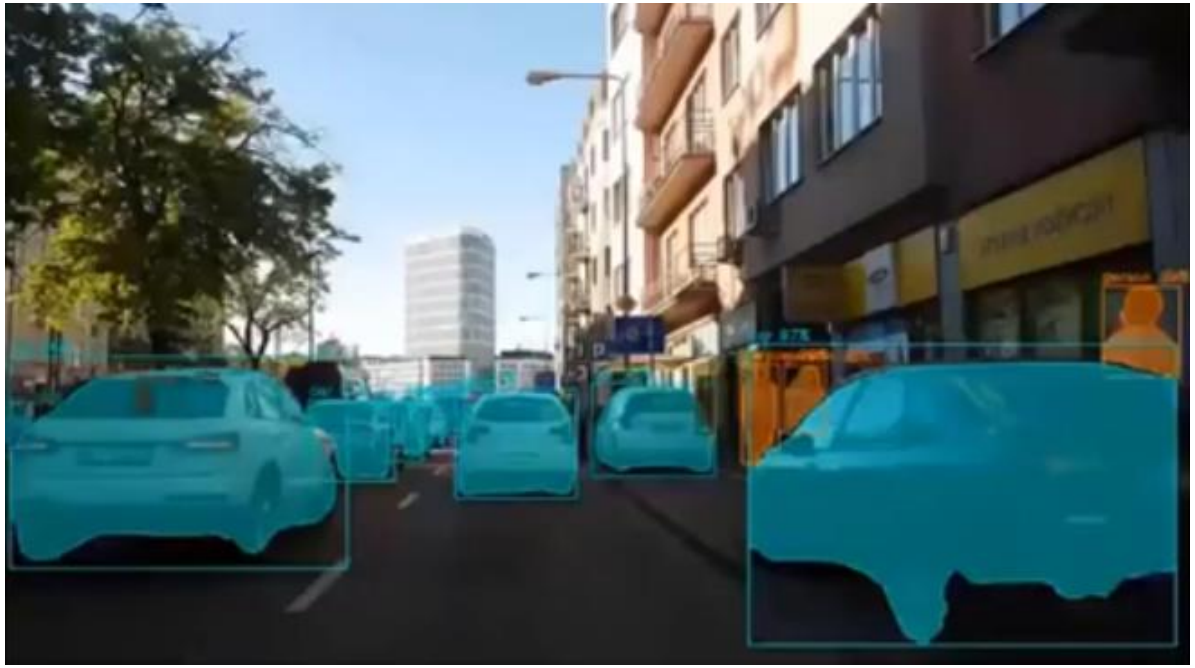
High Level Processing

- Semantic segmentation
  - Classifies pixels based on their semantic meaning
  - Treats all objects within a category as one entity



High Level Processing

- Instance segmentation
  - Distinguishes between different instances of the same class
  - Allows for more precise object identification and differentiation



- Panoptic Segmentation
  - Combines semantic and instance segmentation
  - Labeling each pixel with a class label
  - Identify individual object instances in an image

High Level Processing

Examples:

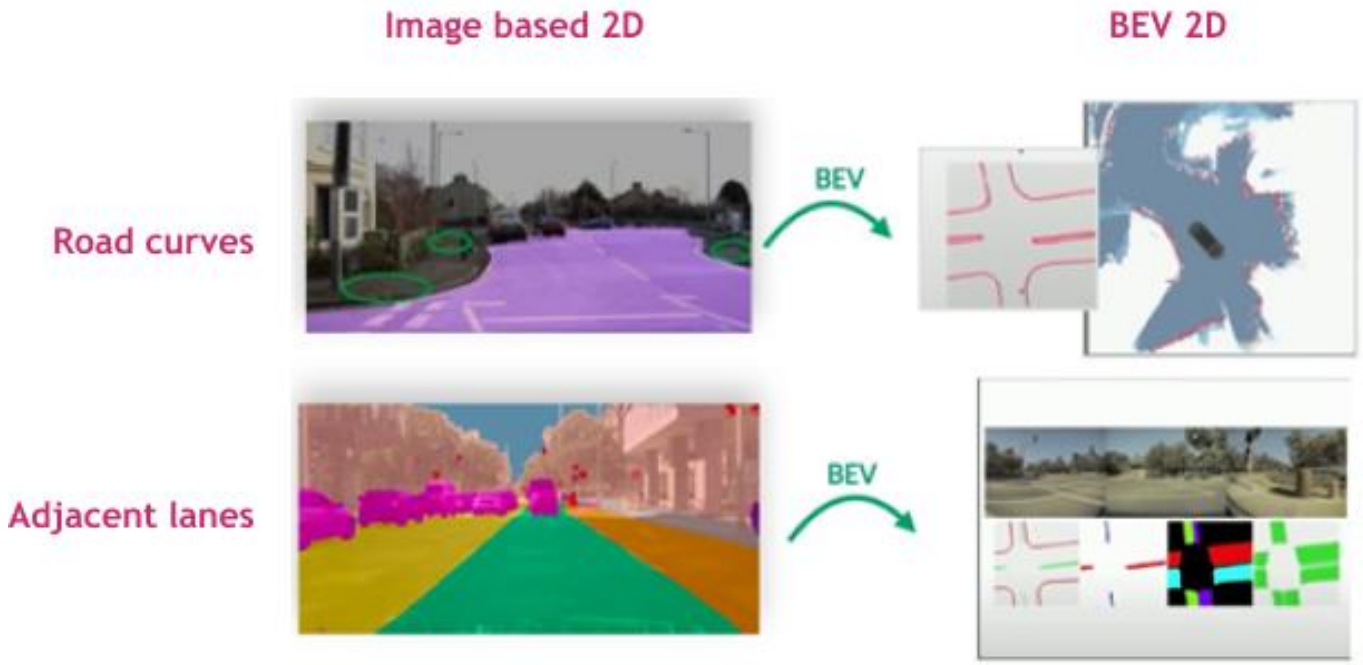
- Semantic Segmentation: Assigning a class label to each pixel in the image, enabling a detailed understanding of the scene
- Instance Segmentation: Differentiating between individual objects of the same class, providing a more granular analysis
- Visual Question Answering: Combining image understanding with natural language processing to answer questions about the visual content

Deep Learning for Image Segmentation

1. CNNs are important deep learning models that helps in image segmentation
2. Object detection algorithms first identify object locations using a region proposal network (RPN), generating candidate bounding boxes
3. CNNs extracts features from the region of interest (ROI) defined by the bounding box
4. feed it into a fully connected convolutional network (FCN) for instance segmentation
5. The FCN outputs a binary mask identifying pixels belonging to the object of interest

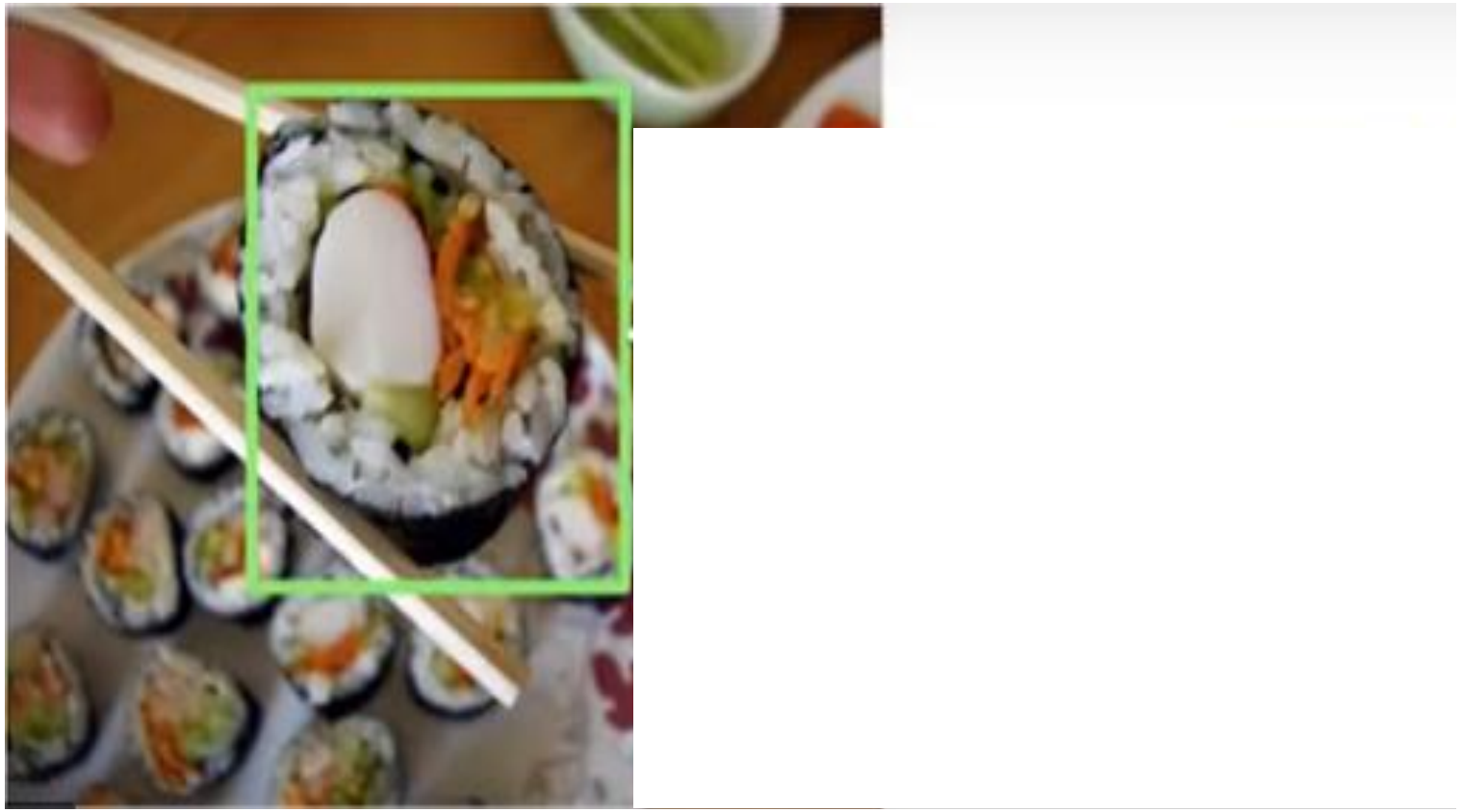
5. Image Segmentation

- Example: Study roads
  - Identify drivable areas, shows where there is free space
  - Point out road curves, giving a closer look at the road environment
  - Segmentation mask is combined with Bird Eye View (BEV) conversion
  - Integration of Panoptic Segmentation with Bird-Eye-View Networks is used for identifying free space and road curves



High Level Processing

Image Understanding



High Level Processing

Image Understanding

High Level Processing

Theme understanding

- Does it have people?
- Is it a market place or football ground or garden?
- Identify location of bicycle

Theme understanding

1. What objects are present in the image?
2. Draw bounding box around each object
3. Classify object as building
4. Is this buiding a house or shop?

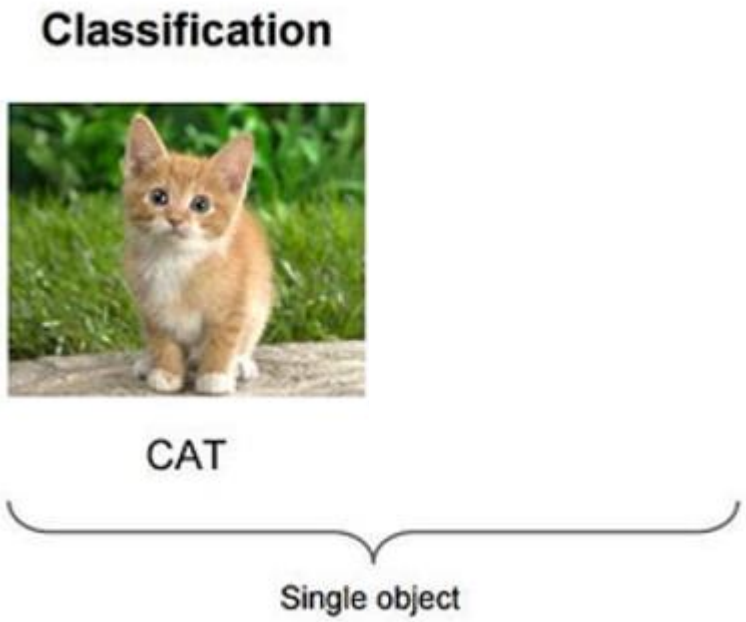
High Level Processing

Visual recognition:

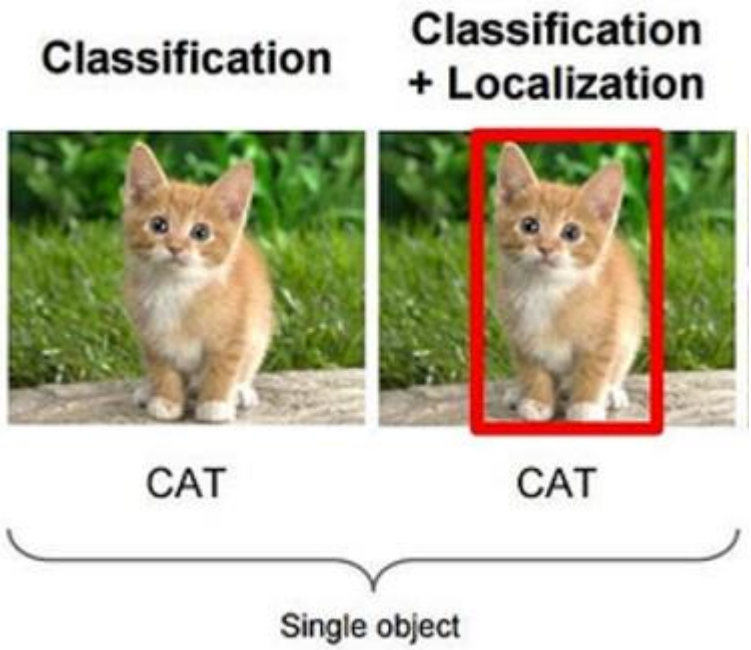
- Classifying images/ videos, localize objects
- Classify human activities

- Identify action
- Predict next action

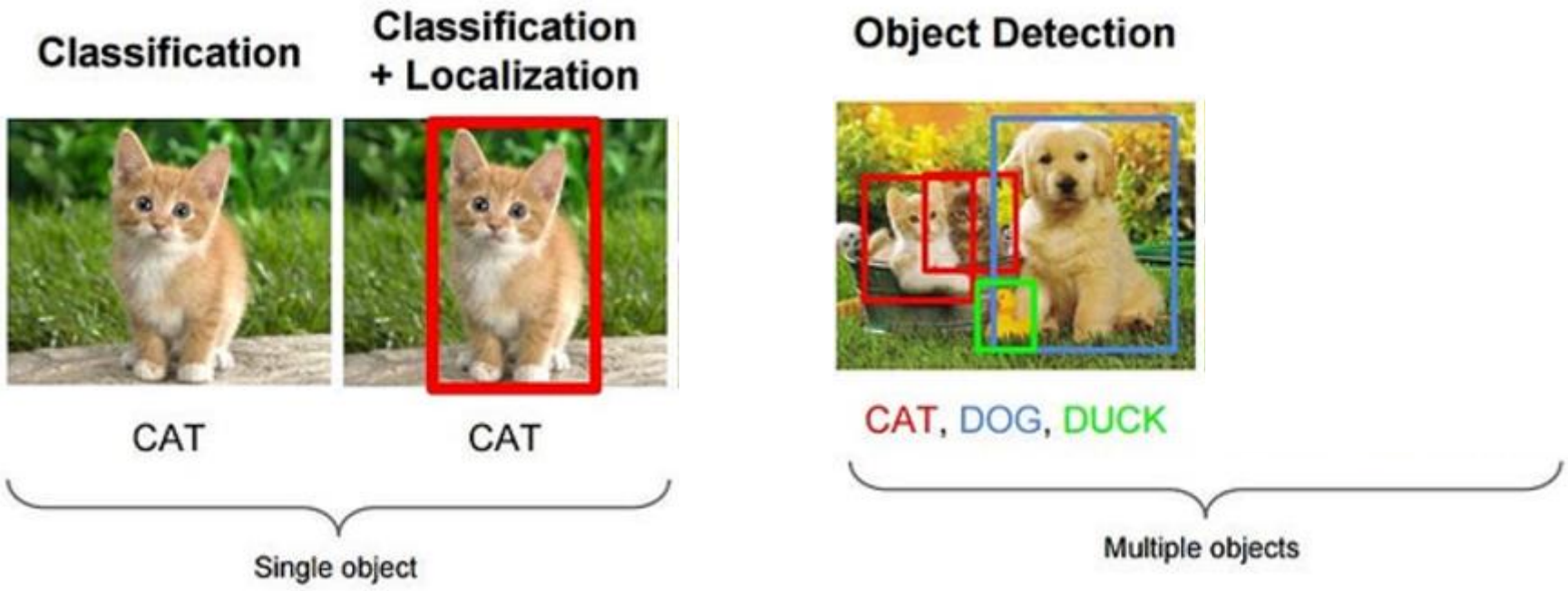
Specific Applications



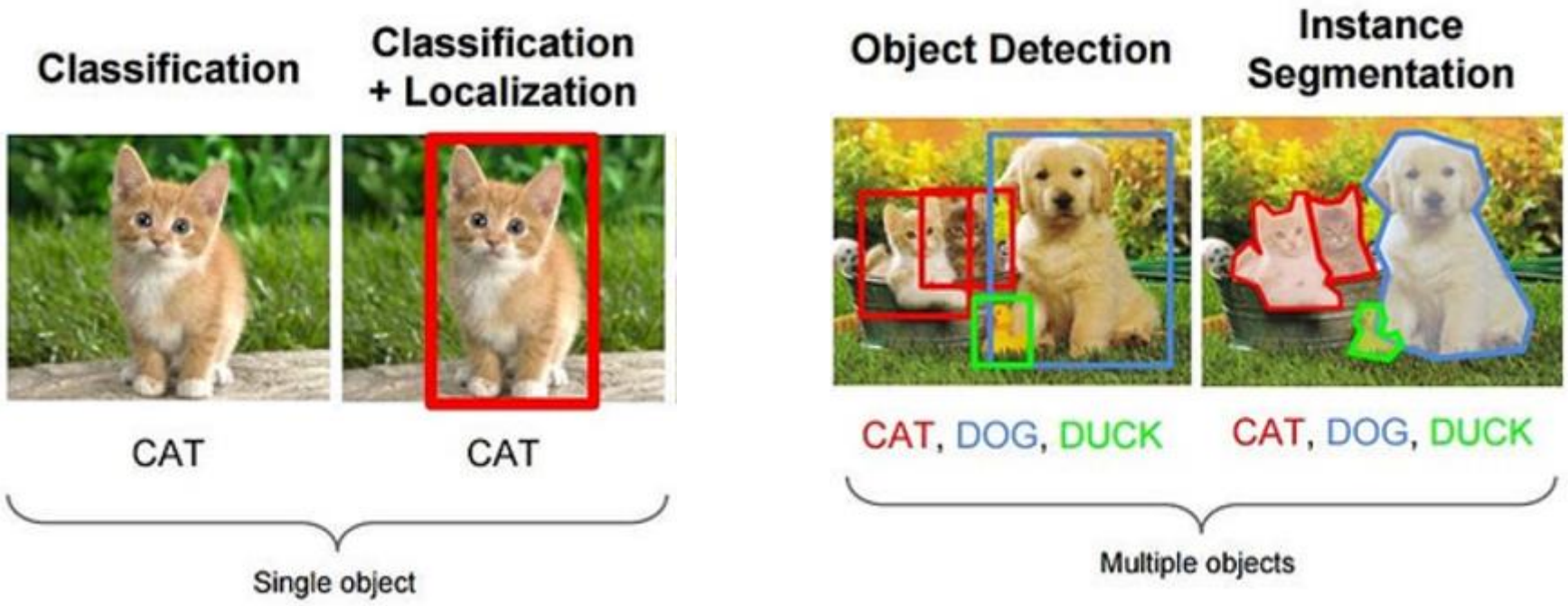
Specific Applications



Specific Applications



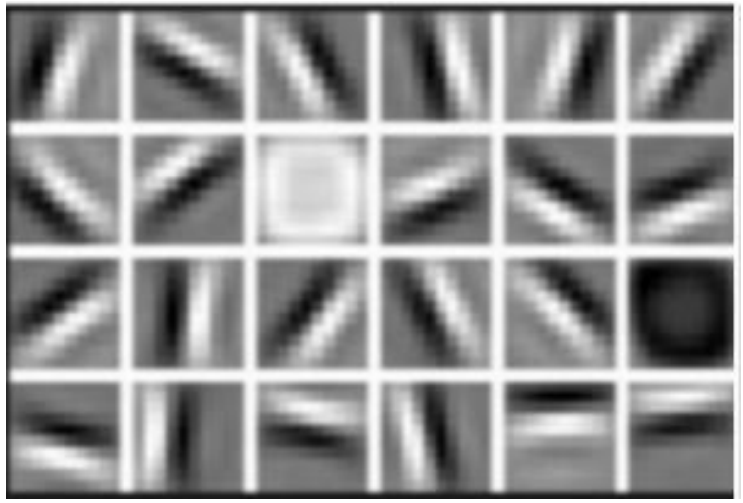
Specific Applications



Most of the applications are based on features of image

Difference between Low and High Level Processing

- Low level processing extracts features such as colors, edges, and textures from an image
- High level processing extracts from low-level features and use it to denote more meaningful concepts



Lines, Corners, Edges

Faces with expressions

Difference between Low and High Level Processing

	Low Level	**High Level**
Content	• Related to the raw pixel data of the image • More sensitive to noise and changes in the image	• More robust • Higher level of understanding of the image content

Difference between Low and High Level Processing

	Low Level	**High Level**
Content	• Related to the raw pixel data of the image • More sensitive to noise and changes in the image	• More robust • Higher level of understanding of the image content
Scale	• Typically retrieved for local region • Vulnerable to little modifications of the picture, like lighting or orientation	• Frequently retrieved at a global scale • Takes into account the whole image/video • More robust

Difference between Low and High Level Processing

	Low Level	**High Level**
Content	• Related to the raw pixel data of the image • More sensitive to noise and changes in the image	• More robust • Higher level of understanding of the image content
Scale	• Typically retrieved for local region • Vulnerable to little modifications of the picture, like lighting or orientation	• Frequently retrieved at a global scale • Takes into account the whole image/video • More robust
Resources	• Feature extraction usually takes fewer system resources	• Requires more advanced techniques

Difference between Low and High Level Features

	Low Level	**High Level**
Content	• Related to the raw pixel data of the image • More sensitive to noise and changes in the image	• More robust • Higher level of understanding of the image content
Scale	• Typically retrieved for local region • Vulnerable to little modifications of the picture, like lighting or orientation	• Frequently retrieved at a global scale • Takes into account the whole image/video • More robust
Resources	• Feature extraction usually takes fewer system resources	• Requires more advanced techniques
Useful	• Image segmentation, object detection, and feature matching	• Image classification, object recognition, and scene understanding

Applications

- Object
  - Classification: Broad category of object in image
  - Identification: Type of a given object in image
  - Detection: Check whether object exists in image
  - Landmark Detection: Identify key points of the objects
  - Segmentation: Identify pixels belonging to objects
  - Recognition: Existence and location of objects
- Video motion analysis to estimate the velocity of objects in a video, or the camera
- Scene reconstruction to create a 3D model of a scene captured in the form of images or video

References

- https://builtin.com/artificial-intelligence/image-recognition#:~:text=Using%20image%20recognition%2C%20a%20computer,within%20it%20and%20respond%20accordingly
- https://randomwalk.ai/blog/the-5-fundamental-processes-in-computer-vision/
- https://www.google.com/url?sa=t&source=web&rct=j&url=https://sites.ecse.rpi.edu/~qji/CV/3dvision_intro.pdf&ved=2ahUKEwidwsC95Kv-AhU9jGMGHQa6DZsQFnoECBQQBg&usg=AOvVaw3hM3mSc376_FoSKN-jNp9F
- https://www.v7labs.com/blog/what-is-computer-vision
- https://towardsdatascience.com/everything-you-ever-wanted-to-know-about-computer-vision-heres-a-look-why-it-s-so-awesome-e8a58dfb641e
- https://www.simplilearn.com/computer-vision-article
- https://www.javatpoint.com/computer-vision
- http://www.cs.williams.edu/~andrea/cs108/Lectures/Lect14/Lect14.html
- https://www.baeldung.com/cs/cv-low-vs-high-level-features#:~:text=In%20the%20area%20of%20computer,examples%20of%20high%2Dlevel%20features.
- https://nptel.ac.in/courses/108103174