Computer Vision

Image Segmentation

Course Instructor:
Dr. Suman Kumar Maji

What is image segmentation?

- Image segmentation is the process of dividing an image into multiple meaningful and homogeneous regions or objects based on their inherent characteristics, such as color, texture, shape, or brightness.
- Image segmentation aims to simplify and/or change the representation of an image into something more meaningful and easier to analyze.
- Here, each pixel is labeled. All the pixels belonging to the same category have a common label assigned to them.

Image segmentation tasks

- The task of segmentation can further be done in two ways:
  1. **Similarity:** The segments are formed by detecting similarity between image pixels. It is often done by thresholding. Machine learning algorithms (such as clustering) are based on this type of approach for image segmentation.
  2. **Discontinuity:** The segments are formed based on the change of pixel intensity values within the image. This strategy is used by line, point, and edge detection techniques to obtain intermediate segmentation results that may be processed to obtain the final segmented image.

Types of Segmentation

- Image segmentation modes are divided into three categories based on the amount and type of information that should be extracted from the image:
  1. Instance
  2. semantic
  3. panoptic
- To understand the three modes of image segmentation, it would be more convenient to know more about objects and backgrounds.
- Objects are the identifiable entities in an image that can be distinguished from each other by assigning unique IDs, while the background refers to parts of the image that cannot be counted, such as the sky, water bodies, and other similar elements.
- By distinguishing between objects and backgrounds, it becomes easier to understand the different modes of image segmentation and their respective applications.

Types of Image segmentation

Types of Image Segmentation

Type	Definition	Characteristics	Applications
Instance Segmentation	Detects and segments each object as a separate entity with unique boundaries.	- Identifies individual objects (e.g., cars, people). - Separates overlapping objects. - Does not require class information.	Object tracking, AR/VR, advanced object detection.
Semantic Segmentation	Labels each pixel with a class category (e.g., "sky," "road," "car").	- Assigns dense class labels to every pixel. - Does not distinguish between different instances of the same class. - Background and objects grouped by class.	Road analysis, medical imaging, scene understanding.
Panoptic Segmentation	Combines semantic and instance segmentation for detailed labeling.	- Provides pixel-wise labeling by class and instance. - Differentiates objects and their instances. - Offers the most granular, high-quality information.	Autonomous vehicles, robotics, interactive AI systems.

Instance Segmentation

- Instance segmentation is a type of image segmentation that involves detecting and segmenting each object in an image.
- It is similar to object detection but with the added task of segmenting the object's boundaries.
- The algorithm has no idea of the class of the region, but it separates overlapping objects.
- Instance segmentation is useful in applications where individual objects need to be identified and tracked.

cont...

Semantic Segmentation

- Semantic segmentation is a type of image segmentation that involves labeling each pixel in an image with a corresponding class label with no other information or context taken into consideration.
- The goal is to assign a label to every pixel in the image, which provides a dense labeling of the image.
- The algorithm takes an image as input and generates a segmentation map where the pixel value (0,1,...255) of the image is transformed into class labels (0,1,...n).
- It is useful in applications where identifying the different classes of objects on the road is important.

cont...

Semantic segmentation - the human and the dog are classified together as mammals and separated from the rest of the background.

Panoptic Segmentation

- Panoptic segmentation is a combination of semantic and instance segmentation.
- It involves labeling each pixel with a class label and identifying each object instance in the image.
- This mode of image segmentation provides the maximum amount of high-quality granular information from machine learning algorithms.
- It is useful in applications where the computer vision model needs to detect and interact with different objects in its environment, like an autonomous robot.

cont...

Image Segmentation Techniques: Traditional Techniques

- Traditional image segmentation techniques have been used for decades in computer vision to extract meaningful information from images.
- These techniques are based on mathematical models and algorithms that identify regions of an image with common characteristics, such as color, texture, or brightness.
- Traditional image segmentation techniques are usually computationally efficient and relatively simple to implement.
- They are often used for applications that require fast and accurate segmentation of images, such as object detection, tracking, and recognition.

cont...

Traditional Segmentation Techniques

Technique	Definition	Characteristics	Applications
Thresholding A value	Divides image pixels into classes based on intensity relative to a threshold value.	- Simple and efficient. - Binary output (foreground/background). - Effective for high-contrast images. - Global or adaptive.	Document scanning, basic object detection.
Region-based Segmentation Similarity based	Groups pixels into regions based on similarity (e.g., color, texture, intensity).	- Can use "split and merge" or "graph-based" methods. - Requires criteria for similarity. - Handles non-uniform regions.	Medical imaging, scene segmentation, region analysis.
Edge-based Segmentation Intensity change	Detects object boundaries based on intensity changes.	- Focuses on edges. - Common methods: Canny, Sobel, LoG. - May struggle with noisy or smooth regions.	Shape detection, contour extraction, feature recognition.
Clustering Clustering algorithms	Groups pixels with similar features into clusters using algorithms like K-means or mean-shift.	- Handles high-dimensional data. - Adaptable to various similarity metrics. - Limited accuracy in complex scenes.	Color-based segmentation, image simplification, feature extraction.

Thresholding

cont...

- Thresholding is one of the simplest image segmentation methods.
- The pixels are divided into classes based on their histogram intensity which is relative to a fixed value or threshold.
- This method is suitable for segmenting objects where the difference in pixel values between the two target classes is significant.
- In low-noise images, the threshold value can be kept constant, but with images with noise, dynamic thresholding performs better.
- In thresholding-based segmentation, the greyscale image is divided into two segments based on their relationship to the threshold value, producing binary images.
- Algorithms like contour detection and identification work on these binarized images.

Thresholding types: Global thresholding

- The two commonly used thresholding methods are:
  1. Global thresholding
  2. Adaptive thresholding
- **Global thresholding** is a technique used in image segmentation to divide images into foreground and background regions based on pixel intensity values.
- A threshold value is chosen to separate the two regions, and pixels with intensity values above the threshold are assigned to the foreground region and those below the threshold to the background region.
- This method is simple and efficient but may not work well for images with varying illumination or contrast.
- In those cases, adaptive thresholding techniques may be more appropriate.

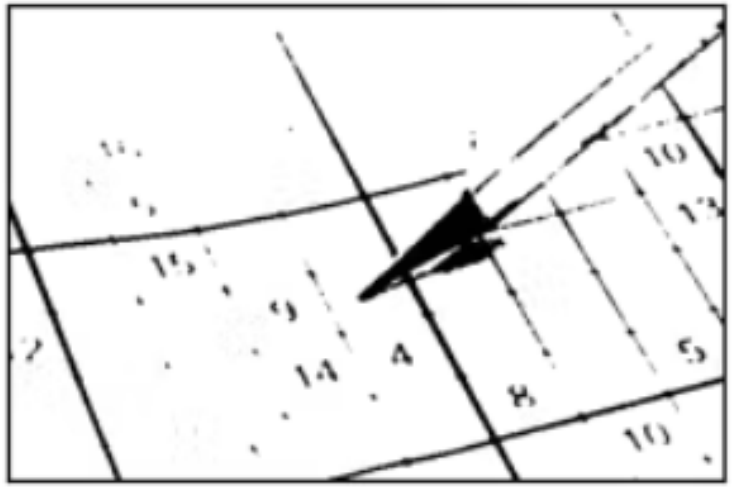
Thresholding types: Adaptive thresholding

- **Adaptive thresholding** is a technique used in image segmentation to divide an image into foreground and background regions by adjusting the threshold value locally based on the image characteristics.
- The method involves selecting a threshold value for each smaller region or block, based on the statistics of the pixel values within that block.
- Adaptive thresholding is useful for images with non-uniform illumination or varying contrast and is commonly used in document scanning, image binarization, and image segmentation.
- The choice of adaptive thresholding technique depends on the specific application requirements and image characteristics.

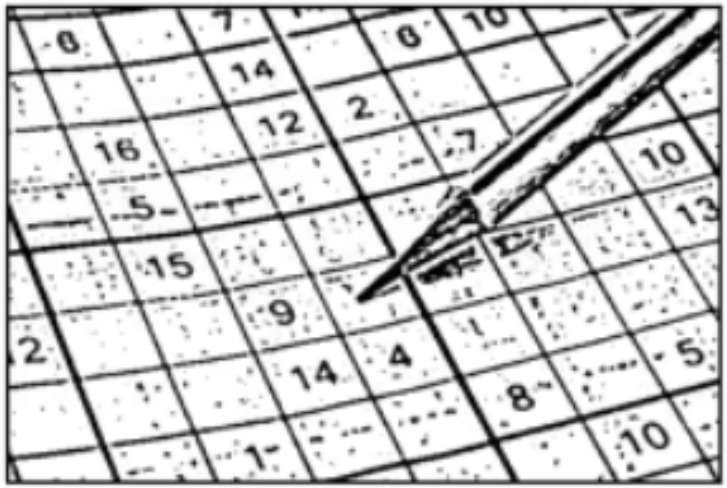
Thresholding types

Original Image

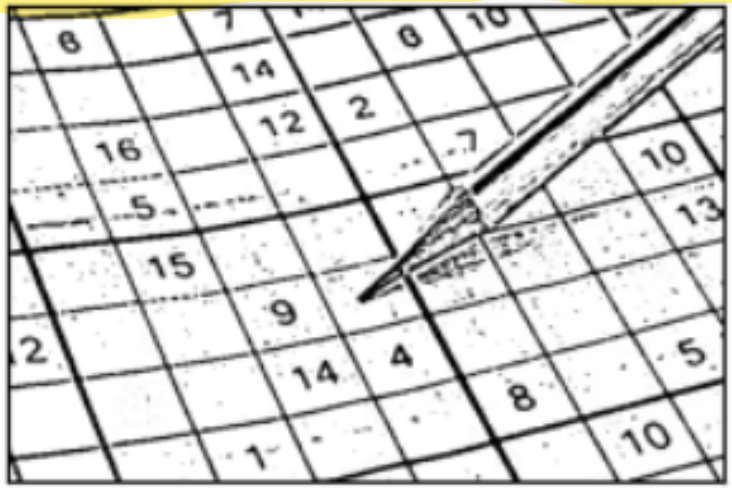
Global Thresholding (v = 127)



Adaptive Mean Thresholding



Adaptive Gaussian Thresholding



Region-based Segmentation

- Region-based segmentation is a technique used in image processing to divide an image into regions based on similarity criteria, such as color, texture, or intensity.
- The method involves grouping pixels into regions or clusters based on their similarity and then merging or splitting regions until the desired level of segmentation is achieved.
- The two commonly used region-based segmentation techniques are:
  1. **Split and merge segmentation**
  2. **Graph-based segmentation**

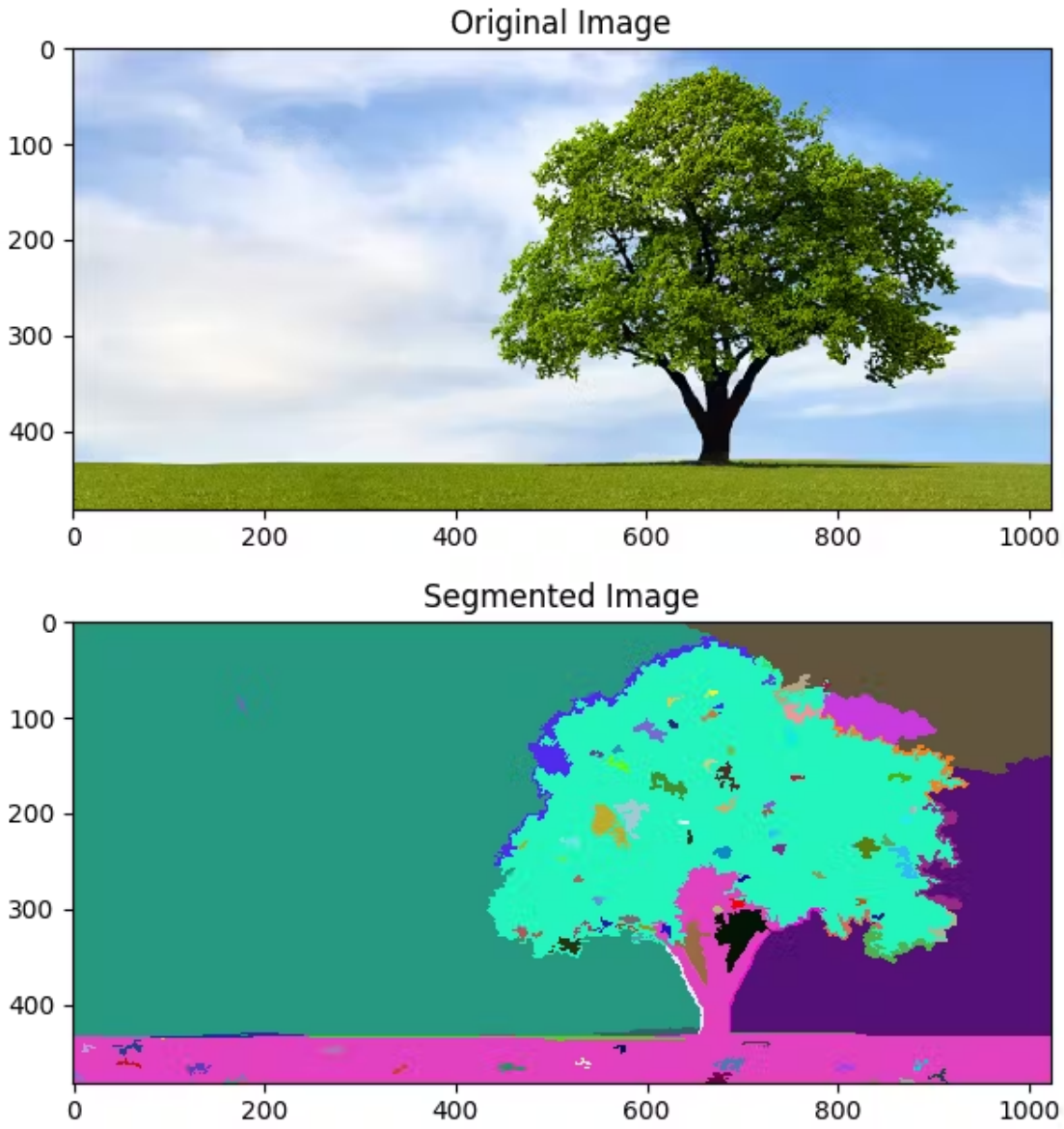
Split and merge segmentaion

- **Split and merge segmentation** is a region-based segmentation technique that recursively divides an image into smaller regions until a stopping criterion is met and then merges similar regions to form larger regions.
- The method involves splitting the image into smaller blocks or regions and then merging adjacent regions that meet certain similarity criteria, such as similar color or texture.
- Split and merge segmentation is a simple and efficient technique for segmenting images, but it may not work well for complex images with overlapping or irregular regions.

Graph-based segmentation

- **Graph-based segmentation** is a technique used in image processing to divide an image into regions based on the edges or boundaries between regions.
- The method involves representing the image as a graph, where the nodes represent pixels, and the edges represent the similarity between pixels.
- The graph is then partitioned into regions by minimizing a cost function, such as the normalized cut or minimum spanning tree.

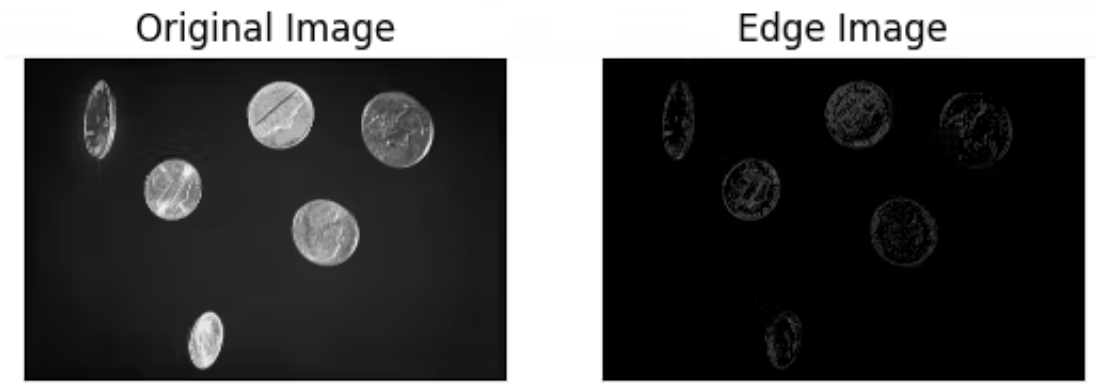
cont...


Original Image
0
100
200
300
400
0
200
400
600
800
1000
Segmented Image
0
100
200
300
400
0
200
400
600
800
1000

Edge-based Segmentation

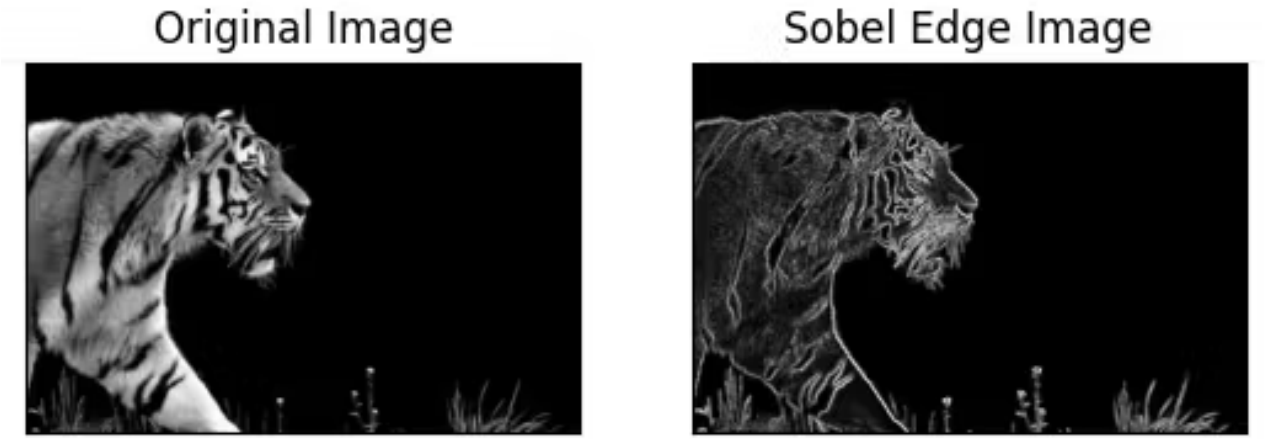
- **Edge-based segmentation** is a technique used in image processing to identify and separate the edges of an image from the background.
- The method involves detecting the abrupt changes in intensity or color values of the pixels in the image and using them to mark the boundaries of the objects.

Canny-edge detection



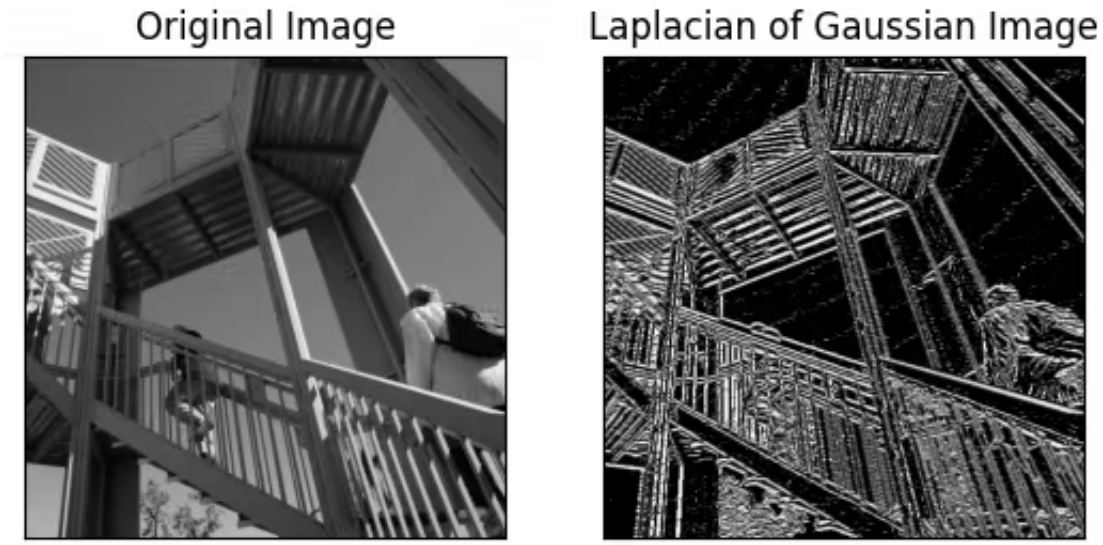
- **Canny edge detection** is a popular method for edge detection that uses a multi-stage algorithm to detect edges in an image.
- The method involves smoothing the image using a Gaussian filter, computing the gradient magnitude and direction of the image, applying non-maximum suppression to thin the edges, and using hysteresis thresholding to remove weak edges.

Sobel edge detection



- **Sobel edge detection** is a method for edge detection that uses a gradient-based approach to detect edges in an image.
- The method involves computing the gradient magnitude and direction of the image using a Sobel operator, which is a convolution kernel that extracts horizontal and vertical edge information separately.

Laplacian of Gaussian (LoG) edge detection



- **Laplacian of Gaussian (LoG) edge detection** is a method for edge detection that combines Gaussian smoothing with the Laplacian operator.
- The method involves applying a Gaussian filter to the image to remove noise and then applying the Laplacian operator to highlight the edges.
- LoG edge detection is a robust and accurate method for edge detection, but it is computationally expensive and may not work well for images with complex edges.

Clustering

- **Clustering** is one of the most popular techniques used for image segmentation, as it can group pixels with similar characteristics into clusters or segments.
- The main idea behind clustering-based segmentation is to group pixels into clusters based on their similarity, where each cluster represents a segment.
- This can be achieved using various clustering algorithms, such as K means clustering, mean shift clustering, hierarchical clustering, and fuzzy clustering.

K-means clustering

- **K-means clustering** is a widely used clustering algorithm for image segmentation.
- In this approach, the pixels in an image are treated as data points, and the algorithm partitions these data points into K clusters based on their similarity.
- The similarity is measured using a distance metric, such as Euclidean distance or Mahalanobis distance.
- The algorithm starts by randomly selecting K initial centroids, and then iteratively assigns each pixel to the nearest centroid and updates the centroids based on the mean of the assigned pixels.
- This process continues until the centroids converge to a stable value.

cont...

Original Image

Segmented Image when K = 10

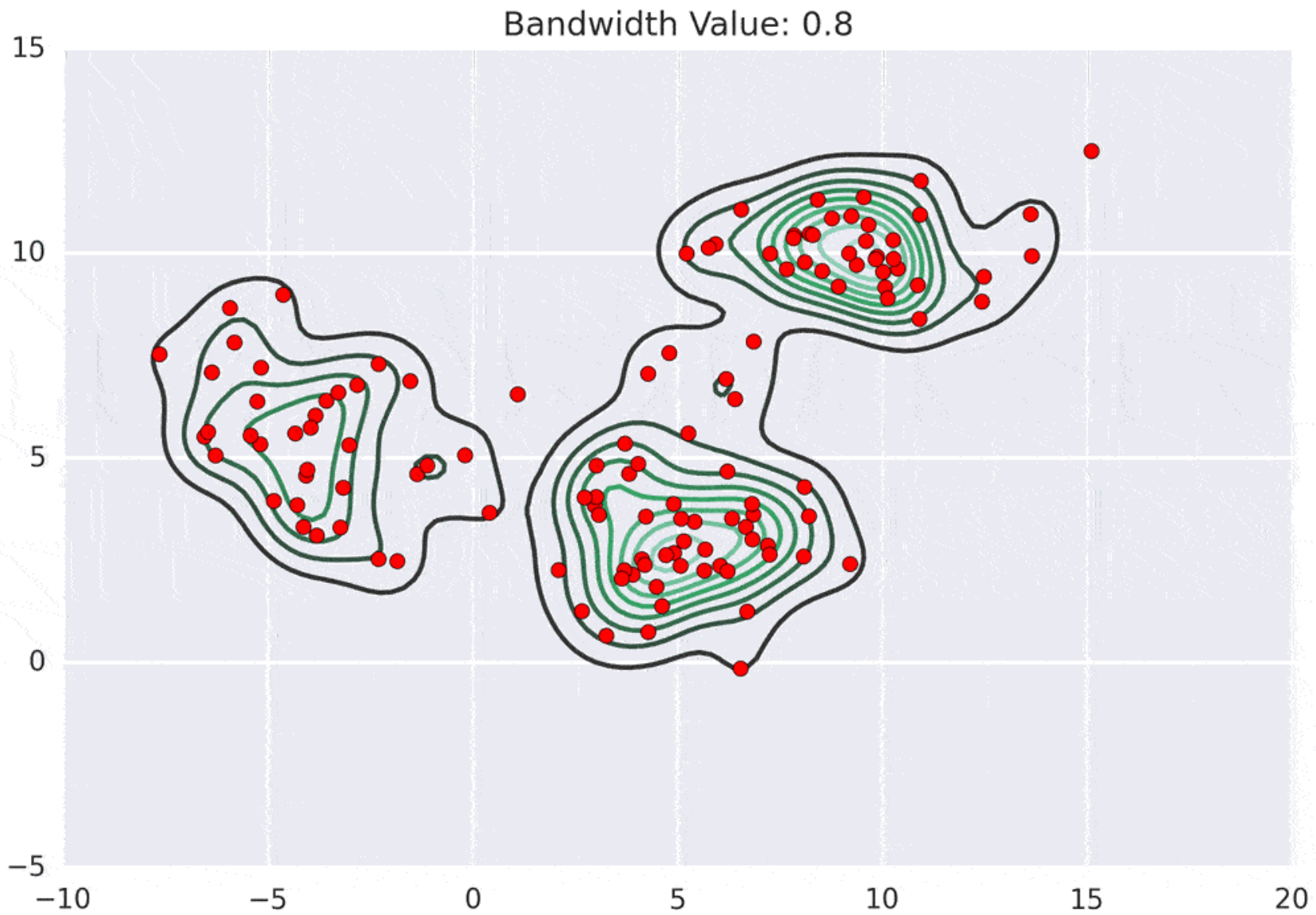
Segmented Image when K = 4

Segmented Image when K = 2

Mean-shift clustering

- **Mean shift clustering** is another popular clustering algorithm used for image segmentation.
- In this approach, each pixel is represented as a point in a high-dimensional space, and the algorithm shifts each point toward the direction of the local density maximum.
- This process is repeated until convergence, where each pixel is assigned to a cluster based on the nearest local density maximum.

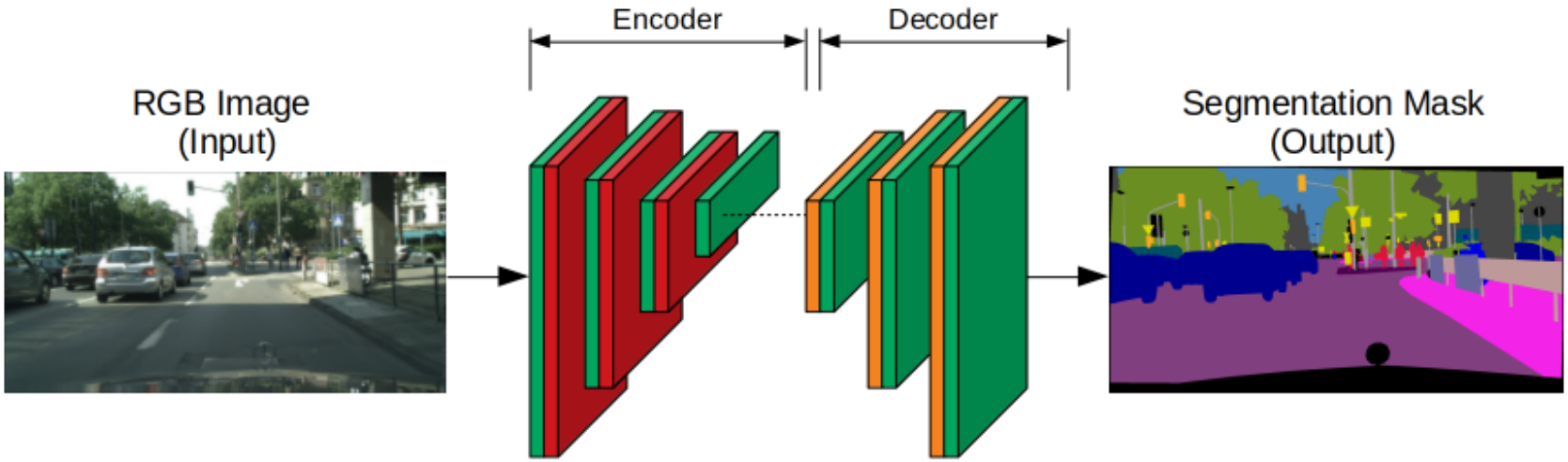
cont...



Deep Learning Techniques

- Neural networks also provide solutions for image segmentation by training neural networks to identify which features are important in an image, rather than relying on customized functions like in traditional algorithms.
- Neural nets that perform the task of segmentation typically use an encoder-decoder structure.
- The encoder extracts features of an image through narrower and deeper filters.
- If the encoder is pre-trained on a task like an image or face recognition, it then uses that knowledge to extract features for segmentation (transfer learning).
- The decoder then over a series of layers inflates the encoder's output into a segmentation mask resembling the pixel resolution of the input image.

cont...



cont...

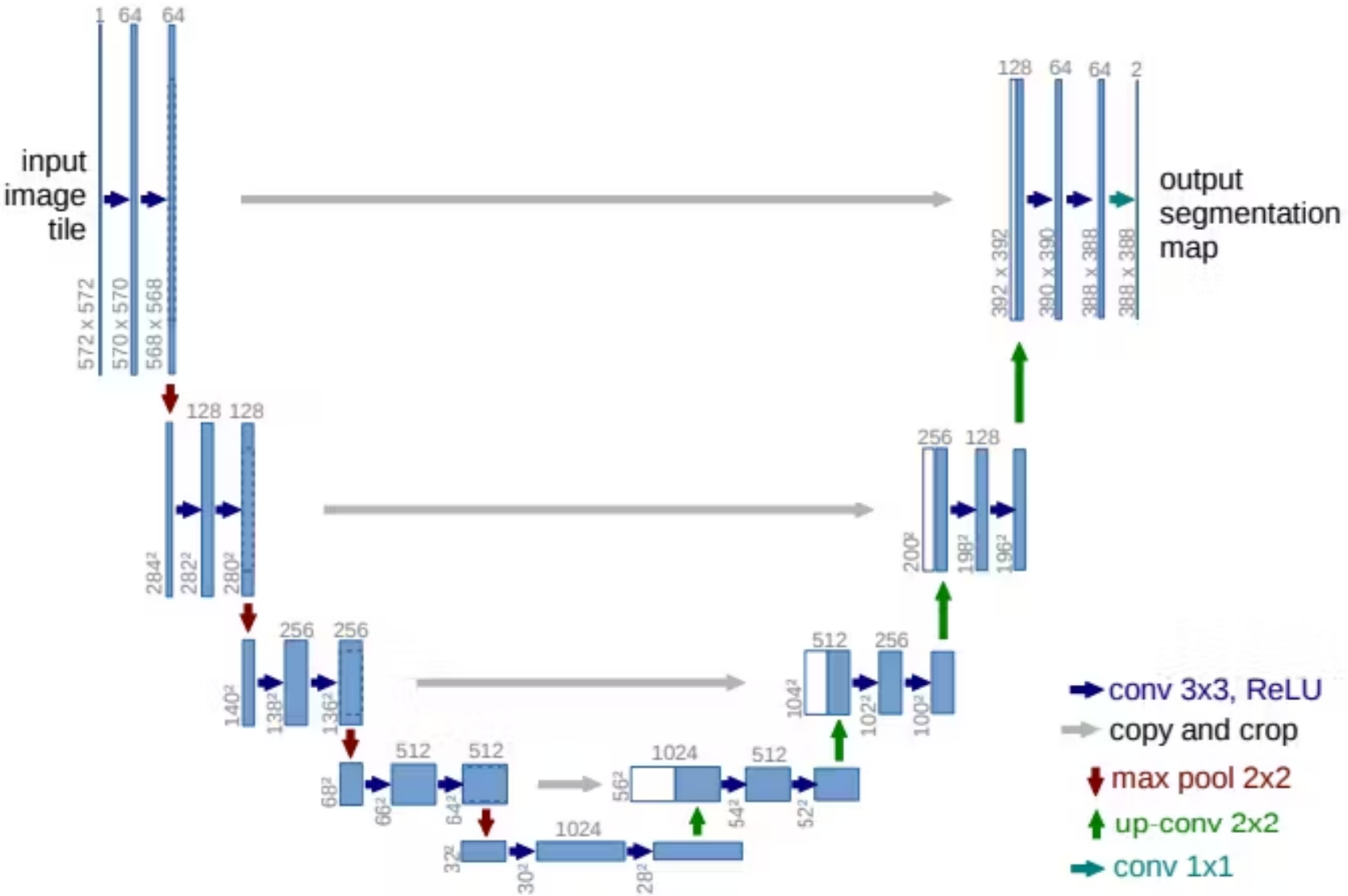
Deep Learning Techniques

Technique	Definition	Key Features	Applications
U-Net	A fully convolutional network designed for precise segmentation tasks, especially in medical imaging.	- Encoder-decoder structure with skip connections. - Combines low-level and high-level features for finer segmentation. - Retains spatial context effectively.	Medical imaging (tumor detection, organ segmentation).
SegNet	A deep convolutional network optimized for semantic segmentation tasks.	- Encoder-decoder structure. - Uses max-pooling indices for efficient upsampling. - Reduces computational complexity.	Scene understanding, autonomous driving.
DeepLab	A CNN architecture that leverages atrous (dilated) convolution for dense segmentation tasks.	- Incorporates features from all convolutional layers. - Atrous convolution for capturing contextual information. - Efficient computation.	Object detection, road scene analysis.

U-Net

- **U-Net** is a modified, fully convolutional neural network.
- It was primarily proposed for medical purposes, i.e., to detect tumors in the lungs and brain.
- It has the same encoder and decoder.
- The encoder is used to extract features using a shortcut connection, unlike in fully convolutional networks, which extract features by upsampling.
- The shortcut connection in the U-Net is designed to tackle the problem of information loss.
- In the U-Net architecture, the encoders and decoders are designed in such a manner that the network captures finer information and retains more information by concatenating high-level features with low-level ones.
- This allows the network to yield more accurate results.

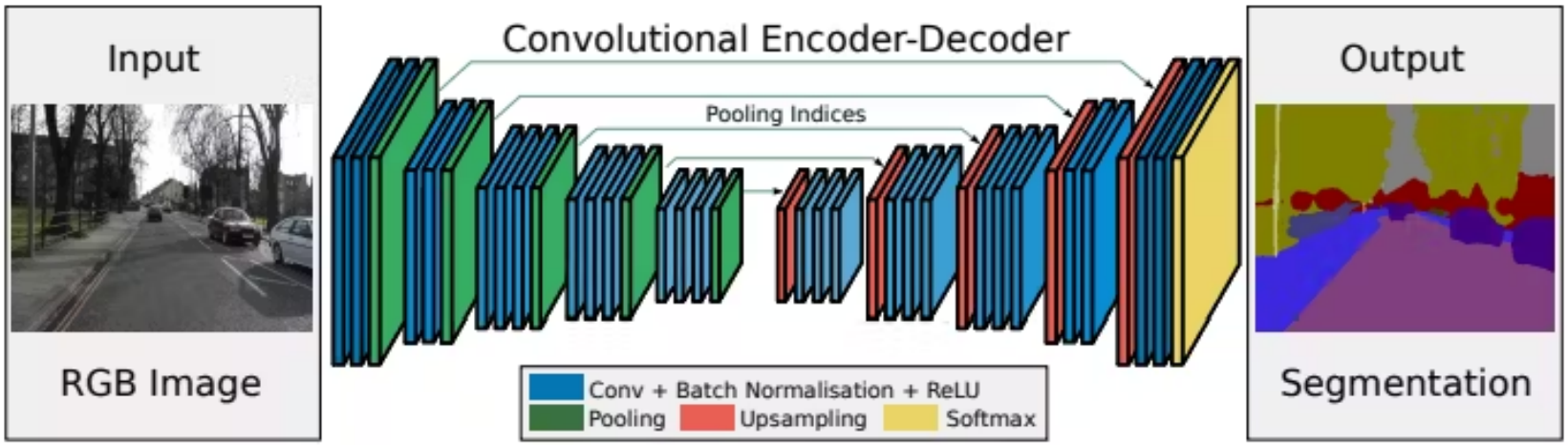
cont...


input image tile
572 x 572
570 x 570
568 x 568
1 64 64
128 128
284²
282²
280²
256 256
140²
138²
136²
512 512
68²
66²
64²
1024
32²
30²
28²
512
56²
54²
52²
256
104²
102²
100²
256 128
200²
198²
196²
128 64 64 2
392 x 392
390 x 390
388 x 388
388 x 388
output segmentation map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

SegNet

- SegNet is also a deep fully convolutional network that is designed especially for semantic pixel-wise segmentation.
- Like U-Net, SegNet's architecture also consists of encoder and decoder blocks.
- The SegNet differs from other neural networks in the way it uses its decoder for upsampling the features.
- The decoder network uses the pooling indices computed in the max-pooling layer which in turn makes the encoder perform non-linear upsampling.
- This eliminates the need for learning to upsample.
- SegNet is primarily designed for scene-understanding applications.

cont...


Input
RGB Image
Convolutional Encoder-Decoder
Pooling Indices
Conv + Batch Normalisation + ReLU
Pooling
Upsampling
Softmax
Output
Segmentation