Computer Vision Project Report

Image Segmentation

Aniket Yadav
IIIT Vadodara (202051025)
Vadodara, Gujarat, India
202051025@iiitvadodara.ac.in

Avnit Kumar Anand
IIIT Vadodara (202051040)
Vadodara, Gujarat, India
202051040@iiitvadodara.ac.in

Ayush Arora
IIIT Vadodara (202051041)
Vadodara, Gujarat, India
202051041@iiitvadodara.ac.in

Kailash Kumar
IIIT Vadodara (202051093)
Vadodara, Gujarat, India
202051093@iiitvadodara.ac.in

***Abstract*— Image segmentation is a fundamental task in computer vision, playing a crucial role in various applications such as object recognition, scene understanding, and medical image analysis. This report provides an overview of recent advancements in image segmentation techniques, highlighting their significance and impact on improving the accuracy and efficiency of computer vision systems.**

The report begins by introducing the concept of image segmentation and its importance in extracting meaningful information from visual data. It then delves into traditional segmentation methods, such as region-based and edge-based techniques, discussing their strengths and limitations. Subsequently, the focus shifts to the emergence of deep learning-based approaches, particularly convolutional neural networks (CNNs), which have revolutionized the field of image segmentation.

A detailed exploration of popular deep learning architectures for image segmentation, such as U-Net, Mask R-CNN, and DeepLab, is presented, accompanied by a discussion on their respective strengths and applications. The report also addresses the challenges associated with training deep neural networks for image segmentation, including the need for large annotated datasets and computational resources..

I. INTRODUCTION

The purpose of image segmentation, a critical problem in computer vision, is to separate a picture into distinct, meaningful sections or objects. In many applications, including robotics, medical imaging, object recognition, tracking, and detection, it is an essential task. Image segmentation can be done using a variety of ways, from conventional procedures to deep learning-based strategies. The development of deep learning has greatly increased picture segmentation's efficiency and accuracy.

II. WHAT IS IMAGE SEGMENTATION?

The technique of splitting a picture into several homogeneous and significant areas or objects according to their innate properties—such as colour, texture, form, or brightness—is known as image segmentation. 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. The task of segmentation can further be done in two ways:

1. Similarity: As the name suggests, the segments are formed by detecting similarity between image pixels. It is often done by thresholding (see below for more on thresholding). Machine learning algorithms (such as clustering) are based on this type of approach for image segmentation.
2. Discontinuity: Here, 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.

.

III. TYPES OF SEGMENTATION

Based on the type and quantity of information that needs to be extracted from the image, image segmentation modalities are classified into three categories: instance, semantic, and panoptic. Let's examine these different image segmentation techniques.

It would also be more convenient to have more knowledge about backgrounds and objects in order to comprehend the three kinds of image segmentation.

While the background of an image refers to things that cannot be counted, such as the sky, water bodies, and other similar elements, objects are the identifiable items in the image that can be separated from each other by assigning unique IDs. It is simpler to comprehend the various image segmentation techniques and their associated uses when objects and backgrounds are distinguished.

A. Instance Segmentation

One kind of image segmentation is instance segmentation, which entails identifying and segmenting every object in an image. With the additional duty of segmenting the object's boundaries, it is comparable to object detection. The algorithm separates overlapping objects without knowing the region's class. Applications that require the identification and tracking of individual objects can benefit from instance segmentation.

B. Semantic Segmentation

A sort of image segmentation known as "semantic segmentation" labels each pixel in a picture with the proper class label without taking any further context or information into account. The aim is to densely label the image by giving a label to each pixel in the picture. Using an image as input, the algorithm creates a segmentation map in which the picture's pixel values (0,1,...255) are converted to class labels (0,1,...n). It is helpful in situations where it's crucial to distinguish between various classes of roadside things.

IV. Image Segmentation Techniques

A. Traditional Techniques

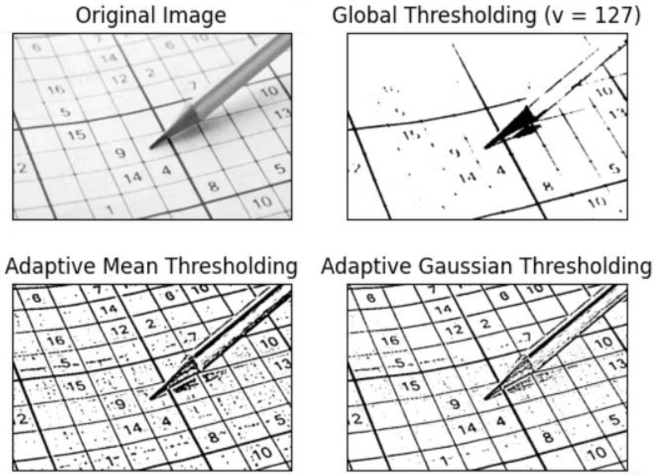
For many years, computer vision has employed conventional image segmentation techniques to extract relevant information from images. These methods find areas of an image with similar features like colour, texture, or brightness using mathematical models and algorithms. Conventional picture segmentation methods are often easy to apply and computationally efficient. They are frequently employed in applications like object detection, tracking, and recognition that call for quick and precise image segmentation. We will examine a few of the most popular methods in this section.

1. Thresholding

Thresholding is one of the simplest image segmentation methods. Here, 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. The two commonly used thresholding methods are:

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.

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.

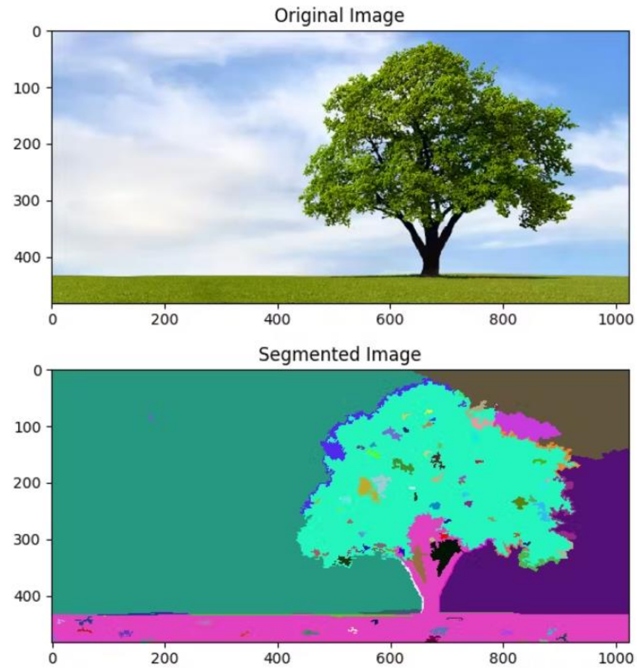


2. Region-based Segmentation

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:

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.

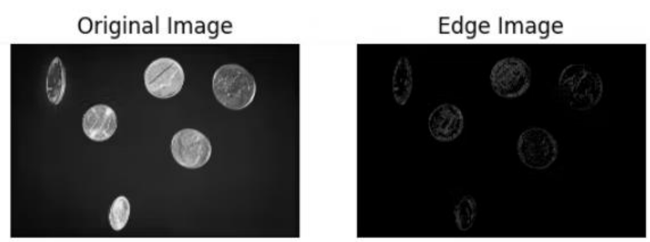
Graph-based segmentation is a technique used in image processing to divide an image into regions based on the edges or boundaries between regions. The method involves representing the image as a graph, where the nodes represent pixels, and the edges represent the similarity between pixels. The graph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



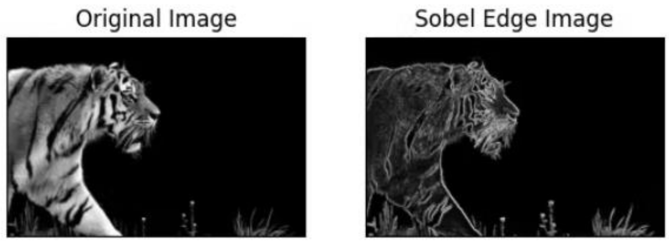
3. Edge-based Segmentation

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. The two most common edge-based segmentation techniques are:

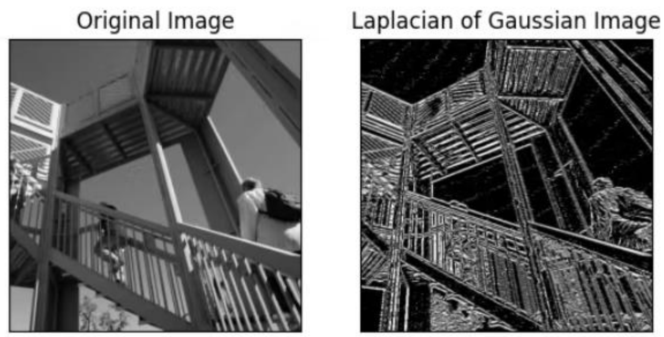
Canny edge detection is a popular method for edge detection that uses a multi-stage algorithm to detect edges in an image. The method involves smoothing the image using a Gaussian filter, computing the gradient magnitude and direction of the image, applying non-maximum suppression to thin the edges,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



Sobel edge detection 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 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.

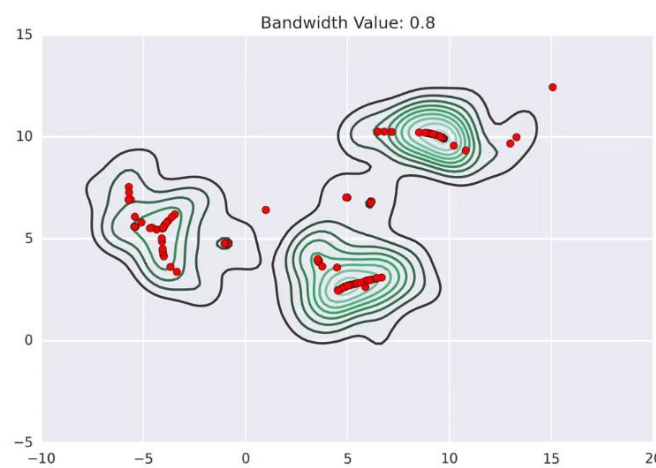


4. Clustering

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.

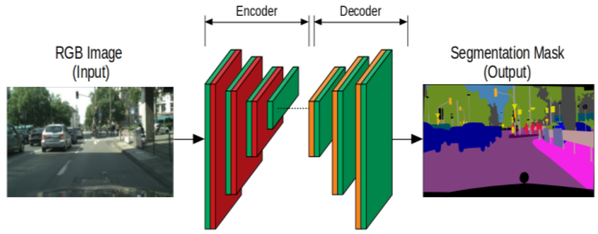
K-means clustering is a widely used clustering algorithm for image segmentation. In this approach, the pixels in an image are treated as data points, and the algorithm partitions these data points into K clusters based on their similarity. The similarity is measured using a distance metric, such as Euclidean 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.

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.



B. Deep Learning Techniques

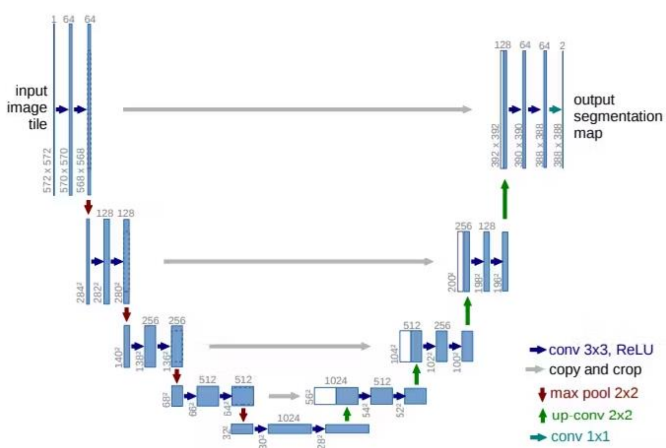
Neural networks also provide solutions for image segmentation by training neural networks to identify which features are important in an image, rather than relying on customized functions like in traditional algorithms. Neural nets that perform the task of segmentation typically use an encoder-decoder 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.



The task of segmentation is one that many deep learning models excel at doing with high reliability. Let's examine a couple of them.

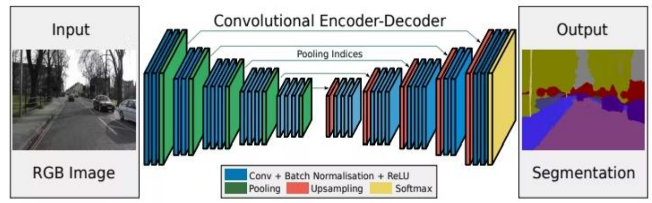
1. U-Net

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.



2. 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.



3. DeepLab

DeepLab is primarily a convolutional neural network (CNN) architecture. Unlike the other two networks, it uses features from every convolutional block and then concatenates them to their deconvolutional block. The neural network uses the features from the last convolutional block and upsamples it like the fully convolutional network (FCN). It uses the atrous convolution or dilated convolution method for upsampling. The advantage of atrous convolution is that the computation cost is reduced while capturing more information.

V. Observation

In the assessment of image segmentation techniques for object recognition in computer vision, it is evident that state-of-the-art algorithms like U-Net and Mask R-CNN offer high accuracy in delineating object boundaries. However, precision is heavily contingent on the quality and diversity of training data, posing challenges in scenarios with complex or uncommon objects. A notable trade-off emerges between the impressive accuracy of models and their computational efficiency, raising concerns for real-time applications in resource-constrained environments. Semantic segmentation encounters difficulties in handling intricate scenes, particularly in scenarios with overlapping or ambiguous objects. Instance segmentation, exemplified by Mask R-CNN, showcases advancements in object counting and tracking, but challenges persist in densely packed scenes. The robustness of segmentation models to varied conditions, transfer learning for generalization, and adaptability to diverse environments are critical considerations for future advancements in this field, aiming to enhance the applicability of segmentation algorithms in real-world settings. Ongoing research efforts focusing on precision, efficiency, and adaptability are pivotal for addressing these challenges.

VI. References

1. https://en.wikipedia.org/wiki/Image_segmentation

2. https://www.analyticsvidhya.com/blog/2019/04/introduction-image-segmentation-techniques-python/