IMAGE SEGMENTATION

An Essay by Ritwik Guha

INTRODUCTION TO IMAGE SEGMENTATION

Image segmentation is a computer vision technique that partitions a digital image into discrete groups of pixels—image segments—to inform object detection and related tasks. Computer vision is a field of artificial intelligence (AI) that uses machine learning and neural networks to teach computers and systems to derive meaningful information from digital images, videos and other visual inputs—and to make recommendations or take actions when they see defects or issues. Image segmentation is thus a necessary prerequisite in various applications such as object recognition, tracking, and detection, medical imaging, and robotics.

The task of segmentation can further be done in the following ways:

* **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**](https://developers.google.com/machine-learning/clustering/overview)) are based on this type of approach for image segmentation.
* **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**](https://encord.com/glossary/edge-detection-definition/) techniques to obtain intermediate segmentation results that may be processed to obtain the final segmented image.

Image thresholding is a technique that simplifies a grayscale image into a binary image by classifying each pixel value as either black or white based on its intensity level or gray-level compared to the threshold value. This technique reduces the image to only two levels of intensity, making it easier to identify and isolate objects of interest. Binary image conversion allows for efficient processing and analysis of images, enabling various computer vision applications such as edge detection and pattern recognition.

Deep learning-based segmentation: Deep learning techniques, such as [Convolutional Neural Networks (CNNs)](https://viso.ai/deep-learning/convolutional-neural-networks/), have revolutionized image segmentation by providing highly accurate and efficient solutions. These techniques use a hierarchical approach to image processing, where multiple layers of filters are applied to the input image to extract high-level features

While exploring OpenCV, I found 2 practical applications of thresholding – Simple Thresholding and Adaptive Thresholding. Further the cv.threshold function and the different types of simple thresholding such as cv.THRESH\_BINARY, cv.THRESH\_BINARY\_INV, etc. To deal with real life images that are not taken in perfect lighting adaptive thresholding is used so that effective processing is done – cv.adaptiveThreshold. Another technique which used histograms was Otsu’s Binarization. The basic idea was that it automatically found a threshold value rather than choosing it beforehand. It seemed to work exceptionally well while processing noisy images.

APPLICATIONS IN AUTONOMOUS DRIVING

Car companies like Tesla, Morris Garages (Mg), Audi, Hyundai, etc., are launching cars with different autonomous levels. This autonomous level starts from 0, with no assistance system, and goes up to level 5, which describes fully autonomous driving. The first step in most autonomous vehicle systems is object recognition or semantic segmentation.

Image segmentation is the partitioning of an image (car dashboard image, surveillance camera image.) into segments or sets of pixels to identify and locate different objects or entities in the image. Each pixel is assigned a label so that pixels with the same label share some common characteristics like colour or texture. Image segmentation can be divided into two groups:

**Semantic segmentation** classifies pixels based on their semantic meaning, treating all objects within a category as one entity.

**Instance segmentation** on the other hand, distinguishes between different instances of the same class, allowing for more precise object identification and differentiation.

Image Segmentation is the partitioning of an image into multiple regions and then extracting the specified region/area, also known as the Region of Interest (ROI). If we want to segregate a road's region or the car's region from that image, then that region is our ROI.

Pixels having similar attributes will be masked with the same colour, and pixels with different characteristics will have a different colour. For example, if we consider the pixels of the road from the image, then all the pixels in the road will have the same attribute, and we will mask it with a single colour. Similarly, we will mask the pixels of the sky with a single colour. After colour clustering, the road will look pink, and the sky will look blue in our images.

Relevant use cases:

In driving assistance systems, the real time implementation of **pedestrian detection system** cannot be easily solved. Semantic image segmentation has shown a promising performance. However, the problem remains in the labelled image data availability. There is still a lack of pedestrian data in terms of semantic image segmentation.

For **lane detection** our input is a set of images, but out labels are the same images, with overlayed segmented lines. Here, each colour representing one line. This means we're not doing just image segmentation, but instance segmentation. Thanks to this, we can tell apart the left and right lines, but also solid from dashed, curves from straight, etc...

Further t**raffic sign detection** can be performed by detecting appropriate colours such as bright red, and **object detection** can be performed similarly to pedestrian detection.

CHALLENGES AND FUTURE DIRECTIONS

One of the main challenges in image segmentation is the variability of images that can affect the quality and accuracy of the segmentation results. Images can vary in terms of size, resolution, contrast, brightness, noise, distortion, occlusion, and illumination. In the case of autonomous driving since the stakes are extremely high, and the margin of error is very low, these challenges need to overcome.

Varying light conditions are a problem which can be potentially solved by integrating headlights into the AI model, and controlling the headlight intensity to keep the camera brightness constant with changing environmental light.

Real time delays also cause issues because it takes a decent amount of time for image to be input, get processed into a set of lines and curves, be fed into the steering/accelerator, and for the action to then take place. This becomes even harder at higher speeds where all this has to keep pace with the increasing velocity of the vehicle. To solve this, increasing camera ranges would help as the input would be received earlier, for straight roads and gradual turns. For immediate responses and quick adjustments, a better image segmentation model or even a software that detects objects based on a priority order and for example processes pedestrians first could help. Rather than waiting for the image to be fully processed and then sending the output, if at each stage a partial output could be sent and as soon as even a hint of a pedestrian/object is detected the brakes could be applied.

Occlusions such as rain, hail, leaves, other vehicles coming extremely close, etc. could possibly hinder camera vision and cause problems. A possible solution would be a sort of indicator to the driver that the camera is being blocked similar to a seatbelt indicator. For fully autonomous vehicles this requires a more complex solution. Changing the location of the camera or having multiple cameras might help.

AUTO-SEGMENTATION WITH SAM

Auto-segmentation refers to the process of automatically segmenting an image without human intervention. Auto-segmentation with Meta’s [**Segment Anything Model**](https://arxiv.org/abs/2304.02643) (SAM) has instantly become popular as it shows remarkable performance in image segmentation tasks. It is a single model that can easily perform both interactive segmentation and automatic segmentation. Since SAM is trained on a diverse, high-quality dataset, it can generalize to new types of objects and images beyond what is observed during training. This ability to generalize means that by and large, practitioners will no longer need to collect their segmentation data and fine-tune a model for their use case.