International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1284-1287 www.ijaem.net ISSN: 2395-5252

Image Segmentation using R-CNN

Ms. Chanchal [1], Naman Jain [2], Shaksham Agarwal [3], Shikhar Goyal [4], Tejveer [5]

[1] *Faculty at Computer Science and Engineering Department, Raj Kumar Goel Institute of Technology, AKTU*
[2,3,4,5] *Student at Computer Science and Engineering Department, Raj Kumar Goel Institute of Technology, AKTU*

Submitted: 10-05-2022 Revised: 19-05-2022 Accepted: 22-05-2022

ABSTRACT
Segmentation is an important part of computer vision. It is used in many applications like medical imaging, computer-guided surgical, machine vision, object detection, surveillance, content-based browsing and augmented reality applications. To simplify and make the video representation more understandable and easier to analyze, it is important to know how to determine plausible segmentation methods and the corresponding algorithmic techniques. The expected segmentation quality of an application is dependent on its granularity, as well as the requirements that are related to shape precision or temporal coherence.

Image-processing techniques are becoming more important due to the advancement of computer technology in many different applications. Image segmentation is an important topic in image processing. It is also a focus and hotspot for image processing techniques. There are many general-purpose algorithms and techniques for image segmentation. These techniques are not able to solve all image segmentation problems. They must be used in conjunction with domain knowledge to solve a specific image segmentation problem.

I. BACKGROUND

Image segmentation is an important step of image processing. It can be found everywhere if you want to analyze the contents of the image. Image segmentation is needed to identify if an indoor image contains a chair or a person. We can also analyze each object individually using image segmentation. Image segmentation is usually used before pattern recognition, feature extraction and compression.

Image segmentation refers to the division of an image into various groups. Image segmentation by clustering has been the subject of much research. There are many methods, but the most widely used is the K-Means Clustering algorithm.

Image segmentation refers to the process of grouping or separating an image into separate parts. These parts correspond to things that humans can easily separate and view as separate objects. Computers are not capable of correctly recognising objects and many methods have been created to segment images. Segmentation is based on different features in an image. This could be information about the color of pixels used to create histograms or information about edges, boundaries, or texture information.

Image segmentation results in a collection of segments that cover the entire image or a set contours taken from the image. Each of the pixels within a given region is similar in terms of a characteristic or computed property. This could be color texture, intensity, or other characteristics. The color of adjacent regions is significantly different from the same characteristic.

II. WHAT IS IMAGE SEGMENTATION?

Image segmentation refers to a type of digital image processing that focuses on the division of an image into parts according to their properties and features. Image segmentation's primary purpose is to simplify an image so that it can be analyzed more easily.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1284-1287 www.ijaem.net ISSN: 2395-5252

Image Segmentation, a computer vision task, is the process of dividing a digital image into multiple segments (also known as image objects). This segmentation can be used to locate boundaries and objects (lines, curves etc.).

III. TWO MAIN TYPES OF IMAGE SEGMENTATION FALL UNDER MASK R–CNN -

1. Semantic Segmentation
2. Segmentation of an instance

3.1 Semantic Segmentation :

Semantic segmentation categorizes each pixel in a set of categories, without distinguishing object instances. In other words, semantic segmentation deals with the identification/classification of similar objects as a single class from the pixel level.

3.2 Segmentation of an instance : Instance Segmentation or Instance Recognition is a method of accurately segmenting an instance and detecting all objects in an image. This is the combination of object detection and object localization. This segmentation, in other words, allows for a clear distinction between objects classified as similar.

IV. TYPES OF IMAGE SEGMENTATION

4.1 Segmentation based on region:-

A simple way to separate objects is to use their pixel values. Important note: If there is a stark contrast between the objects and their background, the pixel values for each object will differ from the background. We can define a threshold value in this instance. You can classify the pixel values that fall below or above this threshold as an object or background. Threshold Segmentation is a technique that allows you to segment the image. We can divide an image into two areas (object and background) by setting a single threshold value. This is called the global threshold. Multiple objects and backgrounds must be defined as multiple thresholds. These thresholds collectively are known as the local threshold.

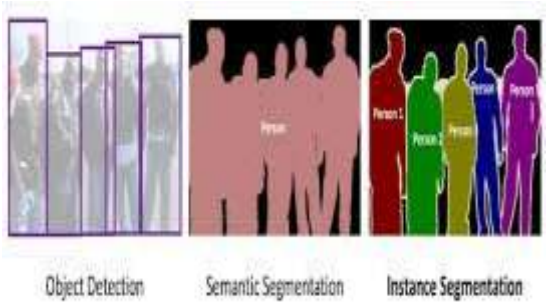
4.2 Edge Detection Segmentation:-

What is the boundary that separates two objects within an image? An edge exists between adjacent areas with different grayscale values (pixel value). These edges are considered to be the discontinuous local features in an image.

This discontinuity can be used to identify edges and thus define the boundary of an object. This allows us to detect the shapes of multiple objects in an image.

4.3 Image Segmentation Based on Clustering:-

Clustering refers to the process of grouping data points into groups so that they are more like data points from other groups. These groups are called clusters. K-means is one of the most popular clustering algorithms.



V. METHODOLOGY

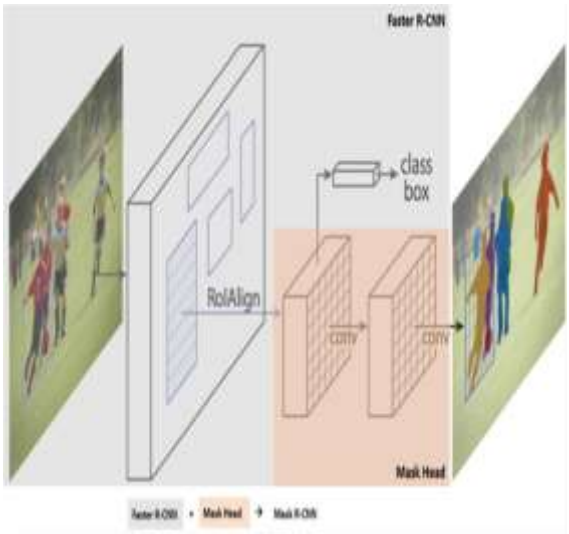
5.1 Mask R-CNN :-

Mask R-CNN extends the Faster R-CNN object recognition architecture. Mask R-CNN is an extension to the Faster R-CNN outputs. Faster R-CNN generates two things from each object in an image: Its class, and its bounding box coordinates. Mask R-CNN adds another branch to this, which also outputs the object mask.

On top of Faster RCN, the Mask R-CNN framework was built. Mask R-CNN will return the object mask for any image.

- First, let's quickly see how Faster RCN works. This will allow us to understand Mask R-CNN.
- R-CNN uses ConvNet for feature mapping extraction from images faster
- The feature maps are then sent through the Region Proposal Network (RPN), which returns the candidate bounding box.
- To bring all candidates to the same size, we apply an RoI pooling layer to these candidate bounding boxes.
- Finally, the proposals are sent to a fully connected layer that classifies and outputs the bounding boxes of objects.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1284-1287 www.ijaem.net ISSN: 2395-5252



5.2 Backbone Model :-

Similar to Faster R-CNN's ConvNet to extract feature maps from an image, Mask R-CNN uses the ResNet 101 architecture for extracting features from images. The ResNet 101 architecture is used to extract features from an image. These features are used as input to the next layer.

5.3 Region Proposal Network :-

We now take the feature maps from the previous step and apply the region proposal network (RPM). This predicts whether an object is found in the region. This step determines which regions or feature maps the model predicts will contain an object.

5.4 Region of Interest (RoI) :-

You might find that the regions created by the RPN have different shapes. We apply a pooling layer to convert all regions to the same form. These regions are then connected to form a network that predicts the class label as well as the bounding boxes.

The steps are very similar to Faster R-CNN. The difference between these frameworks is now. This mask is also generated by Mask R-CNN.

To reduce the time required to compute this, we first calculate the region of interest. We compute the Intersection over Union for all predicted regions using the ground truth boxes. This is how we can compute IoU:

IoU = Area at the intersection/Area of the union

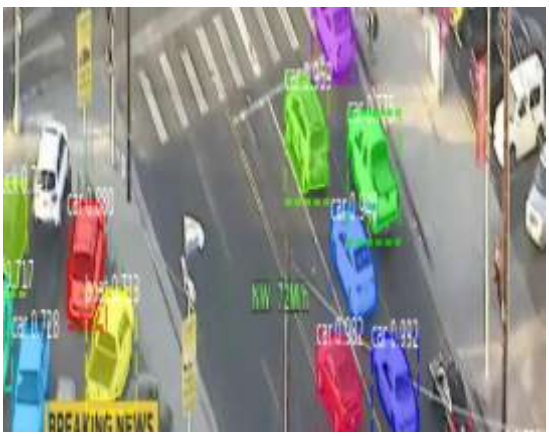
Only if the IoU exceeds or equals 0.5, then we consider it a region of interest. We ignore that region. This is done for all regions. Then, we select only those regions where the IoU exceeds 0.5.

5.5 Segmentation Mask :-

Once we have the RoIs calculated based on IoU values, it is possible to add a mask branch into the existing architecture. This returns the segmentation mask of each object-containing region. This returns the region's segmentation mask in a size of 28 X 28, which is then scaled for inference.

VI. CONCLUSION :-

We have used MASK-RCNN to do image segmentation in some images and the result is as following -



International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 5 May 2022, pp: 1284-1287 www.ijaem.net ISSN: 2395-5252

VII. FUTURE SCOPE :-

Image segmentation is widely used in these applications and there is a lot of scope for development -

7.1 Remote Sensing:-

Remote sensing is a rapidly growing field for data detection. Remote sensors can detect energy reflected by the earth and extract data. Satellite8 is an example . Satellites collect data from the earth's surface and can then be used to analyze it in many ways, such as which areas are green or where there is water. We can also see which areas have more people. The remote sensing of space is responsible for monitoring the changes in our environment like deforestation(cutting down the trees), ecosystem degradation, changes in the forest carbon stocks and carbon recycling, among others. These models capture the dynamics of changes and their impacts on global ecology.

7.2 Medical imaging :-

Medical imaging is the process of creating an optical representation of human organs. Medical imaging is a key component of modern medicine. We can now see the inner structure behind the skin and bones. It can detect any problems and provide a treatment. Medical imaging includes X-rays (ultrasound, CT scans, and MRI). These techniques will allow us to see the internal structure of our bodies. For instance, X-rays can be used to determine if any bones are broken. The MRI can help identify ligament injuries. Each method has its merits and drawbacks. These are the best guidelines.

7.3 Aerial imaging

Aerial imaging refers to the capture of photographs from aircraft or other flying objects. Drones are now used to take photographs. These drones are extremely useful and can be used to capture photographs in any place. Although its primary purpose was to track wartime activities, it wasn't welcomed by many as it was seen as an invasion of privacy. It became clear that there was more to it than just the bad. Slowly, imaging became the norm.

REFERENCES :-

[1]. Cortez, Diogo, et al. "Image segmentation towards new image representation methods." Signal processing: Image communication 6.6 (1995): 485-498.

[2]. Yanowitz, Shimon D., and Alfred M. Bruckstein. "A new method for image segmentation." Computer Vision, Graphics, and Image Processing 46.1 (1989): 82-95.

[3]. Zhang, Yu-Jin. "An overview of image and video segmentation in the last 40 years." Advances in Image and Video Segmentation (2006).

[4]. Dai, Cheng, et al. "Video scene segmentation using tensor-train faster-rcnn for multimedia iot systems." IEEE Internet of Things Journal 8.12 (2020)

[5]. Wang, Wenguan, et al. "A survey on deep learning techniques for video segmentation." arXiv preprint arXiv:2107.01153 (2021)