International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

LIVE CAPTURING BASED IMAGE SEGMENTATION USING MASK R-CNN

Mr. K. Jeevan Ratnakar, Assistant Professor Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

V. Vedasri, S. Aasritha, V. Venkata Deepthi
UG Students, Department of Information Technology,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh. E-Mail ID: jeevanratnakar@vvit.net

ABSTRACT

The primary goal of this project is to divide images into various regions or parts, frequently based on the properties of the pixels in the image. Deep learning systems are more accurate than traditional techniques. Mask R-CNN is used to derive high-level properties from data that are important for machine learning-based semantic segmentation of images. The computer vision method of image segmentation is crucial. To make image analysis simpler, it entails breaking a visual input into segments. Segments are collections of pixels, or "superpixels," that depict objects or portions of objects. There has recently been a significant amount of work targeted at creating image segmentation approaches using deep learning models due to the success of these models in a variety of vision applications. When using CNN to segment images, portions of the picture are fed into the network, and the convolutional neural network labels the pixels as it processes the input. Fully convolutional networks (FCNS) process different input sizes quicker and use convolutional layers to do so. It entails reducing the input image's dimensions before recovering it using orientation invariance skills. The decoder most notable is the R-CNN or region-based convolutional neural networks, and the most recent method called mask R-CNN, which is capable of getting state-of-the-art results on a variety of object detection tasks.

KEYWORDS

Semantic segmentation, Instance segmentation, Convolutional neural networks, Deep learning, and Image segmentation.

INTRODUCTION

A novel category of image segmentation models with notable performance improvements has been introduced in recent years thanks to Deep Learning models. In the field of image segmentation, a paradigm shift has occurred as a result of the highest accuracy rates being frequently attained by deep learning-based models on well-known benchmarks. Image segmentation is the process of breaking up a visual input into segments in order to make picture analysis easier. To divide an image into different parts or regions, image segmentation is a widely used method in digital image processing and analysis. This technique frequently relies on the properties of the image's pixels. With sets of pixels, segments depict objects or portions of objects. Each object in the image is given its pixel-by-pixel mask through the process of image segmentation, which provides a much more detailed knowledge of the image. Objects and boundaries (lines, curves, etc.) in images are usually located using image segmentation. More precisely, it is the process of assigning each pixel in an image a name so that pixels with the same label share particular characteristics. Image segmentation is the extension of image classification, and it includes localization in addition to categorization. As a result, image segmentation is a subset of image classification because it makes use of a model to pinpoint the position of an object by drawing its boundaries.

The literature has developed a large number of image segmentation algorithms, ranging from the earliest techniques, such as thresholding, histogram-based bundling, region-growing, k-means clustering, and watersheds, to more sophisticated techniques, such as active contours, graph cuts, conditional and Markov random fields, and sparsity-based methods.

Deep learning (DL) networks, on the other hand, have recently generated a new breed of image segmentation models with remarkable performance improvements, frequently achieving the highest accuracy rates on common benchmarks, and causing what many regards as a paradigm shift in the industry.

LITERATURE SURVEY

This part presents related work on image segmentation by a variety of authors, each of whom has a unique perspective on image segmentation.

In their survey, Satish Kumar et al.described the different uses for image segmentation as it relates to computer vision, medical imaging, scanning, identification, etc.

P. Sravani et al.'s study gives a summary of different segmentation methods as well as clustering research. Despite the fact that many techniques have been developed, not all of them are applicable to all types of pictures. Similarity-based groups are created during the segmentation of a

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

picture. A similarity measure called distance has a direct impact on how clusters are formed. In his review of the image segmentation research, H. P. Narkhede outlined various approaches and problems related to digital image processing which is used for different types of pattern recognition. According to Rajeshwar Dass et al. In this study, the authors classify and analyze the primary image segmentation algorithms, concluding that the techniques can be divided into categories based on features like image homogeneity, spatial characteristics of the image continuity, image substance, texture, and image similarity. In her paper, PunamThakare describes different image segmentation methods and goes into great depth about edge detection methods and their assessment. The algorithm that is provided combines edge detector detection and assessment. According to the findings, the nature of the image and its underlying truths affect how often it is recognized.

METHODOLOGY

The most commonly used techniques for image segmentation are thresholding, edge segmentation, and cluster-based segmentation. This study proposes a method that divides an image input into segments to understand the image easily. It does so by creating a MASK R-CNN model utilizing learning techniques. The image is first subjected to classification and then to object detection. In the classification step, the whole image is categorized into classes such as animals, humans, and objects. In next step is object detection, the detection of objects in an image and drawing a rectangle around them is done. Therefore, the key elements of the system can be summed up as follows:

Input: An Image.

Output: For each object in the image it provides its class, object mask and bounding box coordinates.

System: The system will classify and detect the various objects in the image.

1. Data Collection

The dataset is taken from Kaggle which is provided by author Julia Elliott. In 2019 there is a competition that contains images for image segmentation gathered from all over the world. The dataset was updated by the author in 2019 and the dataset contains of total 99,999 images and the size of this data set is less than 5GB. Each object of the image is masked after detection. The dataset characteristics are multivariate and attribute characteristics are real. The dataset is split into training and testing data with ratios of 80% and 20%.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

2. Classification

Generally, classification is a machinelearning task that is used to determine the existing objects in the image. This is a training model to identify which classes or objects are present. Classification is mostly useful at the yes or no level of decision, which means whether an object is in the
image or not.

Localization is a different task from classification, this task is used to determine the position of the classified objects in the image.

3. Object Detection

Object detection is the combination of both classification and localization. This is useful to tell what objects are in the image and where the objects are in the image. To generate the bounding boxes around the object it uses classification.

IMPLEMENTATION

A system is developed using Mask R-CNN. Following is the architecture of Mask R-CNN: It is a stage of state-of-the-art model for instance segmentation and it is developed on top of faster R-CNN.

The architecture comprises:

Backbone Network

Region Proposal Network

Mask Representation

RolAlign

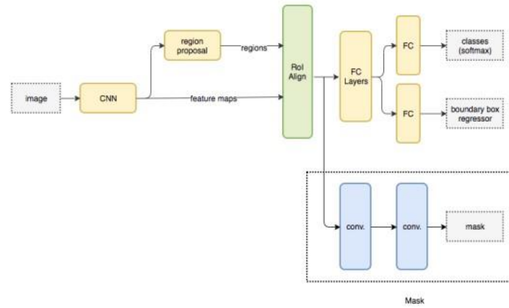

Figure.1. Architecture of Mask R-CNN

Backbone Network: There are two kinds of backbone networks. One is standard ResNet and the other is ResNet-FPN. This consists of multi-layer Rol generation. At every layer, the feature map size is reduced by half and the number of feature maps is doubled.

Region Proposal Network: The previous layer generates the convolution feature map which is passed through the 3*3 convolutional layer. The objectness score is determined by the two parallel branches which take the output of the convolutional layer as input. From this regress of bounding box coordinates are also obtained.

Mask Representation: A mask contains the occupying space information of the object. To predict the mask it uses a fully connected network. An m*m mask representation is obtained as outputs from the ConvNet which takes Rol as input.

Rol Algin: This is used to generate an input for the fully connected network that predicts the mask. The main purpose of the Rol align is to generate a fixed size feature map from the different size feature map.

RESULTS

By implementing the model, below are the result screens.

In the below image, the model identified the object which is a person and set the mask for the image and also set the bounding box in live capturing.

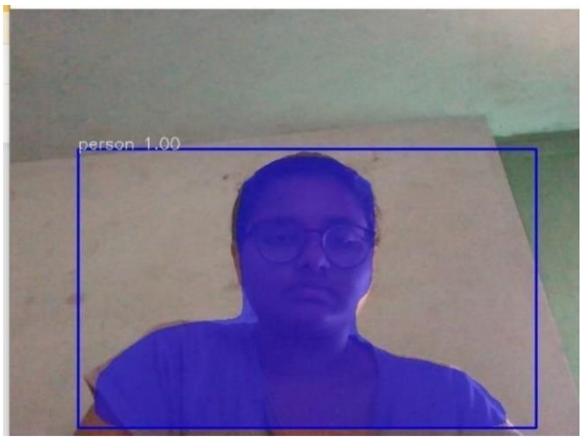

Fig.2. Output when objects are one.

The model has been tested when there are more than one object. Consider the below image, which shows the output when it identifies two objects.

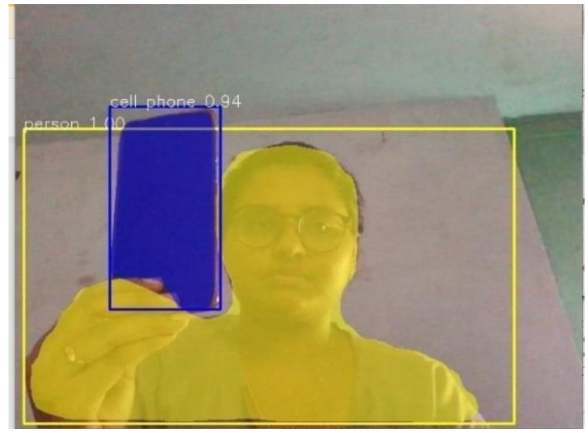

Fig.3. Output when objects are two.

In the Fig.3 the model identified two objects such as person and cell phone. The model has also set the bounding box and mask for the objects. By the results, one can conclude that the model is providing the accurate results.

CONCLUSION

In this article, we have developed a new method for image segmentation using Mask R-CNN by this the objects in the image are identified. Along with the object identification, it also results in the class, object mask, and bounding box coordinates of the object. The Proposed system is an image segmentation model using a regional convolutional neural network for segmenting pre-captured images and live captured images.

LIMITATIONS

However, the Mask R-CNN is a powerful

model, but it also contains some limitations. They are:

Object Obstruction: When an object is covered by another object fully or partially, this can cause inaccurate or missing object instances as result.

Training Data: As most of the deep learning models require large amounts of data for training which results in better performance, Mask R-CNN also falls under these models. This can be difficult because some of the domains do not contain large data sets.

FUTURE SCOPE

As Mask R-CNN is the most accurate and flexible model for image segmentation. To improve its performance and capabilities there is an enhancement, which is Domain Adoption. As already mentioned in the drawbacks some of the domains may not have much training dataset. This could affect the performance of the model, to overcome this problem domain adoption is used. Which means taking knowledge from one domain and using it in other domains.

REFERENCES

[1] F. Meyer, An overview of morphological segmentation. International Journal of Pattern Recognition and Artificial Intelligence, vol. 15, no. 7, pp. 1089-1118, 2001.

[2] J-M. Morel and S. Solimi, Variational Methods in image segmentation; Birkhauser Boston, 1995.

[3] D. Mumford and J. Shah, "Boundary Detection by Minimizing functionals" Image Understanding, 1988.

[4] P. Salembier and F. Marques, "Regionbased representations of image and video: Segmentation tools for multimedia services", IEEE Transacations on circuits and systems for video technology, vol. 9, no. 8, pp. 11471167, Dec. 1999.

[5] J. Serra, "Connectivity on complete lattices", Journal of Mathematical Imaging and Vision 9, pp. 231-251, 1998.

[6] J. Serra, "segmentation", tech. report EMP N-06/03/MM, pp. 25, march 2003.

[7] C. Tzafestas and P. Maragos, "Shape Connectivity", Journal of Mathematical Imaging and Vision, vol. 17, no. 2, pp. 109129, sept 2002.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Comput.

Vis. Pattern Recognit., Jun. 2014, vol. 13, no. 1, pp. 580-587.

[9] R. Girshick, "Fast R-CNN", in Proc, IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015. pp. 1440-1448.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol.
39, no. 6, pp. 1137-1149, Jun. 2017.