IMAGE SEGMENTATION IN ROAD TRAFFICAPPLICATION USING MASK R-CNN

K. Sahithi , J. RajKumari , A. Pranathi , G. SaiPooja

K. Sahithi Student, Information Technology, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India
J. RajKumari Student, Information Technology, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India
A. Pranathi Student, Information Technology, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India
G. SaiPooja Student, Information Technology, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India

ABSTRACT

Automating Road traffic system is a compulsive measure in recent times as the demand of vehicles increasing & congestion is everywhere. An image segmentation mechanism in this application will be useful to automate. The main objective of an image segmentation is to divide an image into many sections for the further analysis, so we can get the only necessary or a segment of information. The partitioning the image will be based on some image features like colour, texture, pixel intensity value etc.

There are several techniques of image segmentation like region based method, edge based method, clustering methods.

This project demonstrates how image segmentation mechanism takes place in any Road traffic application using region based method by Mask R-CNN architecture

Keywords: *Image Localization, Object Detection, Image Segmentation, Machine Learning, Deep Learning.*

1. INTRODUCTION

1.1 IMAGE SEGMENTATION:

In digital image processing and computer vision, **image segmentation** is the process of partitioning a digital image into multiple **image segments**, also known as **image regions** or **image objects** (sets of pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in pixels with the same label share certain characteristics. It is widely used in computer vision tasks such as image annotation, activity recognition, face detection, face recognition, video object co-segmentation. It is also used in tracking objects, for example tracking a ball during a football match, tracking movement of a cricket bat, or tracking a person in a video. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different color respect to the same characteristic(s). When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D constructions with the help of interpolation algorithms like marching cubes. In more traditional ML-based approaches, computer vision techniques are used to look at various features of an image, such as the color histogram or edges, to identify groups of pixels that belong to an object. These features are then fed into a regression model that predicts the location of

the object along with its label and a mask is applied to each object in an image.
On the other hand, deep learning-based approaches employ convolutional neural networks (CNNs) to perform end-to-end, unsupervised image segmentation, in which features don't need to be defined and extracted separately.

1.2. GOALS AND OBJECTIVES:

The main objective of this project is to perform instance of a segmentation to analyze and segment the objects in images ineffective way and with more performance to all available deep learning algorithms. To achieve this, we use Mask R-CNN architecture and some python libraries like pillow, scikit-image which is preferred for object detection tasks. Some of the Road traffic applications where we use this image segmentation mechanism is self-driving cars, vehicle detection, speed detection

1.3. IMPORTANCE OF IMAGE SEGMENTATION:

Another important subject within computer vision is image segmentation. It is the process of dividing an image into different regions based on the characteristics of pixels to identify objects or boundaries to simplify an image and more efficiently analyze it. Segmentation impacts a number of domains, from the filmmaking industry to the field of medicine. For instance, the software behind green screens implements image segmentation to crop out the foreground and place it on a background for scenes that cannot be shot or would be dangerous to shoot in real life. Image segmentation is also used to track objects in a sequence of images and to classify terrains, like petroleum reserves, in satellite images. Some medical applications of segmentation include the identification of injured muscle, the measurement of bone and tissue, and the detection of suspicious structures aid radiologists (Computer Aided Diagnosis, or CAD).But there are important differences image recognition only outputs a class label for an identified object, and image segmentation .

2. Process in Mask R-CNN Architecture:

Image is run through the CNN to generate the feature maps. Region Proposal Network (RPN) uses a CNN to generate the multiple Region of Interest(ROI) using a lightweight binary classifier. It does this using 9 anchors boxes over the image. The classifier returns object/no-object scores.

Fig: Generating Anchor Boxes

Non Max suppression is applied to Anchors with high abjectness score dimension. Warped features are then fed into fully connected layers to make classification using soft max and boundary box prediction is further refined using the regression model**.** Warped features are also fed into Mask classifier, which consists of two CNN's to output a binary mask for each ROI.

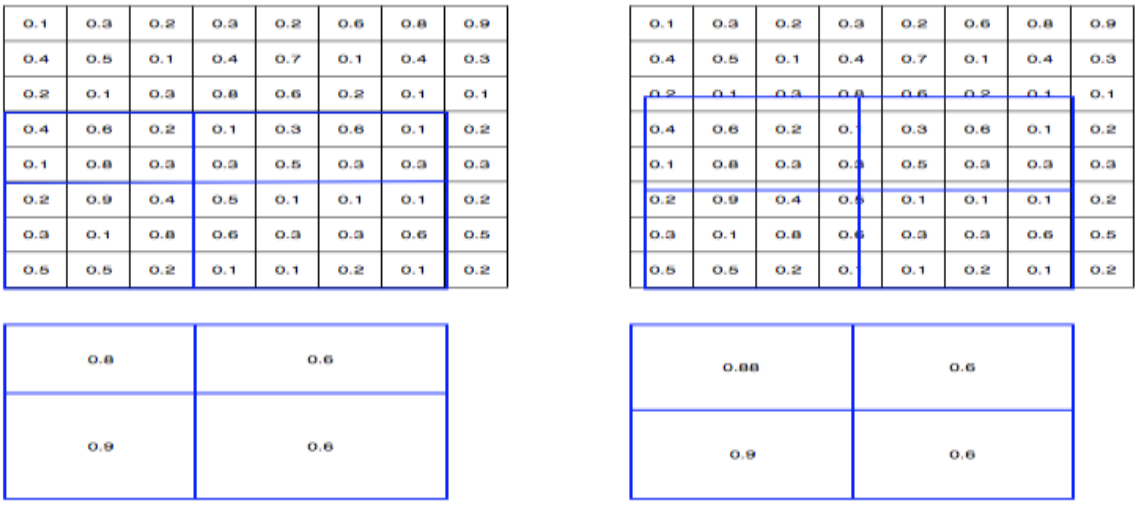


Fig: ROI Alignment

Mask Classifier allows the network to generate masks for every class without competition among classes .neural networks to classify plant diseases.

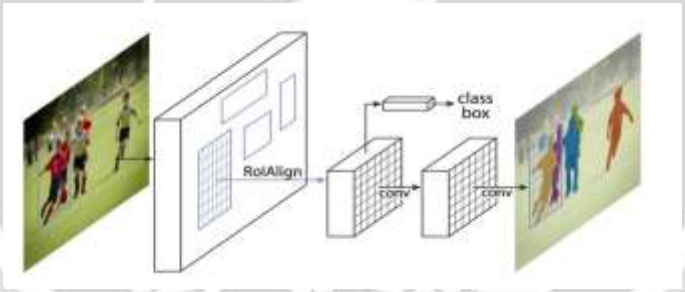


Fig: Mask R-CNN Architecture

Mask R-CNN uses anchor boxes to detect multiple objects, objects of different scales, and overlapping objects in an image. This improves the speed and efficiency for object detection. Anchor *boxes* are a set of predefined bounding boxes of a certain height and width. These boxes are defined to capture the scale and aspect ratio of specific object classes you want to detect.

ROI Align

Another major contribution of Mask R-CNN is the refinement of the ROI pooling. In ROI, the warping is digitalized (top left diagram below): the cell boundaries of the target feature map are forced to realign with the boundary of the input feature maps.

Therefore, each target cells may not be in the same size (bottom left diagram). Mask R-CNN uses and make every target cell to have the same size (bottom right). It also applies interpolation to calculate the feature map values within the cell better. For example, by applying interpolation, the maximum feature value on the top left is changed from 0.8 to 0.88 now.

3. Image segmentation loss functions

Semantic segmentation models usually use a simple cross- categorical entropy loss function during training. However, if you are interested in getting the granular information of an image, then you have to revert to slightly more advanced loss functions. This loss is an improvement to the standard cross-entropy criterion. This is done by changing its shape such that the loss assigned to well-classified examples is down-weighted. Ultimately, this ensures that there is no class imbalance. In this loss function, the cross-entropy loss is scaled with the scaling factors

decaying at zero as the confidence in the correct classes increases. The scaling factor automatically down weights the contribution of easy examples at training time and focuses on the hard ones.

Dice loss

This loss is obtained by calculating smooth dice coefficient function. This loss is

$$\mathrm{FL}(p_t) = -(1-p_t)^{\gamma}\log(p_t).$$ the most commonly used loss is segmentation problems.

Intersection over Union (IoU)-balanced Loss

The IoU-balanced classification loss aims at increasing the gradient of samples with high IoU and decreasing the gradient of samples with low IoU. In this way, the localization accuracy of machine learning model

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$ is increased.

$$IoU = TP/(TP + FP + FN)$$

Boundary loss

One variant of the boundary loss is applied to tasks with highly unbalanced segmentations. This loss's form is that of a distance metric on space contours and not regions. In this manner,
it tackles the problem posed by regional losses for highly imbalanced segmentation tasks.

$$\mathrm{Dist}(\partial G, \partial S) = \int_{\partial G} \|y_{\partial S}(p) - p\|^2 dp$$

Weighted cross-entropy

In one variant of cross-entropy, all positive examples are weighted by a certain coefficient. It is used in scenarios that involve class imbalance.

$$\mathrm{WCE}(p, \hat{p}) = -(\beta p \log(\hat{p}) + (1-p)\log(1-\hat{p}))$$

SoftMax loss

This loss performs direct optimization of the mean intersection-over-union loss in neural networks based on the convex Lovasz extension of sub-modular losses.

$$\text{loss}(f) = \frac{1}{|C|} \sum_{c \in C} \overline{\Delta_{J_c}}(m(c))$$

Other losses worth mentioning are:

TopK loss whose aim is to ensure that networks concentrate on hard samples during the training process. Distance penalized CE loss that directs the network to boundary regions that are hard to segment. Sensitivity-

Specificity (SS) loss that computes the weighted sum of the mean squared difference of specificity and sensitivity. Hausdorff distance(HD) loss that estimated the Hausdorff distance from a convolutional neural network.These are just a couple of loss functions used in image segmentation.

4. Image segmentation datasets

If you are still here, chances are that you might be asking yourself where you can get some datasets to get started.

Common Objects in COntext—Coco Dataset

COCO is a large-scale object detection, segmentation, and captioning dataset. The dataset contains 91 classes. It has 250,000 people with key points. Its download size is 37.57 GiB. It contains 80 object categories. It is available under the Apache 2.0 License and can be downloaded from official website.

PASCAL Visual Object Classes (PASCAL VOC)

PASCAL has 9963 images with 20 different classes. The training/validation set is a 2GB tar file. The dataset can be downloaded from the official website

The Cityscapes Dataset

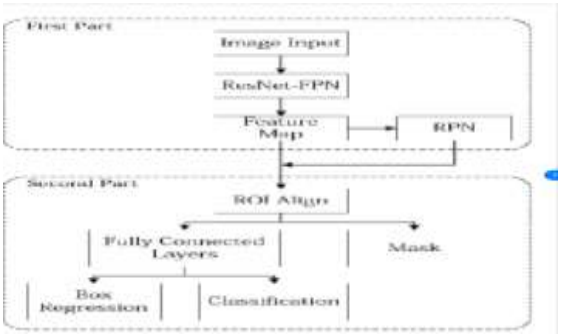
This dataset contains images of city scenes. It can be used to evaluate the performance of vision algorithms in urban scenarios. The dataset can be downloaded from official website

The Cambridge-driving Labeled Video Database—CamVid

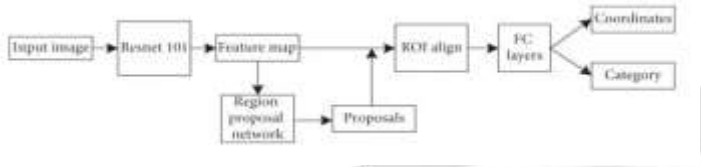
This is a motion-based segmentation and recognition dataset. It contains 32 semantic classes. This link contains further explanations and download links to the dataset.

5. System Architecture:

5.1 MASK R CNN ARCHITECTURE



5.2 Resenet-101 Architecture



.

Feature Extraction:
We utilize the ResNet 101 architecture to extract features from the input image. As a result, we get feature maps which are transmitted to Region Proposed Network.Faster C-NN using RPN & ROI Pooling RPN (Regional Proposal Network),ROI (Region of Interest).

6. RESULTS:

The image segmentation model is created through a neural network that takes in a given Ground Truth. The ground truth is a correctly labeled image that tells the neural network what the expected output is.

INPUT:

OUTPUT:

Grab Cut is an algorithm used to segment foreground objects from the background. Grab Cut worked fairly well but required that we manually supply *where* in the input image the object was so that Grab Cut could apply its segmentation magic.

Mask R-CNN, on the other hand, can *automatically* predict both the **bounding box** and the **pixel-wise segmentation mask** of each object in an input image. The downside is that masks produced by Mask R-CNN aren't always "clean" — there is typically a bit of background that "bleeds" into the foreground segmentation.

7.CONCLUSION:

GrabCut is an algorithm used to segment foreground objects from the background. Grab Cut worked fairly well but required that we manually supply *where* in the input image the object was so that GrabCut could apply its segmentation magic.

Mask R-CNN, on the other hand, can *automatically* predict both the **bounding box** and the **pixel-wise segmentation mask** of each object in an input image. The downside is that masks produced by Mask R-CNN aren't always "clean" — there is typically a bit of background that "bleeds" into the foreground segmentation.

8.Future Scope

The future enhancements of this project includes Semantic maps play a key role in tasks such as navigation of mobile robots. However, the visual SLAM algorithm based on multi-objective geometry does not make full use of the rich semantic information in space. The map point information retained in the map is just a spatial geometric point without semantics. Since the algorithm based on convolutional neural network has achieved breakthroughs in the field of target detection, the target segmentation algorithm MASK-RCNN is combined with the SLAM algorithm to construct the semantic map. However, the MASK-RCNN algorithm easily treats part of the background in the image as foreground, which results in inaccuracy of target segmentation. Moreover, Grubcut segmentation algorithm is time-consuming, but it's easy to take foreground as background, which leads to the excessive edge segmentation. Based on these, our paper proposes a novel algorithm which combines MASK-RCNN and Grubcut segmentation. By comparing the experimental results of MASK-Rcnn, Grubcut and the improved algorithm on the data set, it is obvious that the improved algorithm has the best segmentation effect and the accuracy of image target segmentation is significantly improved. These phenomenons demonstrate the effectiveness our proposed algorithm.

9.References:

[1] J. Dai, K. He, Y. Li, S. Ren and J. Sun, "Instance-sensitive fully convolutional networks", *ECCV*, 2016.

[2] J. Dai, Y. Li, K. He and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks", *NIPS*, 2016

[3] R. Girshick, "Fast R-CNN", *ICCV*, 2015.

[4] Sunkara, J. K.. Santhosh, M.. Cherukuri, S. B.. Krishna, L. G.. (2017). *Object tracking techniques and performance measures — A conceptual survey*. IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, 2017, pp. 2297-2305, doi: 10.1109/ICPCSI.2017.8392127.

[5] A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet classification with deep convolutional neural networks", *NIPS*, 2012.