

## Image Segmentation with Machine Learning

A.Praneetha<sup>a</sup>, L.Mounika<sup>b</sup>, M. Prudhvi Raj<sup>c</sup>, K. Keerthi<sup>d</sup>, N.Mahesh<sup>e</sup>

<sup>a</sup>Assistant Professor, Lakireddy Bali Reddy College of Engineering,

<sup>b,c,d,e</sup>-Students Lakireddy Bali Reddy College of Engineering,

**Article History:** Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 23 May 2021

**Abstract:** As a human we have a powerful brain which is trained to identify and classify things what our eyes perceive in fast and efficient manner. Because our brain is trained in a way to analyze everything at particular level. This helps us to easily find an Apple from a bunch of Oranges. In artificial intelligence, computer vision is the field of training computers to explain to recognize and understand the visual world that humans follow. Image segmentation can provide more detailed information about the shape of the image, which is an extension of the concept of object detection. Through image segmentation, the model can easily detect objects, humans and things. Image segmentation is an important concept in computer vision and image processing. Its application areas include augmented reality, robot perception, medical image analysis for disease detection, video surveillance cameras, board defect detection, and image compression

**Keywords:**

### 1. Introduction

#### Machine Learning:

Machine Learning has proved itself as a prominent field of study by solving the complex problems over the last decade [1]. ML has many applications in many fields. ML algorithms based on trial-and-error method which is quite opposite to the conventional algorithm which follows if-else approach [1]. ML algorithms uses trial and error method. Deep learning is a one of the parts of Machine learning which contains modules and libraries to implement Deep learning models. The success of deep learning models in computer vision and image processing applications, there has been many more ways to work on deep learning models that aimed at developing different image segmentation approaches in future [2].

#### Image Segmentation:

The study aims in the image segmentation is to segment the image into meaningful regions with respect to each instance of an object by partitioning. The partitioning is based on measurements taken from an image and might be colored image, gray level, depth, motion or texture [3]. Identifying the objects in an image for measurements of objects in the image such as size of the object and shape of the object [4]. Identifying the objects in videos for object- based video compression to segment the objects in motion capture. Image Segmentation is using to develop the self-driving cars which can detects the objects, humans, things without any human judgement and also using in video surveillance, circuit board defect detection, medical imaging and many more [3, 4].

### 2. Background Work

Image Segmentation is an advancement to the two techniques.

#### 1. Image localization:

For a single object present in an image, image localization technique is used to draw a bounding box around that object as shown in figure (1).

#### 2. Object detection:

When there is more than one object in an image, object detection provides labels along with the bounding boxes and also, we can predict the class to which each object belongs as shown in figure (1).

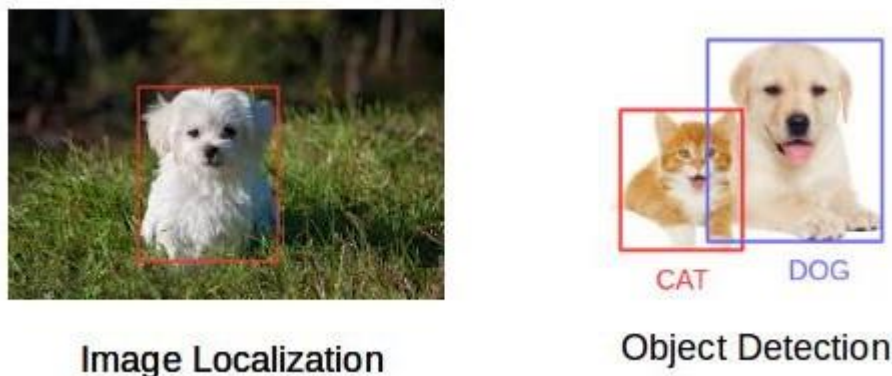


Figure (1) Represents both Image localization and Object Detection [9].

Image localization can help us identify the location of a single object in an image [4]. If there are multiple objects in the image, we rely on a concept called object detection. In object detection, we can predict the location of the image and the category of each object, such as the name of the object. Image segmentation leads to more detailed information about the shape of the image, which is an extension of the concept of object detection [4]. Therefore, we segment the image that is, divide the image into areas of different regions, which helps to distinguish one object from another at a better level. Object detection builds a bounding container like a box which is similar to every object in that identifies in the image however it tells us nothing about the shape of the object. So, we get the set of bounding container coordinates along with the name of that object [9]. Image segmentation, on the other hand, produces a pixelwise mask for each object in the image that the model recognizes [4,9]. This technique gives us a miles more granular expertise of the objects in the image.

Image Segmentation is classified as two types:

**1. Semantic Segmentation:**

Semantic Segmentation is the process of segmenting the image pixels into their respective classes for example, a Cat is associated with black colour [4]; hence all the pixels related to the Cat are coloured black. Many objects belonging to the same class are treated as a single entity and are only represented by the same colour [9,11].

**2. Instance Segmentation:**

Instance segmentation is usually coming into picture when dealing with multiple objects in the image [4]. The main difference to the detected object is masked with a colour so all the pixels associated with the image are given the same colour and it focuses on the number of instances in the image every instance is different to another instance [10]. So, multiple objects of the same class are treated as distinct entities and they are represented with different colours [10,14].

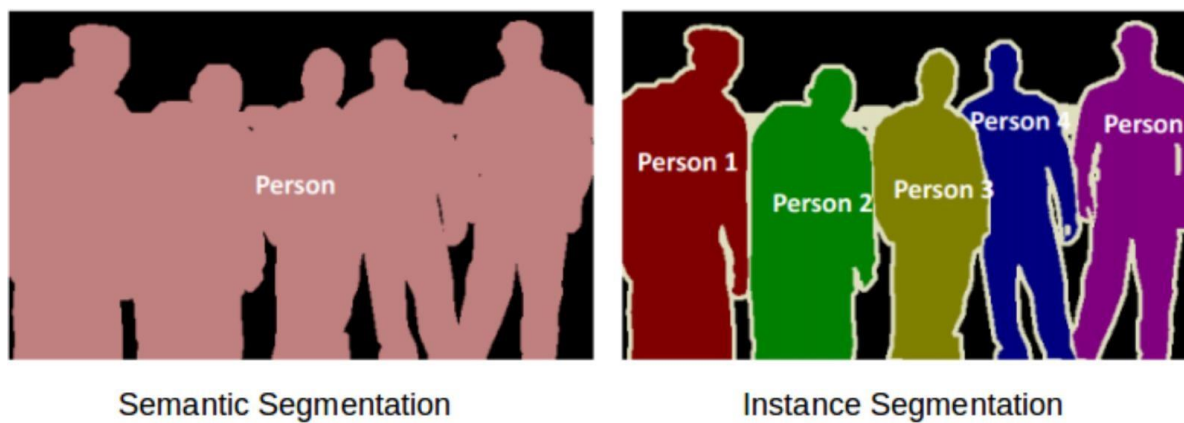


Figure (2). Represents both Semantic and Instance Segmentation [6,9].

The figure (2) shows us the exact difference between Semantic segmentation and Instance segmentation. As shown a figure (2) which contains five people, semantic segmentation would concentrate on classifying all of them as a single case, resulting in all five people being coloured the same colour. Instance segmentation, on the other hand, would classify each of these individuals separately, resulting in a different colour for each of them.

### 3. Materials And Methods

#### Training:

The main objective of this study is to partition an image into meaningful regions by taking input images and also video streaming as an input. The dataset used in this study is pretrained dataset which identifies nearly 80 to 90 objects accurately and it has been obtained by GitHub repository provided by the Matterport [12].

#### CNN:

CNN stands for Convolutional Neural Networks. It divides the image into multiple regions and then classifies each region into different classes, taking more than a minute to predict [9]. It needs a large number of regions to predict correctly, which necessitates a long computation period.

#### R-CNN:

R-CNN stands for Region Based Convolutional Neural Networks and it is an improvement to the CNN. It generates regions using a selective search technique and extracts about 2000 regions from each image [9]. It takes 40-50 seconds to estimate. However, since each region

is passed to the CNN separately, and it uses three different models to make predictions on the input image [7,9], it takes a long time to compute.

#### Fast R-CNN:

Fast R-CNN is an improvement on R-CNN in which each image is only sent to the CNN once, feature maps are extracted, and a selective search technique is applied to these maps to produce predictions. It then combines all three R-CNN models and predicts the output just in 2 seconds [13,9]. However, since the selective search technique is slow, the computation time remains high.

#### Faster R-CNN:

Faster R-CNN is a step forward from Fast R-CNN in that it replaces the selective search technique with an area proposal network, resulting in a much faster algorithm with a prediction time of 0.2 seconds [7,8,9]. However, object proposal takes time, and since various systems operate one after the other, the performance of each system is dependent on the performance of the previous system.

### 4. Proposed Framework

#### Mask R-CNN Framework:

Instance segmentation is difficult and challenging because it necessitates the accurate identification of almost all objects in an image as well as the precise segmentation of each image instance [14]. It incorporates elements from traditional computer vision tasks such as object detection, where the aim is to identify individual objects and localise them using a bounding box, as well as semantic segmentation, where the goal is to classify each pixel into a fixed collection of categories without distinguishing between them based on the object instance.

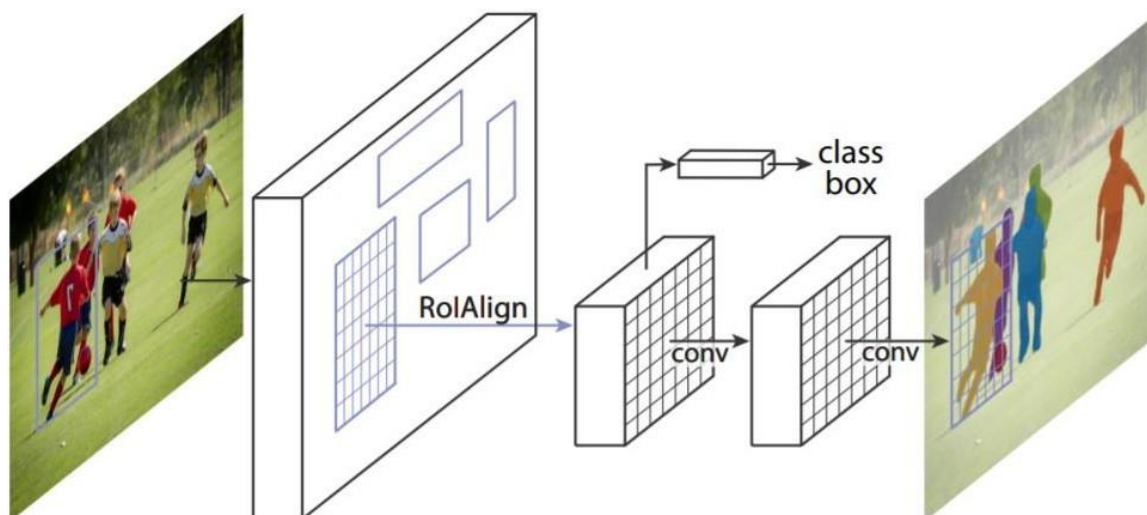


Figure (3). The Mask R-CNN framework for Instance segmentation [4,14].

Mask R-CNN is an evolution of the Faster R-CNN, and it uses a Completely Convolutional Network for pixel-

wise segmentation, similar to its predecessor Faster R-CNN. Mask R-CNN is a variant of Faster R-CNN that adds a branch for predicting segmentation masks on each Region of Interest (RoI) in addition to the classification and bounding box regression branches as shown in figure (3) [14]. The mask branch is a small Fully Convolutional Network that is added to each RoI and predicts pixel-by-pixel a segmentation mask. Given the Faster R-CNN platform, which allows for a wide variety of versatile architecture designs, that is easily trained and implemented by Mask R-CNN framework. Furthermore, the mask branch adds only a small amount of computational overhead, allowing for a fast device and rapid experimentation, and Mask R-CNN predicts the performance in milliseconds [4,7,14].

## 5. Methodology

### Proposed Work Flow:

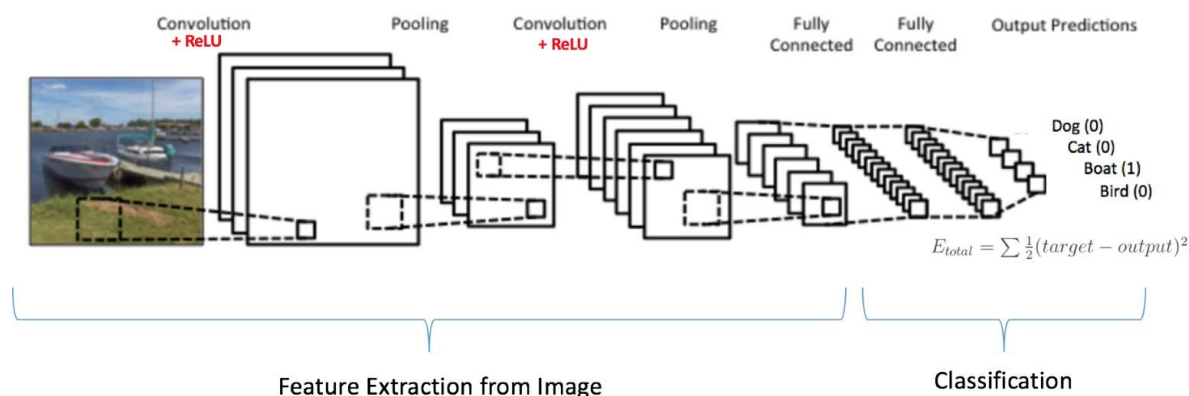


Figure (4). The CNN architecture.

A convolutional neural network (CNN) is a form of deep neural network that is most widely used to analyse visual imagery in deep learning. Based on the convolution kernels shared-weight architecture, which shifts over input features and provides translation equivariant responses. As shown in the figure (4) CNN takes the patch of an image and it undergoes different layers for feature extraction and produces the highest probability outputs that matches. Most convolutional neural networks are equivariant to translation rather than invariant [5]. They are used in image and video recognition, image detection, recommender systems, image segmentation, medical image analysis, natural language processing, and brain-computer interfaces, among other fields [3].

### Convolution layer:

It is a core layer of CNN and also a linear activation layer. It takes input patch from an image which has width, height, depth and this layer apply number of filters (n) on the input patch for example, one filter applied to the colours of an image another filter applied to the edges of an image etc. [14] finally produces the output after applying activation filtering to the input.

### Rectified Linear Unit (ReLU):

ReLU is a non-linear activation layer. If we saw a bench and it is in any direction i.e.; if it is non-linear in visibility, as a human we can easily identify it as a bench. But to achieve that nonlinearity in CNN, ReLU is used.

### Pooling:

Pooling is also a non-linear activation layer. It down samples the input based on the window size because we know that CNN is computation complex and memory complex to reduce that complexity, we are using Pooling by down sampling the output from Convolution layer. Most used techniques are max pooling and average pooling as shown in figure (5) [7,14].

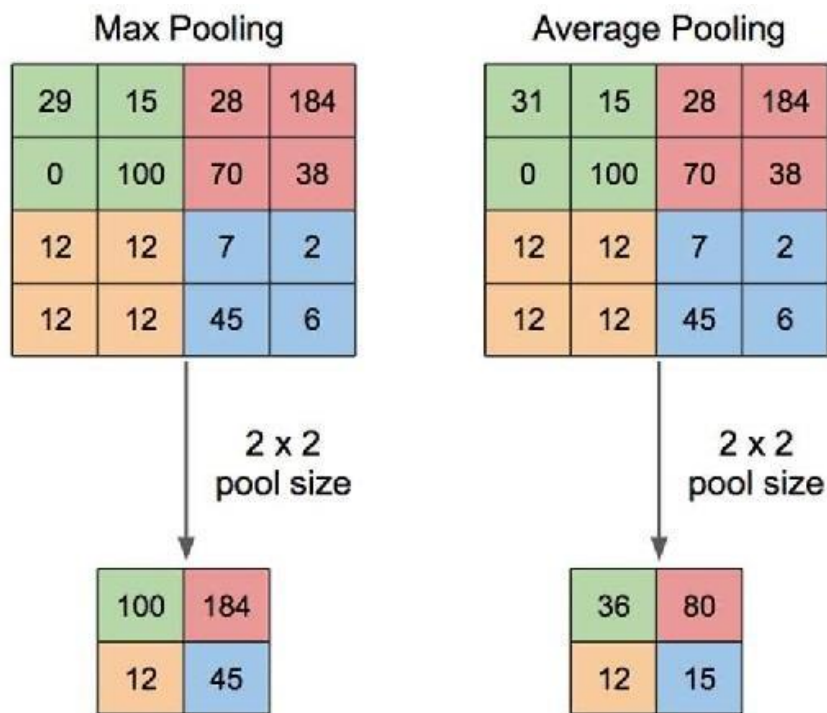


Figure (5). Max pooling and Average pooling.

**Fully Connected Layer:**

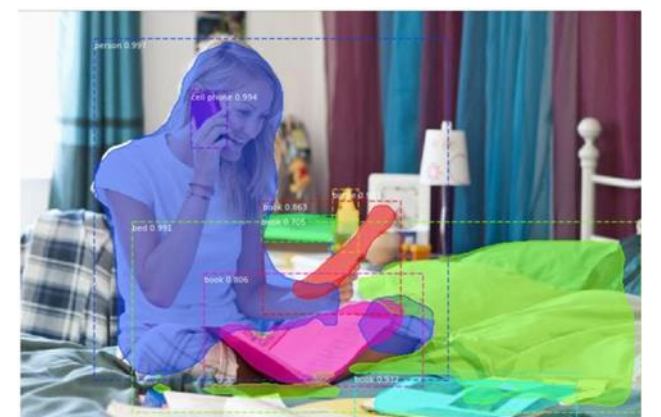
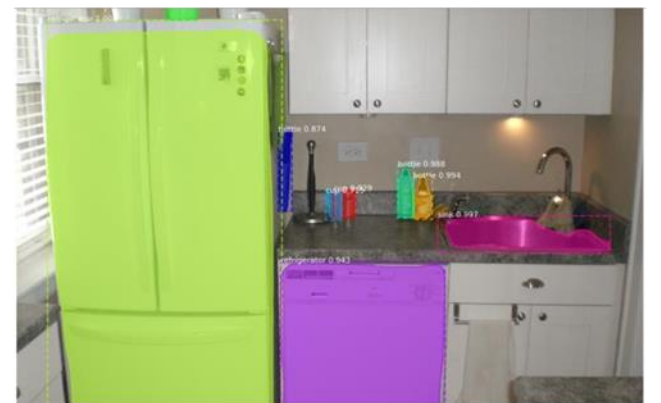
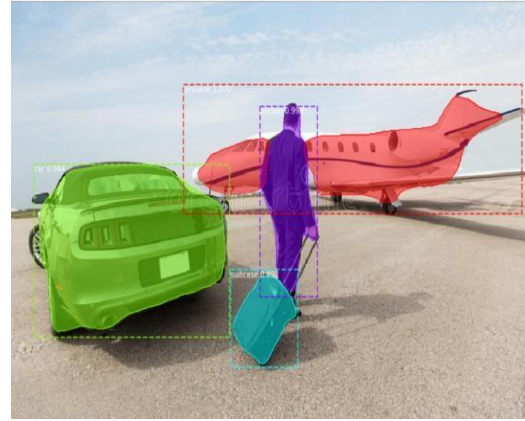
This layer is to detect the final output category and this layer every input is connected and all the coefficients are loaded and from those the top output categories are picked by using SoftMax or SVM algorithms. The output prediction with highest probability is the result.

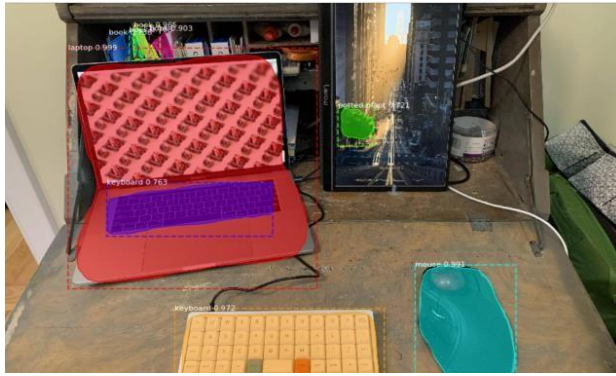
**6. Results**

The following images are the output of image segmentation on the test data images using the Mask R-CNN framework.

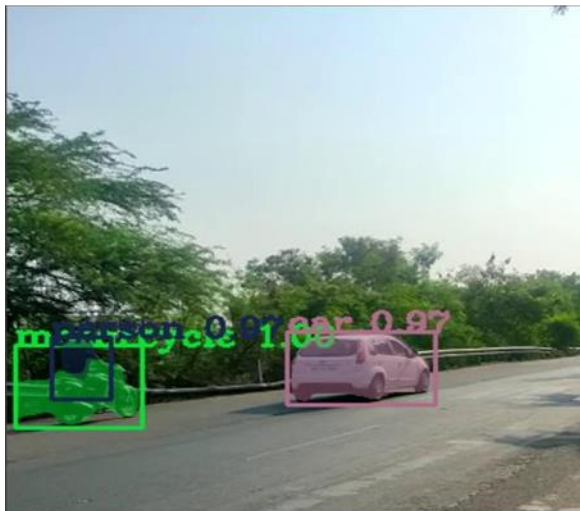








Following are the images after image segmentation on the test video using the Mask R-CNN framework.



## 7. Conclusion

Nowadays, there are many image segmentation strategies to choose from. For image classification, we primarily use semantic and instance segmentation methods in image segmentation. However, in this project instance segmentation is used for image segmentation as well as marking based on the instances of objects since instance segmentation separates every object in the image into instances, even though some objects in the same class are viewed



as separate instances. In this case, instance segmentation for image segmentation produces good results in terms of distinguishing each object. In this project, a deep neural network predicts each object label with high accuracy in the training process, using Mask R-CNN as a pretrained dataset. Mask R-CNN is a more advanced system than its predecessors, with high performance precision and a short prediction time as compared to other frameworks. The COCO dataset, which is a large image dataset designed for object identification, segmentation, and detection of person key points, is used by Mask R-CNN [4,9,12]. In comparison to other image segmentation techniques, it is the best image segmentation technique with instance segmentation based on Mask R-CNN system.

### References

1. The Hundred-Page Machine Learning Book, Author – Andriy Burkov, Publisher – Andriy Burkov.
2. Machine Learning with TensorFlow, Author–Nishant Shukla, Publisher – Manning Publications.
3. Garcia-Garcia, Alberto, et al. "A survey on deep learning techniques for image and video semantic segmentation." *Applied Soft Computing* 70 (2018): 41-65.
4. data-flair.training/blogs/image-segmentation-machine-learning.
5. J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.
6. <https://medium.com/deelvin-machine-learning/human-image-segmentation-experience-from-deelvin-5148a6cc71da>.
7. Image Segmentation Using Deep Learning: A Survey by Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos, 2020.
8. S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *TPAMI*, 2017.
9. [analyticsvidhya.com/blog/2019/04/introduction-image-segmentation-techniques-python](https://analyticsvidhya.com/blog/2019/04/introduction-image-segmentation-techniques-python).
10. S. Liu, J. Jia, S. Fidler, and R. Urtasun. SGN: Sequential grouping networks for instance segmentation. In *ICCV*, 2017.
11. L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv preprint arXiv:1706.05587, 2017.
13. [https://github.com/matterport/Mask\\_RCNN/releases](https://github.com/matterport/Mask_RCNN/releases).
14. R. Girshick. Fast R-CNN. In *ICCV*, 2015.
15. Mask R-CNN by Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick Facebook AI Research (FAIR), 2018.