A Project Report on

LIVE CAPTURING BASED IMAGE SEGMENTATION USING MASK RCNN

Submitted in partial fulfilment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

INFORMATION TECHNOLOGY

Submitted By

VELIVELLI VEDASRI (19BQ1A12H6)

SAKHAMURI AASRITHA (19BQ1A12D9)

VELPULA VENKATA DEEPTHI (19BQ1A12H7)

Under supervision of

Mr.K. JEEVAN RATNAKAR

Assistant Professor of IT

DEPARTMENT OF INFORMATION TECHNOLOGY

VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY

Permanently Affiliated to JNTU Kakinada, Approved by AICTE

Accredited by NAAC with A ‘Grade, ISO 9001:2008 Certified

NAMBUR (V), PEDAKAKANI (M), GUNTUR – 522 508

Tel no: 0863-2118036, url: www.vvitguntur.com

April 2023

VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY: NAMBUR

BONAFIDE CERTIFICATE

This is to certify that this project report **“LIVE CAPTURING BASED IMAGE SEGMENTATION USING MASK RCNN”** is the bonafide work by **“VELIVELLI VEDASRI (19BQ1A12H6), SAKHAMURI AASRITHA (19BQ1A12D9), VELPULA VENKATA DEEPTHI (19BQ1A12H7)”** who carried out the project under my guidance during the year 2023 towards partial fulfillment of the requirements of the Degree of Bachelor of Technology from Jawaharlal Nehru Technological University, Kakinada. The results embodied in this report have not been submitted to any other University for the award of any degree.

Signature of the Head of the Department
Dr. KALAVATHI ALLA
Professor, HOD
Department of Information Technology

Signature of the Supervisor
Mr.K.JEEVAN RATNAKAR
PROJECT GUIDE
Assistant Professor
Department of Information Technology

External Viva voce conducted on ____________________

Internal Examiner **External Examiner**

VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY:: NAMBUR

CERTIFICATE OF AUTHENTICATION

I Solemnly declare that this project report **“LIVE CAPTURING BASED IMAGE SEGMENTATION USING MASK R-CNN”** is the bonafide work done by purely by us, carried out under the supervision of Mr.K.Jeevan Ratnakar towards partial fulfillment of the requirements of the requirements of Degree of Bachelor of Technology in Information Technology from Jawaharlal Nehru Technological University, Kakinada during the year 2022-2023.

It is further certified that this work has not been submitted, either in part or in full, to any department of the Jawaharlal Nehru Technological University, Institution or elsewhere, or for the publication in any form.

Signature of the Student

DECLARATION

We **Velivelli Vedsri, Sakhamuri Aasritha, Velpula Venkata Deepthi** hereby declare that the project entitled "**LIVE CAPTURING BASED IMAGE SEGMENTATION USING MASK R-CNN**" done by us under the guidance of Mr.K. Jeevan Ratnakar, Assistant Professor in VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY is submitted in fulfillment of their acquirements for the award of degree in INFORMATION TECHNOLOGY. The results embodied in this project have not been submitted in any other university or college for the award of any degree or diploma.

DATE:

PLACE: NAMBUR

SIGNATURE OF THE CANDIDATES:

(VELIVELLI VEDASRI – 19BQ1A12H6)

(SAKHAMURI AASRITHA – 19BQ1A12D9)

(VELPULA VENKATA DEEPTHI– 19BQ1A12H7)

ACKNOWLEDGEMENT

Firstly, we would like to express our sincere gratitude to our beloved chairman, **Sri V. Vidya Sagar.** We would be grateful to our Principal, **Dr. Y. Mallikarjuna Reddy.** We would be grateful to our beloved Dean of Academics, **Dr. K. Giri Babu.** And people who made it easier with words of encouragement, motivation, discipline, and faith by offering different places to look forward in expanding our ideas and helped us towards the successful completion of this project work.

We take this opportunity to express our deepest gratitude and appreciation to our HOD, **Dr.A.Kalavathi, Professor,** Guide, **Mr.K.Jeevan Ratnakar, Asst Professor** and project coordinators **Mrs.M.Rajya Lakshmi, Associate Professor,** Department of Information Technology, for their motivating suggestions, insightful advice, valuable guidance, help and support in successful completion of this project and also for encouragement and advice throughout B.Tech program.

We would like to take this opportunity to express our thanks to the **teaching and non-teaching staff** of the Department of Information Technology, VVIT, for their valuable support. We would like to thank the Department of Information Technology, VVIT, for providing all facilities for completing this project. We are also grateful to all our classmates for their help, encouragement, and valuable suggestions. We would be grateful to our parents for their incessant support and cooperation. Finally, we would like to thank all those whose direct and indirect supports helped us in completing the project work on time.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS

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 "super pixels," 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. Traffic-based image segmentation is a technique used to identify and isolate vehicles, pedestrians, and other objects in traffic scenes captured by cameras. This project helps in automating the signal change irrespective of time. Though the signal is green in the way in the traffic, if the vehicles are not present, then it's of no use. So, to avoid time wasting, we came up with an idea to implement the current traffic system. After segmenting the image, if no vehicle is present in that way, it automatically changes the green signal to the next mode.

Keywords: Semantic Segmentation, Instance Segmentation, Convolutional Neural Network, Deep Learning and Image Segmentation.

CHAPTER-1

INTRODUCTION

1.1 ABOUT THE PROJECT

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 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 many 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 regard as a paradigm shift in the industry.

1.2 HARDWARE REQUIREMENTS

The hardware requirements are the requirements of a hardware device. Most hardware only has operating system requirements or compatibility. For example, a printer may be compatible with Windows XP but not compatible with newer versions of Windows like Windows 10, Linux, or Apple macOS.

If a hardware device is not compatible with your computer, it is up to the manufacturer to release drivers. Unfortunately, many manufacturers only release updated drivers to fix problems with older drivers and often do not release drivers for newer operating systems or alternative operating systems. If a hardware device doesn't have drivers for your operating system, the only solution may be to get a more up-to-date replacement device.

• Operating system	: Windows, Linux
• Processor	: Minimum intel i3
• Ram	: Minimum 4 GB
• Hard disk	: Minimum 250 GB

1.3 SOFTWARE REQUIREMENTS

The functional requirements or the overall description documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.

The appropriation of requirements and implementation constraints gives the general overview of the project regarding what the areas of strength and deficit are and how to tackle them.

- Python idle 3.7 version (or)
- Anaconda 3.7 (or)
- Jupyter Notebook (or)
- Google colab

PYTHON IDLE:

IDLE (short for Integrated Development and Learning Environment) is an integrated development environment for Python, which has been bundled with the default implementation of the language since 1.5.2b1. It is packaged as an optional part of the Python packaging with many Linux distributions. It is completely written in Python and the Tkinter GUI toolkit (wrapper functions for Tcl/Tk).

IDLE is intended to be a simple IDE and suitable for beginners, especially in an educational environment. To that end, it is cross-platform, and avoids feature clutter.

According to the included README, its main features are:

- Multi-window text editor with syntax highlighting, autocompletion, smart indent and other.
- Python shell with syntax highlighting.
- Integrated debugger with stepping, persistent breakpoints, and call stack visibility.

Author Guido van Rossum says IDLE stands for "Integrated Development and Learning Environment”, and since Van Rossum named the language Python after the British comedy group Monty Python, the name IDLE was probably also chosen partly to honor Eric Idle, one of Monty Python's founding members.

ANACONDA 3.7

Anaconda is an open-source distribution of the Python and R programming languages for data science that aims to simplify package management and deployment. Package versions in Anaconda are managed by the package management system, conda, which analyzes the current environment before executing an installation to avoid disrupting other frameworks and packages.

The Anaconda distribution comes with over 250 packages automatically installed. Over 7500 additional open-source packages can be installed from PyPI as well as the conda package and virtual environment manager. It also includes a GUI (graphical user interface), Anaconda Navigator, as a graphical alternative to the command line interface. Anaconda Navigator is included in the Anaconda distribution and allows users to launch applications and manage conda packages, environments and channels without using command-line commands. Navigator can search for packages, install them in an environment, run the packages and update them.

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda® Distribution that allows you to launch applications and manage conda packages, environments, and channels without using command line interface (CLI) commands.

JUPYTER NOTEBOOK

The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter.

Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

GOOGLE COLABORATORY

Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries which can be easily loaded in your notebook.

As a programmer, we can perform the following using Google Colab.

- Write and execute code in Python.
- Document your code that supports mathematical equations.
- Create/Upload/Share notebooks.
- Import/Save notebooks from/to Google Drive
- Import/Publish notebooks from GitHub.
- Import external datasets e.g., from Kaggle.
- Integrate PyTorch, TensorFlow, Keras, OpenCV
- Free Cloud service with free GPU

CHAPTER -2

LITERATURE SURVEY

A.Mr. V. Neethi Devan Assistant professor, Mepco Schlenk College, “Image Segmentation for object Detection using Mask R-CNN in colab”.

In the last few years researche across the global applied deep learning concept in computer vision applications. The authors addressed the problems in using two neural networks architecture LeNet and network in Network (NN) performance of the architectures to study computational efficiency by using classification and detection problems. They used multiple databases. The recent development in Deep Learning made more progress in research activities in Digital Image Processing. The authors analyzed the various pros and cons of each approach. The focus is to promote knowledge of classical computer vision techniques. Also exploring how other options of computer vision can be combined. Many hybrid methodologies are studied and proved to improve computer vision performance. To process big data applications, a large amount of space is needed in industry. Also, more space is required for the video streams from CCTV cameras and social media data, sensor data, agriculture data, medical data and data evolved from space research. The authors conducted a survey which starts from object recognition, action recognition, crowd analysis and finally violence detection in a crowd environment. The various problems in the existing methods were identified and summarized.

B.Dr. G. Chandrasekaran, “Image segmentation in Mask R-CNN”, Global Research and development Journal publication.

The authors used an approach on how to use Deep learning for all types of well-known applications such as Speech recognition, Image processing and NLP. In Deep learning, a pre-trained neural network is used to identify and remove noise from images. The processing of images using deep learning is processed for image pre-processing and image augmentation for Various applications with better results. The authors used the latest image classification techniques based on deep neural network architectures to improve the identification of highly boosted electroweak particles with respect to existing methods. Also, they introduced new methods to visualize and interpret the high-level features learned by deep neural networks that provide discrimination beyond physics derived variables, adding a new capability to understand physics and to design more powerful classification methods at the LHC. The authors proposed an analysis of tracking-by-detection approach which includes detection by YOLO and tracking by SORT algorithm. This paper has information about a custom image dataset being trained for 6 specific classes using YOLO and this model is being used in videos for tracking by SORT algorithm. Recognizing a vehicle or pedestrian in an ongoing video is helpful for traffic analysis. The goal of this paper is for analysis and knowledge of the domain.

C.Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, NasserKehtarnavaz, Demetri Terzopoulos, "Image Segmentation Using Deep Learning".

Image segmentation is a key topic in image processing and computer vision with applications such as scene understanding, medical image analysis, robotic perception, video surveillance, augmented reality, and image compression, among many others. Various algorithms for image segmentation have been developed in the literature. Recently, due to the success of deep learning models in a wide range of vision applications, there has been a substantial amount of works aimed at developing image segmentation approaches using deep learning models. In this survey, we provide a comprehensive review of the literature at the time of this writing, covering a broad spectrum of pioneering works for semantic and instance-level segmentation, including fully convolutional pixel-labeling networks, encoder-decoder architectures, multi-scale and pyramid-based approaches, recurrent networks, visual attention models, and generative models in adversarial settings. We investigate the similarity, strengths and challenges of these deep learning models, examine the most widely used datasets, report performances.

Minervini et al.2020, “Mask R-CNN for Plant Phenotyping: An Assessment on Segmentation Accuracy, Speed and Robustness”.

This paper evaluates Mask R-CNN for the task of plant segmentation, which is important for plant phenotyping studies. The authors demonstrate that Mask R-CNN performs well in terms of accuracy and speed compared to other state of the art methods.

CHAPTER-3

SYSTEM ANALYSIS

3.1 INTRODUCTION

It is a process of collecting and interpreting facts, identifying the problems, and decomposition of a system into its components. System analysis is conducted for the purpose of studying a system or its parts in order to identify its objectives. It is a problem- solving technique that improves the system and ensures that all the components of the system work efficiently to accomplish their purpose. System analysis is used in every field where something is developed. Analysis can also be a series of components that perform organic functions together, such as system engineering. System engineering is an interdisciplinary field of engineering that focuses on how complex engineering projects should be designed and managed.

3.2 EXISTING SYSTEM

Image segmentation originally started from Digital Image Processing coupled with optimization algorithms.

Thresholding:

Thresholding is one of the easiest methods of image segmentation where a threshold is set for dividing pixels into two classes. Pixels that have values greater than the threshold value are set to 1 while pixels with values lesser than the threshold value are set to 0.

The image is thus converted into a binary map, resulting in the process often termed binarization. Image thresholding is very useful in case the difference in pixel values between the two target classes is very high, and it is easy to choose an average value as the threshold.

- **Drawback:** Threshold selection is not always accurate and might miss the data.

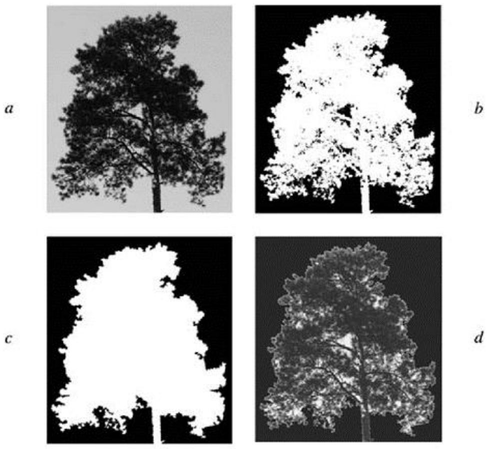


Fig1: Thresholding

Edge Segmentation:

Edge segmentation, also called edge detection, is the task of detecting edges in images. From a segmentation-based viewpoint, we can say that edge detection corresponds to classifying which pixels in an image are edge pixels and singling out those edge pixels under a separate class correspondingly.

- **Drawback:** Could not be applied on images having many edges.
- When using edge segmentation, if two or more objects are overlapped the edge computation becomes hard.

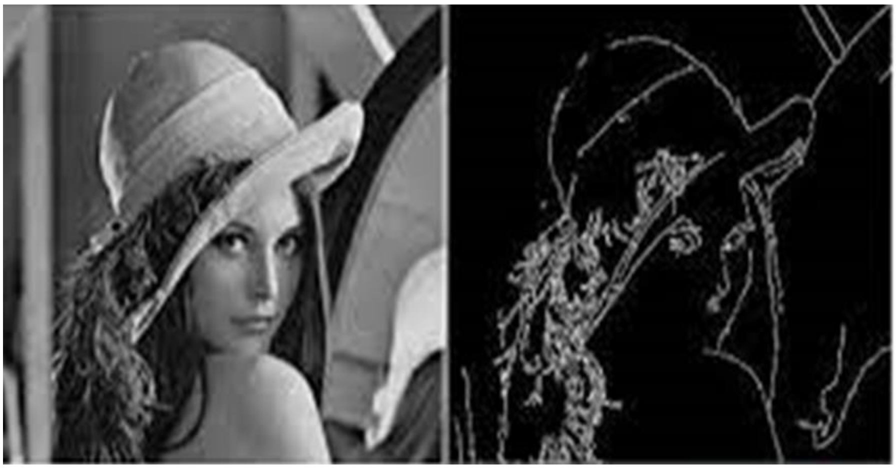


Fig2: Edge Segmentation

Clustering Based Segmentation:

Clustering algorithms perform better than their counterparts and can provide reasonably good segments in a small amount of time. Popular algorithms like the K-means clustering algorithms are unsupervised algorithms that work by clustering pixels with common attributes together as belonging to a particular segment. K-means clustering, in particular, takes all the pixels into consideration and clusters them into “k” classes. Differing from region-growing methods, clustering-based methods do not need a seed point to start segmenting from.

- **Drawback:** In clustering if the objects are overlapped the clusters cannot be divided properly.

3.3 PROPOSED SYSTEM

The target of our project is to apply the Mask R CNN algorithm for image segmentation. There are three levels of image analysis: Classification, Object detection segmentation. Classification: categorizing the entire image into class such as animals, humans, objects. Object detection: detecting objects in an image and drawing a rectangle around them. Segmentation: Identifying the parts of the image and understanding what object it belongs to. Mask R-CNN is a state-of-the-art deep neural network architecture used for image segmentation. Using Mask R-CNN, we can automatically compute pixel-wise masks for objects in the image, allowing us to segment the foreground from the background. Mask R-CNN 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.

Advantages of proposed system:

- Simplicity: Mask R-CNN is simple to train.
- Performance: Mask R-CNN outperforms all existing, single-model entries on every task.
- Efficiency: The method is very efficient and adds only a small overhead to Faster R-CNN.
- Flexibility: Mask R-CNN is easy to generalize to other tasks.

3.4 MODULES

3.4.1 DATA COLLECTION

Collect sufficient data samples and legitimate software samples. In 2019 there is a competition that contains images for image segmentation gathered from all over the world. The dataset obtained from Kaggle contains 9957 augmented images of different objects in JPEG format. This dataset is accompanied by an additional dataset containing the original pre-augmented 410 images as well as two sub-type labels and also bounding boxes for labeling (JPEG+XML) each of the object in these images. For this project, 80% of training data is used and 20% of data is used for testing. The dataset used for this project is COCO data set.

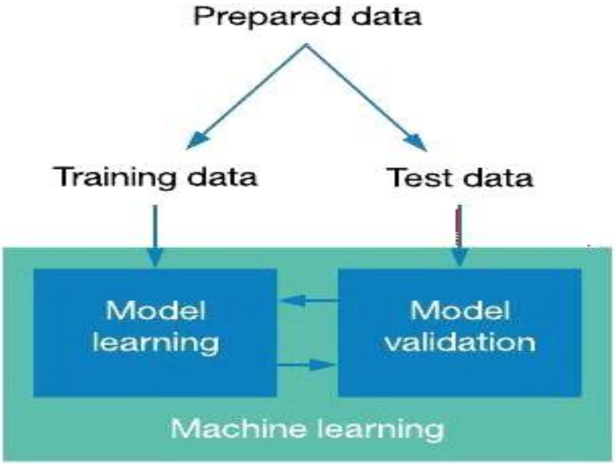


Fig 3: Collection of data

3.4.2 DATA PRE-PROCESSING

Since the dataset we have is small, we decided to augment the images by rotation, reflection about the horizontal axis and shifting both horizontally and vertically. We must ensure that computation time without losing too much accuracy. Hence, the size of the input image was reduced to 80 × 60. Scaling transformations are not applied because the correct identification of the type of image depends on the pixel. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running any analysis. Often, data preprocessing is the most important phase of a machine learning project, especially in computational biology.

3.4.3 MODEL TRAINING AND EVALUATION

To generate an output prediction, the RCNN processes the incoming image layer by layer. To configure a Mask R-CNN network for transfer learning, specify the class names and anchor boxes when you create Mask RCNN object. You can optionally specify additional network properties including the network input size and the ROI pooling sizes. The Mask R-CNN network consists of two stages. The first stage is a region proposal network (RPN), which predicts object proposal bounding boxes based on anchor boxes. The second stage is an R-CNN detector that refines these proposals, classifies them, and computes the pixel-level segmentation for these proposals. After training, the model is evaluated on a test set to see how well it performs on new data. Particularly when working with large datasets and complicated models, the training process can be long and computationally demanding. Yet, it is feasible to enhance the training procedure and attain correct output on a specific job by modifying the architecture, hyperparameters, and optimization techniques. After training, the model I evaluated on a test set to how well it performs on new data.

3.4.4 CLASSIFICATION AND PREDICTION

Following training, the model can classify unlabeled images of different objects. When an image is fed into the model, it compares the extracted features with those of the training and testing images and generates an output specifying the exact label of different objects.

3.5 ALGORITHM

CONVOLUTIONAL NEURAL NETWORK:

Computer vision is an interdisciplinary field that has been gaining huge amounts of traction in recent years (since CNN) and self-driving cars have taken center stage. Another integral part of computer vision is object detection. Object detection aids in pose estimation, vehicle detection, surveillance etc. The difference between object detection algorithms and classification algorithms is that in detection algorithms, we try to draw a bounding box around the object of interest to locate it within the image. Also, you might not necessarily draw just one bounding box in an object detection case, there could be many bounding boxes representing different objects of interest within the image and you would not know how many beforehand. To bypass the problem of selecting a huge number of regions, Ross Girshick et al. proposed a method where we use selective search to extract just 2000 regions from the image and he called them region proposals. Therefore, now, instead of trying to classify a huge number of regions, you can just work with 2000 regions. The better performance of region with convolutional neural networks with inputs such as images, speech, or audio signals sets them apart from other neural networks.

They consist of the following four layers:

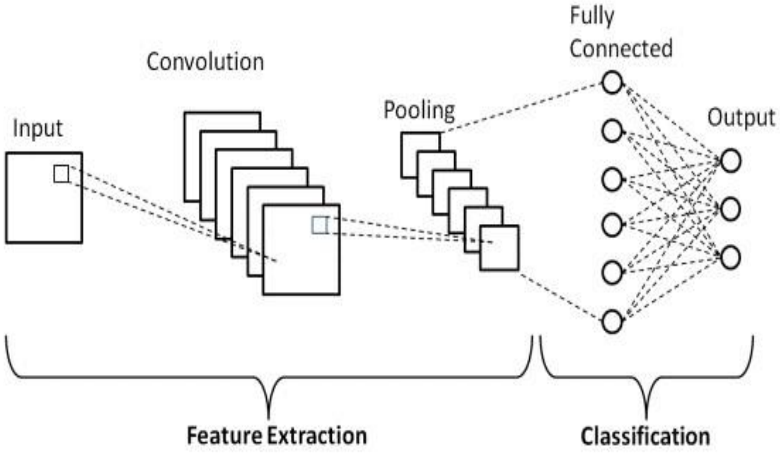


Fig 4: Convolutional Neural Network

- Convolutional layer
- ReLu layer
- Pooling layer
- Fully connected (FC) layer

A. Convolutional layer

The layer of convolution is the central component of the CNN algorithm along with it's here that the bulk in calculation takes place [10]. It needs raw data, a filter, and a feature map, among other things. A color picture made up of a three-dimensional pixel matrix will be used as the input. As a result, the data being provided will have the following parameters: length, breadth, and broad terms, which are equivalent to RGB in an image's format. Additionally, we have a feature detector, also referred to as a core or organize, which will travel throughout the receptive fields of an image and determine whether an attribute is present. Convolution describes this procedure [11].

B. ReLu Layer

Rectified Linear Unit (ReLu) is a trigger function that only starts a node if its value exceeds a predetermined threshold Otherwise, the outcome is zero. However, the connection between the data input and the variable of interest becomes linear once the input goes above a particular boundary [10]. The primary goal is to clear the convergence of every adverse value. All of the benefits stay identical, but zero is substituted for every lower one.

C. Pooling layer

The key objective is to clear the convergence of every adverse value. All of the benefits stay identical, but zero is substituted for every lower one. In contrast, the result matrix is filled by the kernel using an accumulation function applied to the numbers in the field that is open. The combination generally comes in two flavors:

- Max pooling: The one with the highest value is chosen by this filter to be sent to the result array as it travels throughout the input stream.
- Average pooling: The mean amount within the field of reception is determined as the filter moves across the input, and it is then sent to the output array.

D. Fully connected (FC) layer

Each cell in each layer below it is connected to every neuron in the completely connected layer. Based on the characteristics that were extracted from the prior levels and their various effects, this layer conducts the labeling task.

Mask R-CNN:

Mask RCNN, is a Convolutional Neural Network (CNN) and state-of-the-art in terms of image segmentation and instance segmentation. Mask R-CNN was developed on top of Faster R-CNN, a Region-Based Convolutional Neural Network. The first step to understanding how Mask R-CNN works requires an understanding of the concept of Image Segmentation. The computer vision task Image Segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as image objects). This segmentation is used to locate objects and boundaries (lines, curves, etc.).

There are 2 main types of image segmentation that fall under Mask R-CNN:

1. Semantic Segmentation
2. Instance Segmentation

Semantic segmentation:

Semantic segmentation classifies each pixel into a fixed set of categories without differentiating object instances. In other words, semantic segmentation deals with the identification/classification of similar objects as a single class from the pixel level. Semantic segmentation is otherwise known as background segmentation because it separates the subjects of the image from the background.

Instance segmentation:

Instance Segmentation, or Instance Recognition, deals with the correct detection of all objects in an image while also precisely segmenting each instance. It is, therefore, the combination of object detection, object localization, and object classification. In other words, this type of segmentation goes further to give a clear distinction between each object classified as similar instances.

Mask R-CNN was built using Faster R-CNN. While Faster R-CNN has 2 outputs for each candidate object, a class label and a bounding-box offset, Mask R-CNN is the addition of a third branch that outputs the object mask. The additional mask output is distinct from the class and box outputs, requiring the extraction of a much finer spatial layout of an object.

The key element of Mask R-CNN is the pixel-to-pixel alignment, which is the main missing piece of Fast/Faster R-CNN. Mask R-CNN adopts the same two-stage procedure with an identical first stage (which is RPN). In the second stage, in parallel to predicting the class and box offset, Mask R-CNN also outputs a binary mask for each RoI. This contrasts with most recent systems, where classification depends on mask predictions.

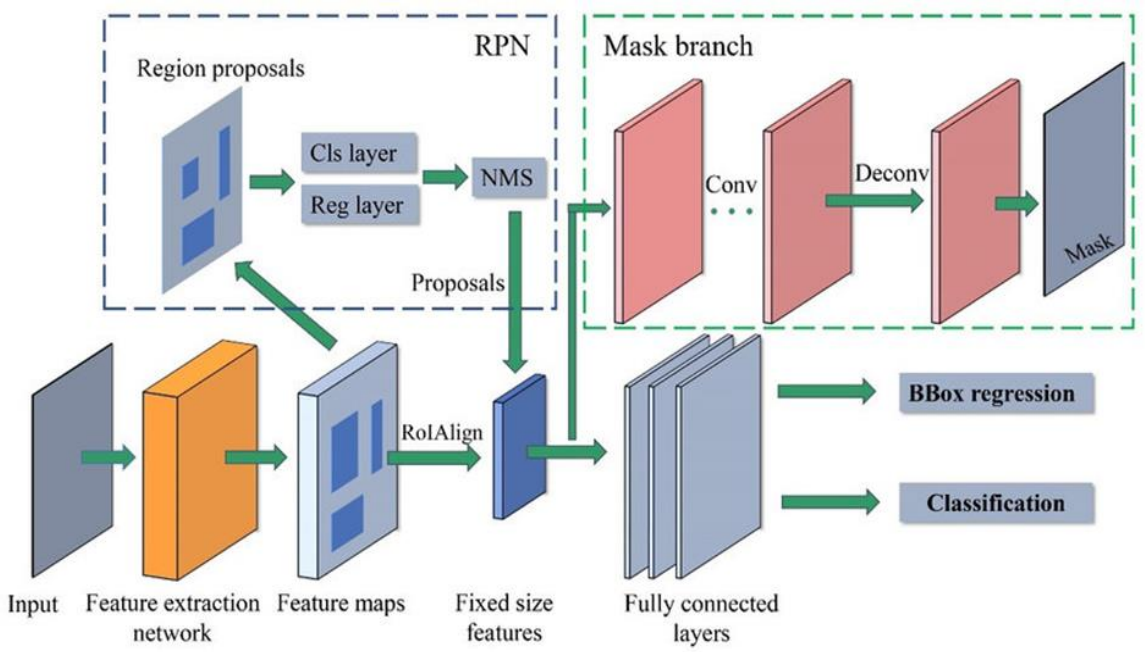


Fig 5: Mask RCNN Layers

ROI Align:

Region Of Interest Alignment is an operation of extracting a small feature map from each RoI in detection and segmentation-based tasks. It removes the quantization of RoI Pool, properly aligning the extracted features with the input. To avoid any quantization of the RoI boundaries or bins, RoIAlign uses bilinear interpolation to compute the exact values of the input features at four regularly sampled locations in each RoI bin, and the result is then aggregated using max or average.

SUPPORT VECTOR MACHINE ALGORITHM:

Support Vector Machine is a supervised learning algorithm, which is used for Classification as well as Regression problems. SVM is a popular algorithm because it has ability to handle complex datasets with high dimensional feature spaces, and it can be used for both linear and nonlinear data. However, primarily, it is used for Classification problems in Machine Learning. In the SVM algorithm, we plot each data item as a point in n-dimensional space with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the optimal hyper-plane that differentiates the two classes very well. The support vector machines in scikit-learn support both dense and sparse sample vectors as input. To use an SVM to make predictions for sparse data, it must have been fit on such data.

In SVM, the algorithm tries to find the best separating hyperplane between two classes in a dataset. The best hyperplane is the one that maximizes the distance between the closest points from each class. These closest points are called support vectors. The SVM algorithm can also be extended to handle nonlinear data by mapping the data into a higher dimensional space using a kernel function. This allows the SVM to separate the data even if it cannot be separated by a linear hyperplane in the original space. It is effective in high dimensional spaces, it can handle nonlinear data, and it has a unique solution, which makes it more robust to overfitting.

CHAPTER-4
SYSTEM DESIGN

4.1 INTRODUCTION OF SYSTEM DESIGN

System design is the process of defining architecture, modules, interfaces and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development. There is some overlap with the disciplines of systems analysis, systems architecture and systems engineering.

The basic study of system design is the understanding of component parts and their subsequent interaction with one another.

System design is the act of taking the marketing information and creating the design of the product to be manufactured. Systems design is therefore the process of defining and developing systems to satisfy specified requirements of the user.

4.2 MASK RCNN ARCHITECTURE

The Mask RCNN architecture gives an overview of the working of RCNN. The working of this is described as follows:

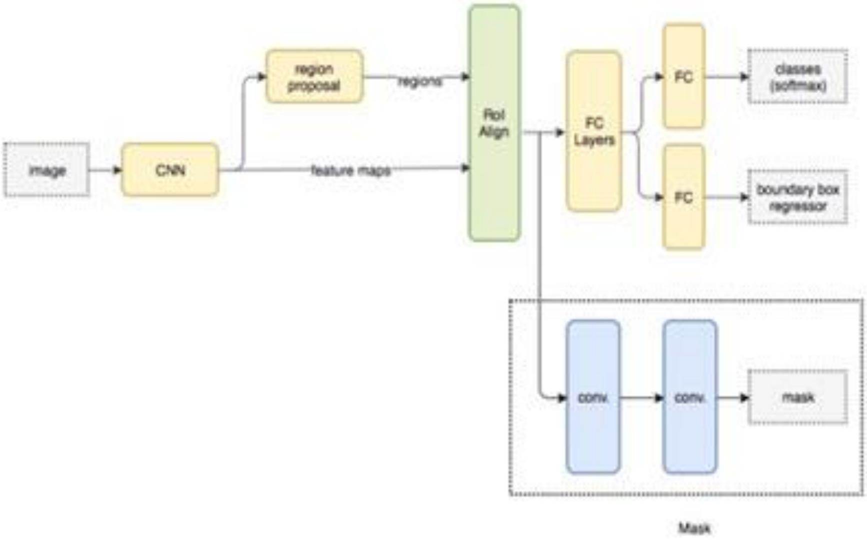


Fig 6: Mask RCNN Architecture

4.3 USECASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

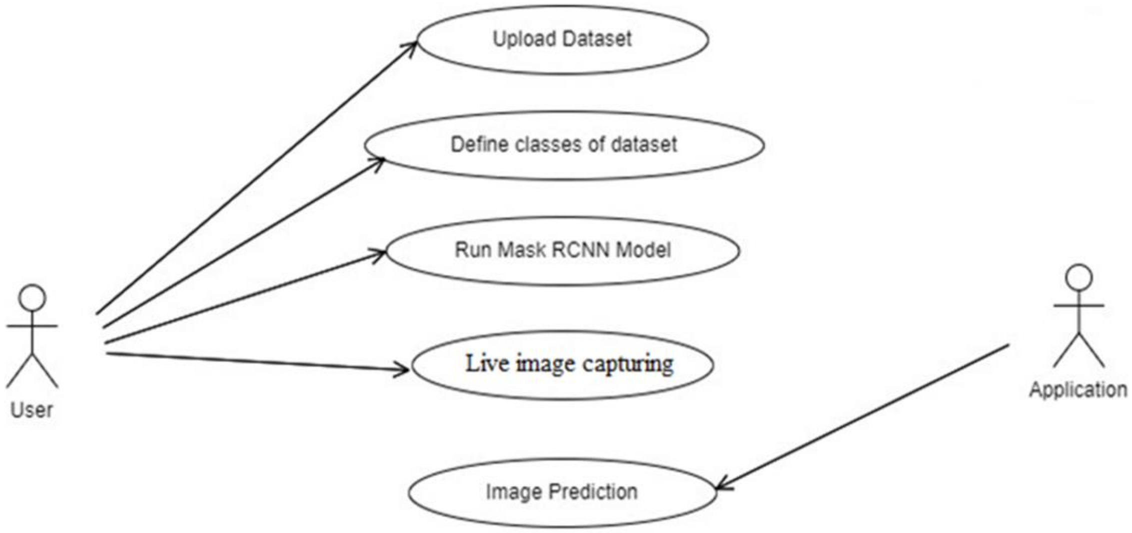


Fig 7: Use case Diagram

4.4 SEQUENCE DIAGRAM

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a message Sequence Chart. Sequence diagrams are sometimes called event diagram. These diagrams are widely used by businessmen and software developers to document and understand requirements for new and existing systems.

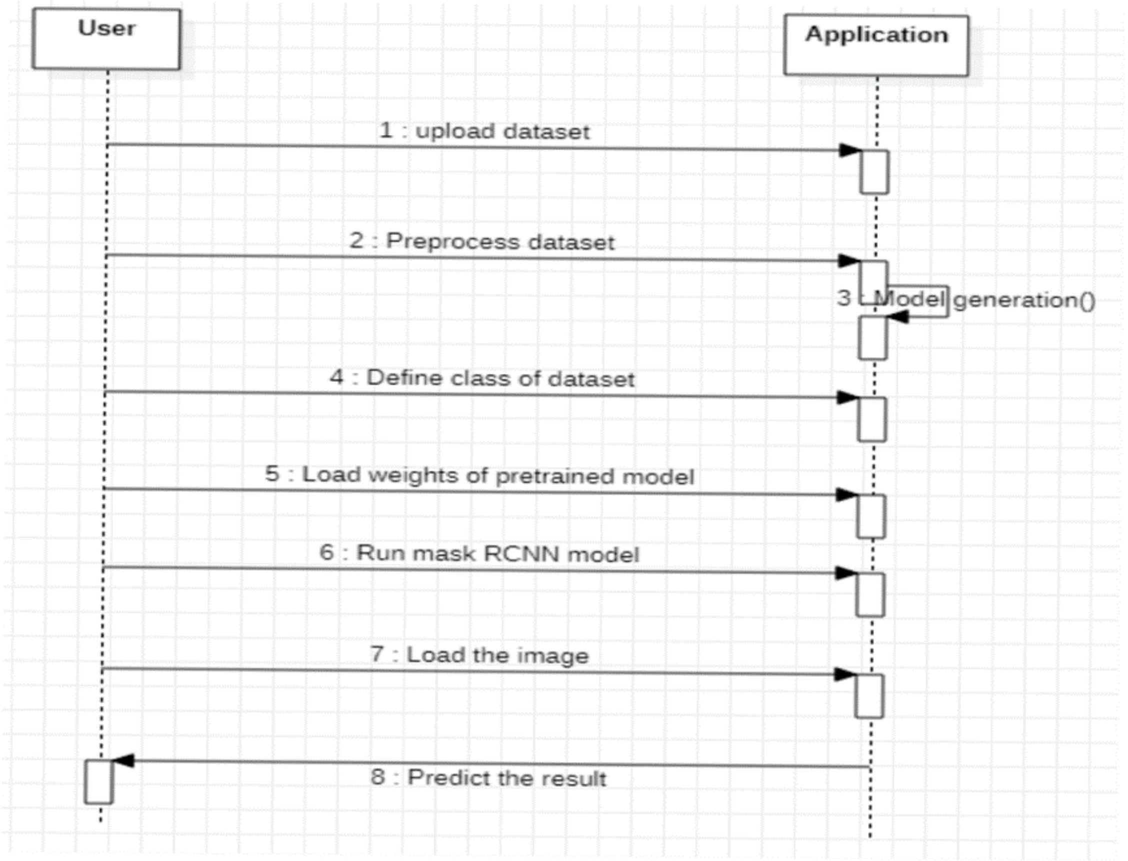


Fig 8: Sequence Diagram

4.5 CLASS DIAGRAM

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

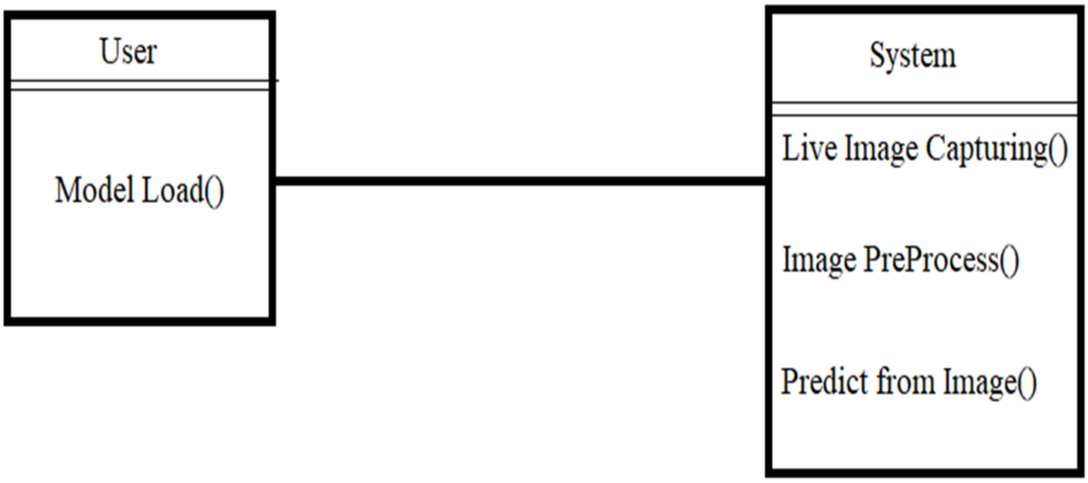


Fig 9: Class Diagram

4.6 ACTIVITY DIAGRAM

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.

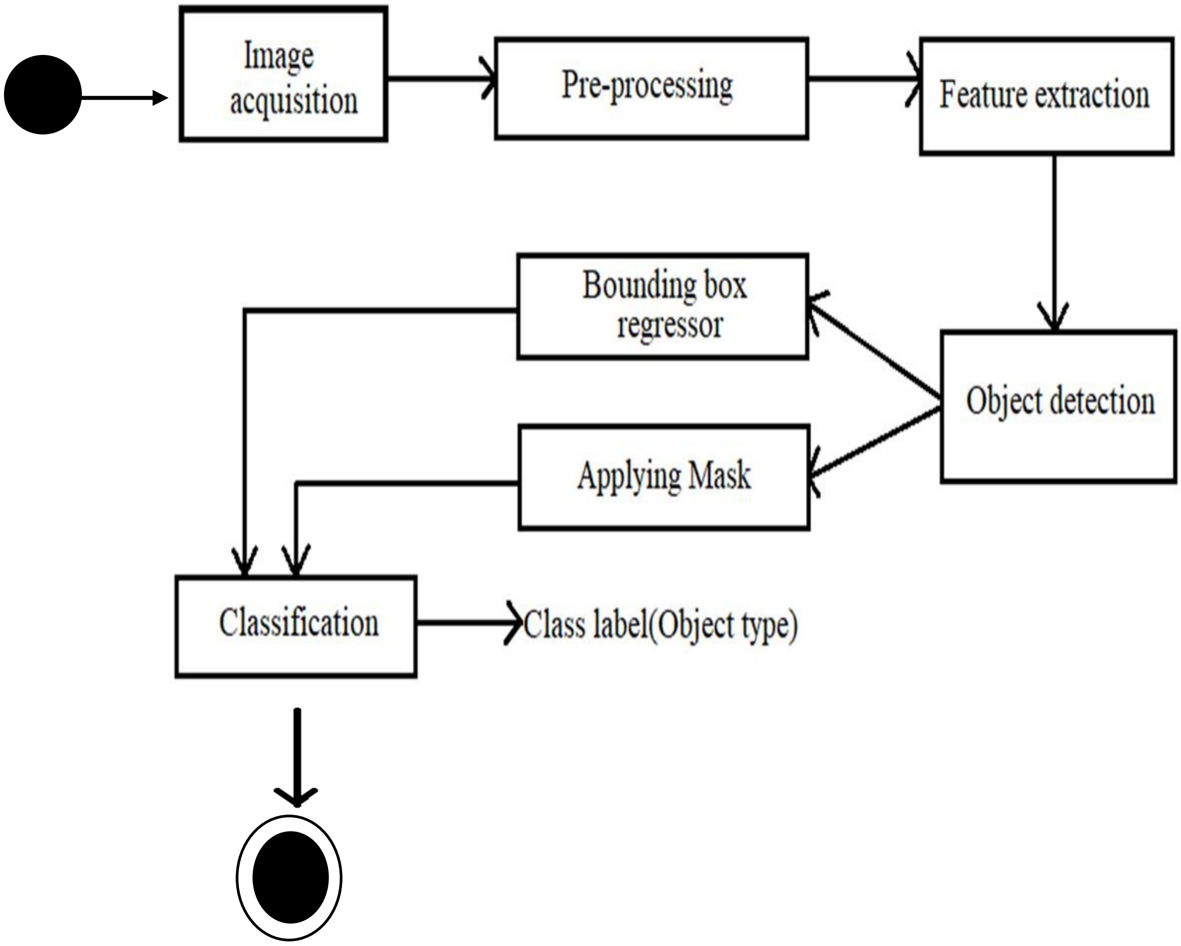


Fig 10: Activity Diagram

CHAPTER-5
SYSTEM IMPLEMENTATION

5.1 INTRODUCTION TO PYTHON

Introduction to Python: This book is about python a Web development and application development that saves you time and makes Web development a joy. Using python framework, you can build and maintain high-quality Web applications with minimal fuss. At its best, Web development is an exciting, creative act; at its worst, it can be a repetitive, frustrating nuisance. Django lets you focus on the fun stuff — the crux of your Web application — while easing the pain of the repetitive bits. In doing so, it provides high-level abstractions of common Web development patterns, shortcuts for frequent programming tasks, and clear conventions for how to solve problems. At the same time, Django tries to stay out of your way, letting you work outside the scope of the framework as needed. The goal of this book is to make you a Django expert. The focus is twofold. First, we explain, in depth, what Django does and how to build Web applications with it. Second, we discuss higher-level concepts where appropriate, answering the question "How can I apply these tools effectively in my own projects?" By reading this book, you'll learn the skills needed to develop powerful Web sites quickly, with code that is clean and easy to maintain.

Python Features

Python's features include –

- **Easy-to-learn** − Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- **Easy-to-read −** Python code is more clearly defined and visible to the eyes.
- Easy-to-maintain − Python's source code is fairly easy-to-maintain.
- **A broad standard library** − Python's bulk of the library is very portable and cross platform compatible on UNIX, Windows, and Macintosh.

- **Interactive Mode** − Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.

- Portable − Python can run on a wide variety of hardware platforms and has the same interface on all platforms.

- **Extendable** − You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.

- **Databases** − Python provides interfaces to all major commercial databases.

- **GUI Programming** − Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.

- **Scalable** − Python provides a better structure and support for large programs than shell scripting.

- Apart from the above-mentioned features, Python has a big list of good features, few are listed below

- It supports functional and structured programming methods as well as OOP.

- It can be used as a scripting language or can be compiled to byte-code for building large applications.

- It provides very high-level dynamic data types and supports dynamic type checking.

- IT supports automatic garbage collection.

- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

5.2 LIBRARIES OR PACKAGES

1. **Requests**. The most famous http library written by Kenneth Reitz. It’s a must have for every python developer.

2. **Scrapy.** If you are involved in web scraping then this is a must have library for you. After using this library, you won’t use any other.

3. **wxPython.** A Gui toolkit for python. I have primarily used it in place of tkinter. You will really love it.

4. **Pillow.** A friendly fork of PIL (Python Imaging Library). It is more user friendly than PIL and is a must have for anyone who works with images.

5. **SQLAlchemy.** A database library. Many love it and many hate it. The choice is yours.
6. **BeautifulSoup.** I know it’s slow, but this xml and html parsing library is very useful for beginners.

7. **Twisted.** The most important tool for any network application developer. It has a very beautiful api and is used by a lot of famous python developers.

8. **NumPy.** How can we leave this very important library? It provides some advanced math functionalities to python.

9. **SciPy.** When we talk about NumPy then we have to talk about SciPy. It is a library of algorithms and mathematical tools for python and has caused many scientists to switch from ruby to python.

10. **Matplotlib.** A numerical plotting library. It is very useful for any data scientist or any data analyzer.

11. **Pygame.** Which developer does not like to play games and develop them? This library will help you achieve your goal of 2d game development.

12. **Pyglet.** A 3d animation and game creation engine. This is the engine in which the famous python port of Minecraft was made.

13. **pyQT.** A GUI toolkit for python. It is my second choice after wxpython for developing GUIs for my python scripts.

14. **pyGtk.** Another python GUI library. It is the same library in which the famous Bittorrent client is created.

15. **Scapy.** A packet sniffer and analyzer for python made in python.

16. **pywin32.** A python library which provides some useful methods and classes for interacting with windows.

17. **nltk. Natural Language Toolkit** – I realize most people won’t be using this one, but it’s generic enough. It is a very useful library if you want to manipulate strings. But its capacity is beyond that. Do check it out.

18. **Nose.** A testing framework for python. It is used by millions of python developers. It is a must have if you do test driven development.

19. **SymPy.** SymPy can do algebraic evaluation, differentiation, expansion, complex numbers, etc. It is contained in a pure Python distribution.

20. **IPython.** I just can’t stress enough how useful this tool is. It is a python prompt on steroids. It has completion, history, shell capabilities, and a lot more. Make sure that you look at it. NumPy’s main object is the homogeneous multidimensional array. It is a table of elements (usually numbers), all of the same type, indexed by a tuple of positive integers. In NumPy dimensions are called axes.

The number of axes is rank.

- Offers Matlab-ish capabilities within Python
- Fast array operations
- 2D arrays, multi-D arrays, linear algebra etc. matplotlib
- High quality plotting library.

PYTHON MODULES

Python allows us to store our code in files (also called modules). This is very useful for more serious programming, where we do not want to retype a long function definition from the very beginning just to change one mistake. In doing this, we are essentially defining our own modules, just like the modules defined already in the Python library.

To support this, Python has a way to put definitions in a file and use them in a script or in an interactive instance of the interpreter. Such a file is called a module; definitions from a module can be imported into other modules or into the main module.

5.3 CODE:

Visualize_cv2.py

```
import cv2
import numpy as np
import os
import sys
from mrcnn import utils
from mrcnn import model as modellib
```

```
ROOT_DIR = os.path.abspath("../")
MODEL_DIR = os.path.join(ROOT_DIR, "logs")
sys.path.append(os.path.join(ROOT_DIR,
"samples/coco/"))
import coco
COCO_MODEL_PATH = os.path.join(ROOT_DIR,
"mask_rcnn_coco.h5")
if not os.path.exists(COCO_MODEL_PATH):
    utils.download_trained_weights(COCO_MODEL
    _PATH)

    class InferenceConfig(coco.CocoConfig):
    GPU_COUNT = 1
    IMAGES_PER_GPU = 1
    config = InferenceConfig()
    config.display()
    model = modellib.MaskRCNN(
    mode="inference",model_dir=MODEL_DIR,
    config=config
    )
    model.load_weights(COCO_MODEL_PATH,
    by_name=True)
    class_names = [
       'BG', 'person', 'bicycle', 'car', 'motorcycle',
    'airplane',
       'bus', 'train', 'truck', 'boat', 'traffic light',
       'fire hydrant', 'stop sign', 'parking meter', 'bench',
    'bird',
       'cat', 'dog', 'horse', 'sheep', 'cow', 'elephant',
    'bear',
       'zebra', 'giraffe', 'backpack', 'umbrella',
    'handbag', 'tie',
       'suitcase', 'frisbee', 'skis', 'snowboard', 'sports
    ball',
       'kite', 'baseball bat', 'baseball glove', 'skateboard',
        'surfboard', 'tennis racket', 'bottle', 'wine glass',
      'cup',
        'fork', 'knife', 'spoon', 'bowl', 'banana', 'apple',
        'sandwich', 'orange', 'broccoli', 'carrot', 'hot dog',
      'pizza',
        'donut', 'cake', 'chair', 'couch', 'potted plant', 'bed',
        'dining table', 'toilet', 'tv', 'laptop', 'mouse',
      'remote',
        'keyboard', 'cell phone', 'microwave', 'oven',
      'toaster',
        'sink', 'refrigerator', 'book', 'clock', 'vase',
      'scissors',
        'teddy bear', 'hair drier', 'toothbrush'
      ]

def random_colors(N):
   np.random.seed(1)
    colors = [tuple(255 * np.random.rand(3)) for _ in
range(N)]
    return colors
    colors = random_colors(len(class_names))
    class_dict = {
        name: color for name, color in zip(class_names,
      colors)
      }

def apply_mask(image, mask, color, alpha=0.5):
        """apply mask to image"""
        for n, c in enumerate(color):
           image[:, :, n] = np.where(
              mask == 1,
              image[:, :, n] * (1 - alpha) + alpha * c,
              image[:, :, n]
           )
        return image
```

```
def display_instances(image, boxes, masks, ids, names,
scores):
    """
        take the image and results and apply the mask,
  box, and Label
    """
    n_instances = boxes.shape[0]

    if not n_instances:
        print('NO INSTANCES TO DISPLAY')
    else:
        assert boxes.shape[0] == masks.shape[-1] ==
  ids.shape[0]

    for i in range(n_instances):
        if not np.any(boxes[i]):
            continue

        y1, x1, y2, x2 = boxes[i]
        label = names[ids[i]]
        color = class_dict[label]
        score = scores[i] if scores is not None else
  None
        caption = '{} {:.2f}'.format(label, score) if
  score else label
        mask = masks[:, :, i]

        image = apply_mask(image, mask, color)
        image = cv2.rectangle(image, (x1, y1), (x2,
  y2), color, 2)
        image = cv2.putText(
            image, caption, (x1, y1),
  cv2.FONT_HERSHEY_COMPLEX, 0.7, color, 2
        )
        return image
if __name__ == '__main__':
        """
           test everything
        """
        capture = cv2.VideoCapture(0)
      # these 2 lines can be removed if you dont have a
1080p camera.
     capture.set(cv2.CAP_PROP_FRAME_WIDTH,
     1920)
     capture.set(cv2.CAP_PROP_FRAME_HEIGHT,
     1080)
     while True:
           ret, frame = capture.read()
           results = model.detect([frame], verbose=0)
           r = results[0]
           frame = display_instances(
              frame,   r['rois'],   r['masks'],   r['class_ids'],
     class_names, r['scores']
           )
           cv2.imshow('frame', frame)
           if cv2.waitKey(1) & 0xFF == ord('q'):
           break

capture.release()
cv2.destroyAllWindows()
```

video_demo.py

```
import cv2
from visualize_cv2 import model, display_instances, class_names
import sys
args=sys.argv
if(len(args)<2):
  print("run command: python video_demo.py 0 or video file name")
```

```
sys.exit(0)
name=args[l]
if(len(args[l] == l)):
  name = int(args[l])
stream=cv2.VideoCapture(name)
while True:
  ret,frame=stream.read()
  if not ret:
    print("unable to fetch frame")
    break
  results = model.detect([frame], verbose=l)
  r=results[0]
  masked_image=display_instances(frame, r['rois'], r['masks'], r['class_ids'],
class_names, r['scores'])
  cv2.imshow("masked_image",masked_image)
  if(cv2.waitKey(l) & 0xFF == ord('q')):
    break
stream.release()
cv2.destroyWindow("masked_image
```

CHAPTER-6

TESTING AND RESULTS

6.1 SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

6.1.1 Types of Testing

1. White Box Testing
2. Black Box Testing
3. Unit testing
4. Integration Testing
5. Alpha Testing
6. Beta Testing
7. Performance Testing

White Box Testing

Testing technique is based on knowledge of the internal logic of an application's code and includes tests like coverage of code statements, branches, paths, conditions. It is performed by software developers.

Black Box Testing

Blackbox testing is testing the functionality of an application without knowing the details of its implementation including internal program structure, data structures etc. Test cases for black box testing are created based on the requirement specifications. Therefore, it is also called as specification-based testing.

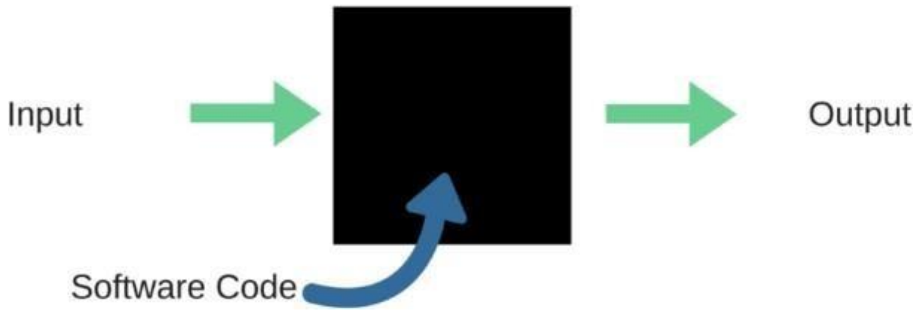


Fig 11: Black Box Testing

When applied to machine learning models, black box testing would mean testing machine learning models without knowing the internal details such as features of the machine learning model, the algorithm used to create the model etc. The challenge, however, is to verify the test outcome against the expected values that are known beforehand.

Table 1: Black box testing

Input	**Actual Output**	**Predicted Output**
[16,6,324,0,0,0,22,0,0,0,0,0,0]	0	0
[16,7,263,7,0,2,700,9,10,1153,832,9,2]	1	1

The model gives out the correct output when different inputs are given which are mentioned in Table. Therefore, the program is said to be executed as expected or correct program.

Unit Testing

Software verification and validation method in which a programmer tests if individual units of source code are fit for use. It is usually conducted by the development team.

Integration Testing

The phase in software testing in which individual software modules are combined and tested as a group. It is usually conducted by testing teams.

Integration testing is also further divided into two types.

1.Incremental testing

2.Non incremental testing

Incremental Testing

Whenever there is a clear relationship between modules, we go for incremental integration testing. Suppose we take two modules and analysis the data flow between them to see if they are working fine or not.

If these modules are working fine, then we can add one more module and test again. And we can continue with the same process to get better results.

Non-Incremental Testing

Whenever the data flow is complex and very difficult to classify a parent and a child, we will go for the non-incremental integration approach. The non-incremental method is also known as the Big Bang method.

Alpha Testing

Type of testing a software product or system conducted at the developer's site. Usually, it is performed by the end users.

Beta Testing

Final testing before releasing application for commercial purpose. It is typically done by end- users or others. Beta testing helps in minimization of product failure risks, and it provides increased quality of the product through customer validation. It is the last test before shipping a product to the customers. One of the major advantages of beta testing is direct feedback from customers.

Performance Testing

Functional testing was conducted to evaluate the compliance of a system or component with specified performance requirements. It is usually conducted by the performance engineer.

6.2 RESULTS

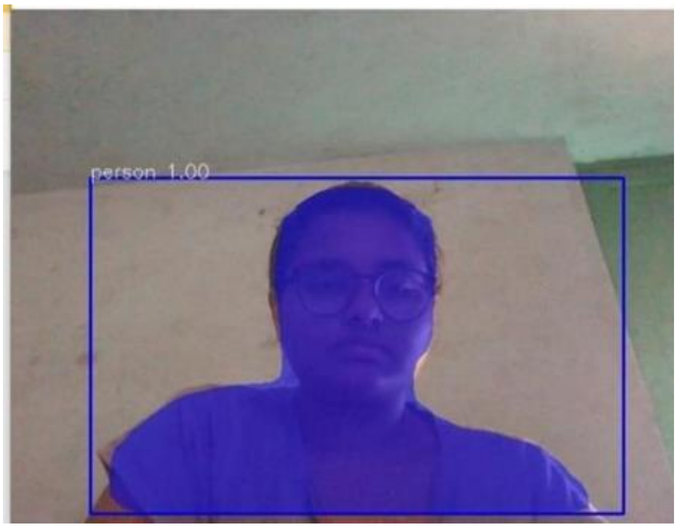


Fig 12.a): Output when there is single object.

By implementing the model, above are the result screens. In the above 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.

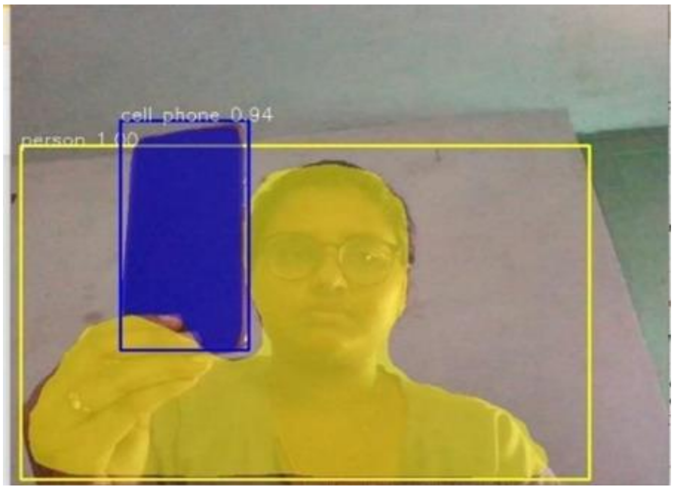


Fig 12.b): Output with multiple objects

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

In Fig 12.b the model identified two objects such as a 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 accurate results**.**

6.3 TEST CASES

A test case is a set of actions performed on a system to determine if it satisfies software requirements and functions correctly. The purpose of a test case is to determine if different features within a system are performing as expected and to confirm that the system satisfies all related standards, guidelines and customer requirements. The process of writing a test case can also help reveal errors or defects within the system.

Table 2: Test Cases

S.No	Test Case Description	Expected Output	Actual Output	Result
1.	Image containing an object	Detection of object and applying bounding box, mask.	Detection of object and applying bounding box, mask.	Pass
2.	Image containing multiple objects	Detection of all the objects and applying bounding box, mask to each and every object separately.	Detection of all the objects and applying bounding box, mask to each and every object separately.	Pass
3.	Image containing zero objects	Does not apply any mask or bounding box.	Did not apply any mask or bounding box.	Pass

Test cases are typically written by members of the quality assurance (QA) team or the testing team and can be used as step-by-step instructions for each system test. Testing begins once the development team has finished a system feature or set of features. A sequence or collection of test cases is called a test suite.

CHAPTER-7
CONCLUSION AND FUTURE SCOPE

Mask RCNN is a deep neural network aimed to solve instance segmentation problems in machine learning or computer vision. Mask R-CNN is a conceptually simple, flexible, and general framework for object instance segmentation. It can efficiently detect objects in an image while simultaneously generating a high-quality segmentation mask for each instance. It does object detection and instance segmentation and can also be extended to human pose estimation. It extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. Mask R-CNN is simple to train and adds only a small overhead to Faster R-CNN, running at 5 fps. In this project we used both real time image and pre captured images to implement image segmentation using the Mask RCNN. The desired output is obtained. In the forthcoming future, this project can be modified and implemented for auto pilot cars, cancer cell shape detection and object detection from satellite images. As taking basis of this project, all the above projects can be implemented. Due to advances in image processing and related technologies there will be millions and millions of robots in the world in a few decades' time, transforming the way the world is managed. Advances in image processing and artificial intelligence will involve spoken commands, anticipating the information requirements of governments, translating languages, recognizing and tracking people and things, diagnosing medical conditions, performing surgery, reprogramming defects in human DNA, and automatic driving all forms of transport. With increasing power and sophistication of modern computing, the concept of computation can go beyond the present limits and in future, image processing technology will advance, and the visual system of man can be replicated.

CHAPTER-8
REFERENCES

[1] **F. Meyer**, An overview of morphological segmentation. International Journal of Pattern Recognition and Artificial.

[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.

PUBLISHED ARTICLE

International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

COPY RIGHT

ELSEVIER

SSRN

IJIEMR Transactions, online available on 10th Apr 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/101

Title LIVE CAPTURING BASED IMAGE SEGMENTATION USING MASK R-CNN

Volume 12, ISSUE 04, Pages: 814-819

Paper Authors

Mr. K. Jeevan Ratnakar, V. Vedasri, S. Aasritha, V. Venkata Deepthi

International Journal for Innovative Engineering and Management Research

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

LIVE CAPTURING BASED IMAGE SEGMENTATION USING MASK R-CNN

Mr. K. Jeevan Ratnakar[1], Assistant Professor Department of Information Technology,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

V. Vedasri[2], **S. Aasritha**[3], **V. Venkata Deepthi**[4]
[2,3,4] UG Students, Department of Information Technology,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
jeevanratnakar@vvit.net[1], saiveda12335@gmail.com[2], aasrithasakhamuri2948@gmail.com[3],
venkatadeepthiv@gmail.com[4]

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 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%.

2. Classification

Generally, classification is a machine-learning 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

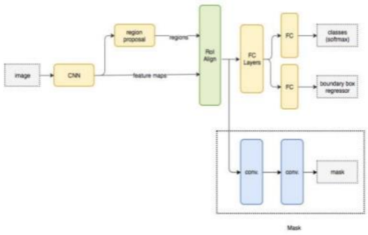


Figure.1. Architecture of Mask R-CNN

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.

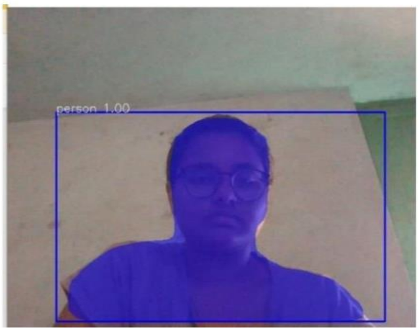


Fig.2. Output when objects are one.

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.

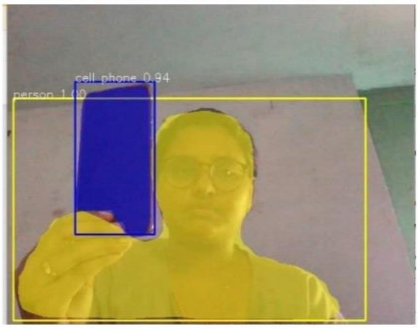


Fig.3. Output when objects are two.

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, "Region-based representations of image and video: Segmentation tools for multimedia services", IEEE Transacations on circuits and systems for video technology, vol. 9, no. 8, pp. 1147-1167, Dec. 1999.

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

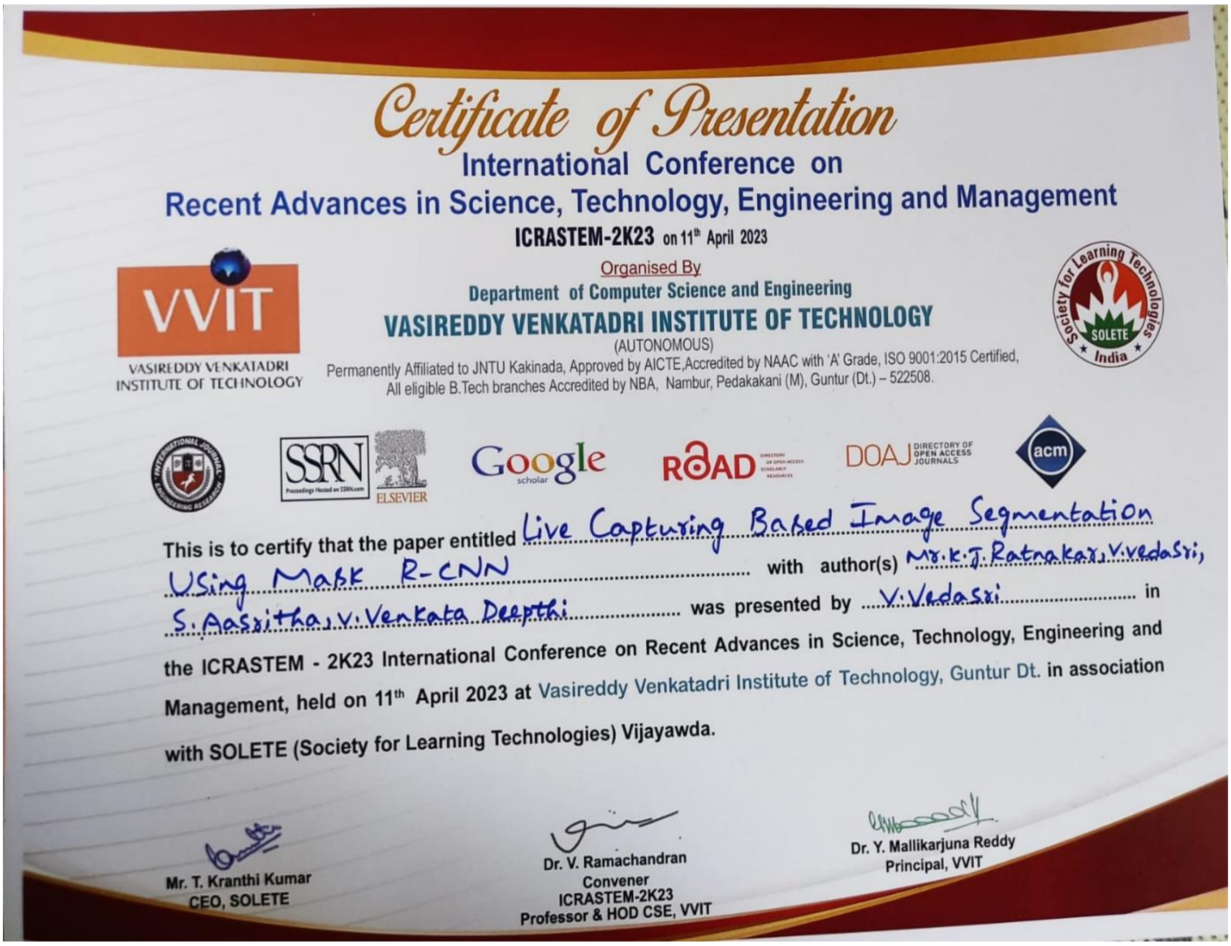
[6] J. Serra, "segmentation", tech. report EMP N-06/03/MM, pp. 25, march 2003.

[7] C. Tzafestas and P. Maragos, "Shape Connectivity", Journal of Mathematical Imaging and Vision, vol. 17, no. 2, pp. 109-129, 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.



Certificate of Presentation

International Conference on

Recent Advances in Science, Technology, Engineering and Management

ICRASTEM-2K23 on 11th April 2023

Organised By

Department of Computer Science and Engineering

VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY

(AUTONOMOUS)

Permanently Affiliated to JNTU Kakinada, Approved by AICTE,Accredited by NAAC with 'A' Grade, ISO 9001:2015 Certified,
All eligible B.Tech branches Accredited by NBA, Nambur, Pedakakani (M), Guntur (Dt.) – 522508.

VVIT
VASIREDDY VENKATADRI
INSTITUTE OF TECHNOLOGY

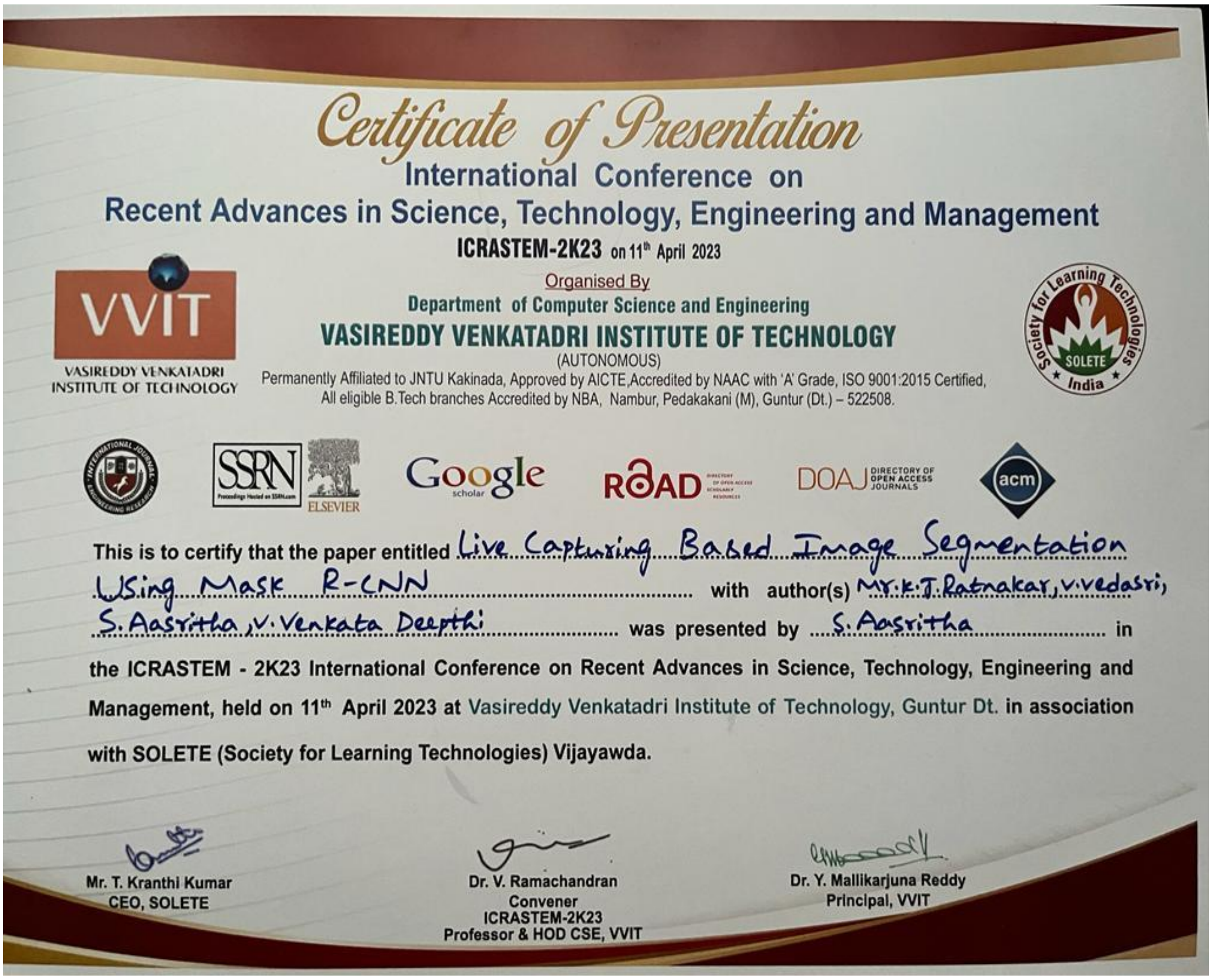
Society for Learning Technologies
SOLETE
India

This is to certify that the paper entitled Live Capturing Based Image Segmentation Using Mask R-CNN with author(s) Mr.K.J.Ratnakar, V.vedasri, S.Aasritha, V.Venkata Deepthi was presented by V.Vedasri in the ICRASTEM - 2K23 International Conference on Recent Advances in Science, Technology, Engineering and Management, held on 11th April 2023 at Vasireddy Venkatadri Institute of Technology, Guntur Dt. in association with SOLETE (Society for Learning Technologies) Vijayawda.

Mr. T. Kranthi Kumar
CEO, SOLETE

Dr. V. Ramachandran
Convener
ICRASTEM-2K23
Professor & HOD CSE, VVIT

Dr. Y. Mallikarjuna Reddy
Principal, VVIT



Certificate of Presentation

International Conference on

Recent Advances in Science, Technology, Engineering and Management

ICRASTEM-2K23 on 11th April 2023

Organised By

Department of Computer Science and Engineering

VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY

(AUTONOMOUS)

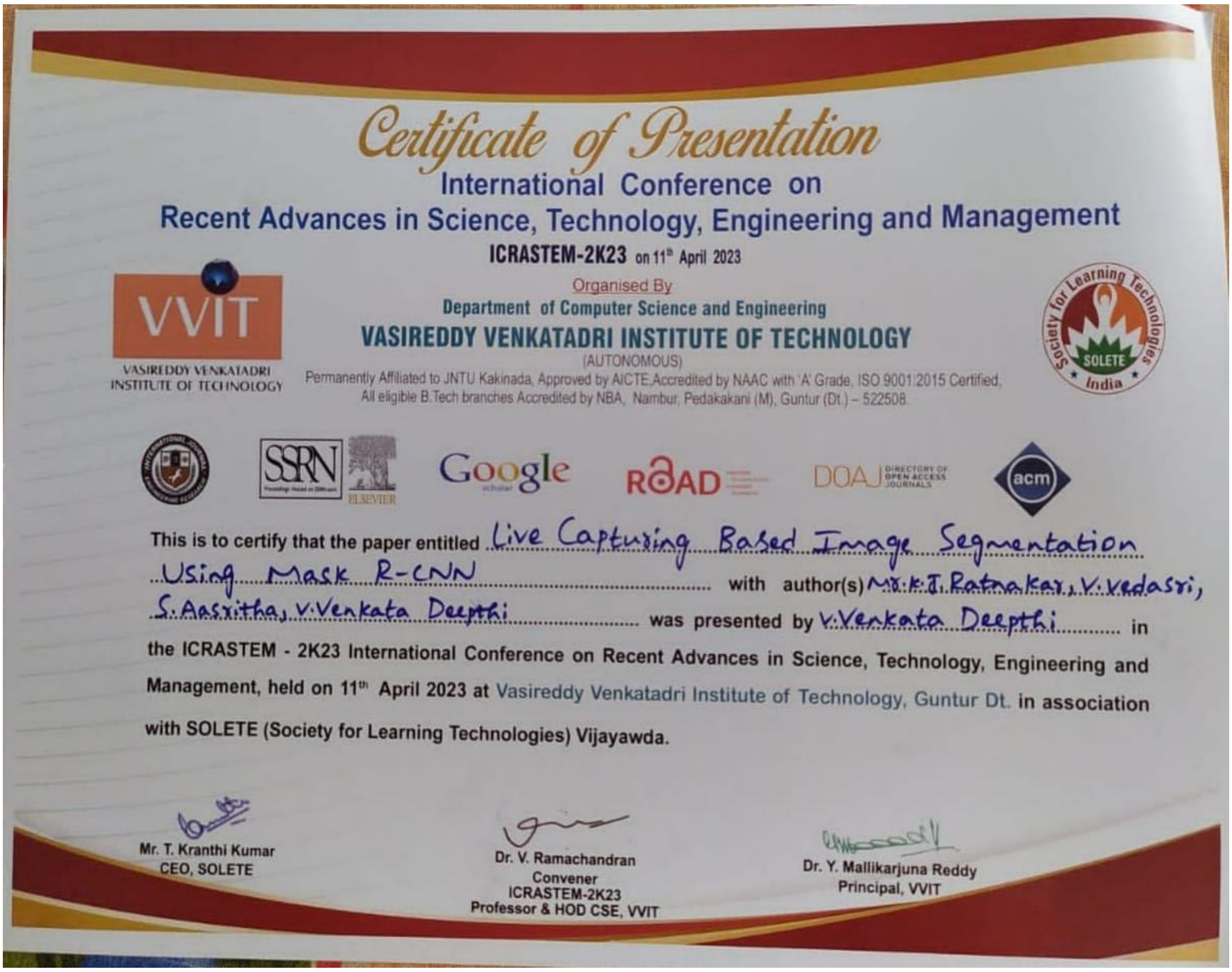
Permanently Affiliated to JNTU Kakinada, Approved by AICTE,Accredited by NAAC with 'A' Grade, ISO 9001:2015 Certified,
All eligible B.Tech branches Accredited by NBA, Nambur, Pedakakani (M), Guntur (Dt.) – 522508.

This is to certify that the paper entitled Live Capturing Based Image Segmentation Using Mask R-CNN with author(s) Mr.k.T.Ratnakar, v.vedasri, S.Aasritha, v.Venkata Deepthi was presented by S.Aasritha in the ICRASTEM - 2K23 International Conference on Recent Advances in Science, Technology, Engineering and Management, held on 11th April 2023 at Vasireddy Venkatadri Institute of Technology, Guntur Dt. in association with SOLETE (Society for Learning Technologies) Vijayawda.

Mr. T. Kranthi Kumar
CEO, SOLETE

Dr. V. Ramachandran
Convener
ICRASTEM-2K23
Professor & HOD CSE, VVIT

Dr. Y. Mallikarjuna Reddy
Principal, VVIT



Certificate of Presentation

International Conference on

Recent Advances in Science, Technology, Engineering and Management

ICRASTEM-2K23 on 11th April 2023

Organised By

Department of Computer Science and Engineering

VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY

(AUTONOMOUS)

Permanently Affiliated to JNTU Kakinada, Approved by AICTE,Accredited by NAAC with 'A' Grade, ISO 9001:2015 Certified,
All eligible B.Tech branches Accredited by NBA, Nambur, Pedakakani (M), Guntur (Dt.) – 522508.

VVIT
VASIREDDY VENKATADRI INSTITUTE OF TECHNOLOGY

Society for Learning Technologies SOLETE India

SSRN ELSEVIER Google scholar ROAD DOAJ DIRECTORY OF OPEN ACCESS JOURNALS acm

This is to certify that the paper entitled Live Capturing Based Image Segmentation Using Mask R-CNN with author(s) Mr.K.J.Ratnakar, V.Vedasri, S.Aasritha, V.Venkata Deepthi was presented by V.Venkata Deepthi in the ICRASTEM - 2K23 International Conference on Recent Advances in Science, Technology, Engineering and Management, held on 11th April 2023 at Vasireddy Venkatadri Institute of Technology, Guntur Dt. in association with SOLETE (Society for Learning Technologies) Vijayawda.

Mr. T. Kranthi Kumar
CEO, SOLETE

Dr. V. Ramachandran
Convener
ICRASTEM-2K23
Professor & HOD CSE, VVIT

Dr. Y. Mallikarjuna Reddy
Principal, VVIT