SMART SEGMENTATION

Submitted in partial fulfillment of the requirements of the degree of

T.E in Computer Engineering

By

Roll no. 05 Tushar Avhad

Roll no. 24 Prathamesh Kadam

Roll no. 44 Bhavesh Patil

Roll no. 72 Mahendra Valvi

Supervisor:

Prof. Kanchan Doke

(Department of Computer Engineering)

Bharati Vidyapeeth College of Engineering, Navi Mumbai

Table of Contents

Chapter – One

Introduction

1.1 Abstract

Smart segmentation is a very important task in computer vision, which involves breaking down an image into smaller parts that represent objects or regions of interest in a meaningful and consistent way. This task is critical in a wide range of applications, including object recognition, scene understanding, and image analysis. The main objective of Smart Segmentation is to divide an image into several segments, with each segment representing a unique object or region of interest. Lately, deep learning techniques, specifically something called convolutional neural networks or CNNs, have been really promising for this task.

This article discusses the fundamentals of Smart Segmentation and provides an overview of deep learning techniques used for this task. The article also covers the implementation of Smart Segmentation using popular deep learning frameworks like TensorFlow. The project involves training a CNN model on a large dataset of images to perform Smart segmentation, which involves labelling each pixel in the image with a class label. Overall, the results show that deep learning is a really powerful technique for image segmentation.

Keywords: Mask R-CNN (Region-Based Convolutional Neural Network), CNN (Convolutional Neural Networks), Deep Learning, Image processing, Segmentation Models, Computer Vision, FCN, Tensorflow, Keras.

1.2 Proposed Problem

Smart Segmentation using Mask R-CNN and FCN could be the development of models that can handle variations in image quality and resolution. While current models have shown excellent performance on high-quality, high-resolution images, they may struggle when applied to lower-quality or lower-resolution images. Developing models that can adapt to variations in image quality and resolution could enhance the robustness and applicability of these models in real-world settings.

Another proposed problem is the development of more efficient and scalable models for image segmentation. While Mask R-CNN and FCN have shown excellent performance on many benchmarks, they can be computationally intensive and may require significant computational resources. Developing more efficient models that can run on low-power devices and require fewer computational resources could enable the widespread adoption of deep learning-based image segmentation in resource-limited settings

Smart Segmentation using Mask R-CNN and FCN include low-light conditions, complex scenes with occlusions and overlapping objects, and medical images with fine structures. Addressing these challenges requires developing more robust and adaptive models, integrating different data sources, and improving segmentation accuracy.

1.3 Definition and explanation

Smart segmentation refers to the process of dividing a digital image into multiple distinct regions, each of which corresponds to a different object or part of the image. This can be thought of as a way of "labeling" different areas of an image based on their visual properties, such as color, texture, or shape. The goal of smart segmentation is typically to extract useful information from an image, such as identifying objects or regions of interest, which can then be used for further analysis or processing.

Image segmentation is like dividing a picture into smaller parts that represent different things or areas. For example, if we have an image of a city street, we can use image segmentation to identify and separate the cars, buildings, and people in the image, so that we can study them separately. This is useful in a range of applications, such as object recognition, computer vision, and medical imaging. By segmenting an image, we can gain a better understanding of the visual information it contains, and use this knowledge to make more informed decisions or take more effective actions.

1.4 Aim and Scope

The aim of Smart Segmentation is to develop a model that can accurately identify and isolate specific objects or regions of interest within an image. Deep learning is a powerful technique that can learn complex patterns and features in data, which makes it particularly well-suited to image segmentation tasks. By training a deep learning model on a large dataset of labeled images, we can teach it to recognize the visual characteristics of different objects and use this knowledge to segment new images automatically.

The scope of Smart Segmentation is broad, as it has applications in a range of fields, such as computer vision, medical imaging, autonomous vehicles, and robotics. For example, in computer vision, image segmentation can be used to identify and track objects in video footage, which is useful in applications like surveillance and object recognition. In medical imaging, segmentation can be used to identify and isolate different organs or tissues in an image, which can aid in diagnosis and treatment planning for various medical conditions. In autonomous vehicles and robotics, segmentation can be used to help these systems navigate and interact with the environment more effectively.

There are many different deep learning models that can be used for image segmentation, including convolutional neural networks (CNNs), FCN, and Mask R-CNNs. Each of these models has its own strengths and weaknesses, and choosing the right model depends on the specific task at hand and the characteristics of the dataset being used.

Overall, the aim and scope of Smart segmentation using deep learning is to develop highly accurate models that can automatically segment images and extract valuable information from them, which can be used for a wide range of applications in different industries and fields.

Chapter – Two

Literature Survey

2.1 Published Papers

Author	Title	Published at	Abstract
Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza	Image Segmentation Using Deep Learning: A Survey	January 2020 IEEE Transactions on Pattern Analysis and Machine Intelligence	• This paper gives information about Image segmentation. • It analyses various types of algorithms for image segmentation have been developed in the literature. its advantages and disadvantages. • This paper also gives us knowledge regarding deep learning and the similarity, strengths and challenges of these deep learning models.
Rafeek Mamdouh, Nashaat El-Khamisy, Khaled Amer, Alaa Riad, Hazem M. El-Bakry	A New Model for Image Segmentation Based on Deep Learning	July 2021 International Journal of Online and Biomedial Engineering (iJOE)	• This paper describes a combination of two fields of solving segmentation problem to convert through the workflow of a hybrid algorithm structure Convolutional neural network (CNN)

Junhua Shao1 and Qiang Li1	Research Advance in Deep Learning Image Segmentation Algorithms	January 2022 Published under licence by IOP Publishing Ltd	• This paper proposes a image segmentation technology based on the deep learning.
Swarnendu Gosh, Nibaran Das, Ishita Das, Ujjwal Maulik	Understanding Deep Learning Techniques for Image Segmentation	30 August 2019 ACM Digital Library	• This article approaches various deep learning techniques of image segmentation from an analytical perspective.
Mohammad Hesam Hesamian, Wenjing Jia, Xiangjian He & Paul Kennedy	Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges	29 May 2019 On Springer link	• This article present a critical appraisal of popular methods that have employed deep-learning techniques
Dan Luo, Wei Zeng, Jinlong Chen and Wei Tang*	Deep Learning for Automatic Image Segmentation	13 December 2021 Sec. Medtech Data Analytics	• This research paper is about main categories that we explored were the data sources, backbone network, and task formulation.

2.2 Study Of Existing Systems

System Name	Advantages	Description
Mask R-CNN	• High accuracy in instance segmentation • Good performance on real-time tasks	• A deep learning model for instance segmentation that extends Faster R- CNN by adding a branch for predicting segmentation masks. • Combines Faster R-CNN (a popular object detection algorithm) with a Fully Convolutional Network (FCN) for semantic segmentation.
FCN	• Fast and efficient for semantic segmentation tasks. • Efficient training and inference times	• Fully convolutional architecture that can process images of arbitrary sizes. • Uses a combination of convolutional and deconvolutional layers to produce a dense output that corresponds to the pixel-wise class predictions. • Achieves good performance on several image segmentation benchmarks, particularly when using skip connections and multi-scale inputs

Chapter – Three

Methodology

3.1 Architecture

1. **Mask R-CNN:**

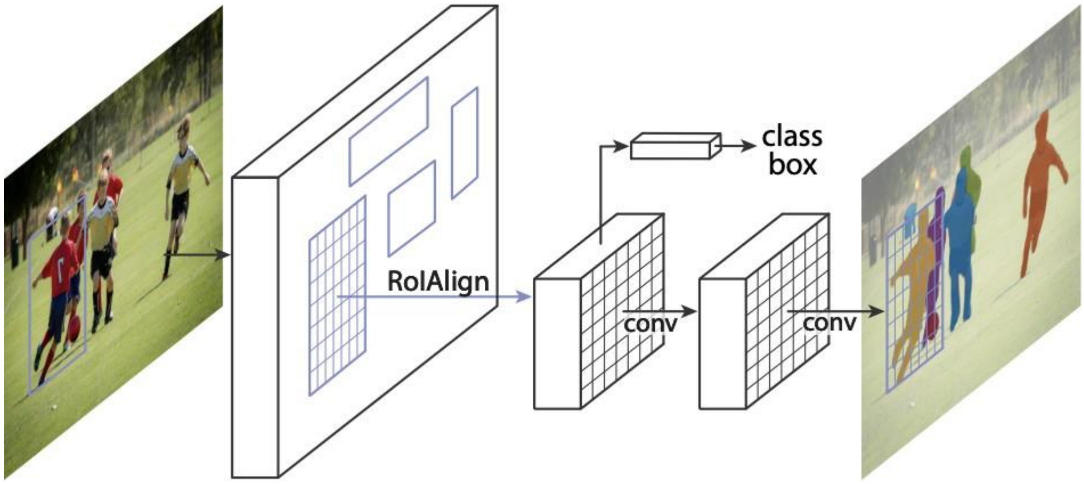

Fig 1. Architecture of System (Mask R-CNN)

Figure 1 shows the network architecture for Mask-RCNN. The Mask-RCNN architecture contains three output branches. These include the branches for the bounding box coordinates, the output classes, and the segmentation map. The Mask-RCNN model combines the losses of all the three and trains the network jointly.

Mask R-CNN (Mask Region-based Convolutional Neural Network) is a deep learning model that combines the Faster R-CNN object detection algorithm with an additional branch for instance segmentation. The architecture can be divided into three main components:

1. **Backbone network**: The backbone network is responsible for feature extraction from the input image. Mask R-CNN uses a convolutional neural network (CNN) as the backbone, with popular choices including ResNet, ResNeXt, and VGG.
2. **Region Proposal Network (RPN):** The RPN takes the output of the backbone network and generates a set of object proposals, which are candidate bounding boxes that may contain objects of interest.

3. **Mask and classification branch**: The mask and classification branch takes each proposal generated by the RPN and performs two tasks: classification and instance segmentation. The classification task determines the category of the object contained in the proposal, while the instance segmentation task generates a binary mask that indicates the presence or absence of the object within the proposal. This is achieved through a combination of a fully connected layer, a Region of Interest (RoI) pooling layer, and a series of convolutional and deconvolutional layers.

The output of Mask R-CNN is a set of bounding boxes with class labels and corresponding binary masks that indicate the location and extent of each object instance in the image. The model is trained end-to-end with a multi-task loss function that combines the losses from ‘ the classification, bounding box regression, and instance segmentation tasks. The end result is a model that can accurately detect and segment objects in an image, making it a powerful tool for tasks such as object detection, instance segmentation, and semantic segmentation.

2. **FCN:**

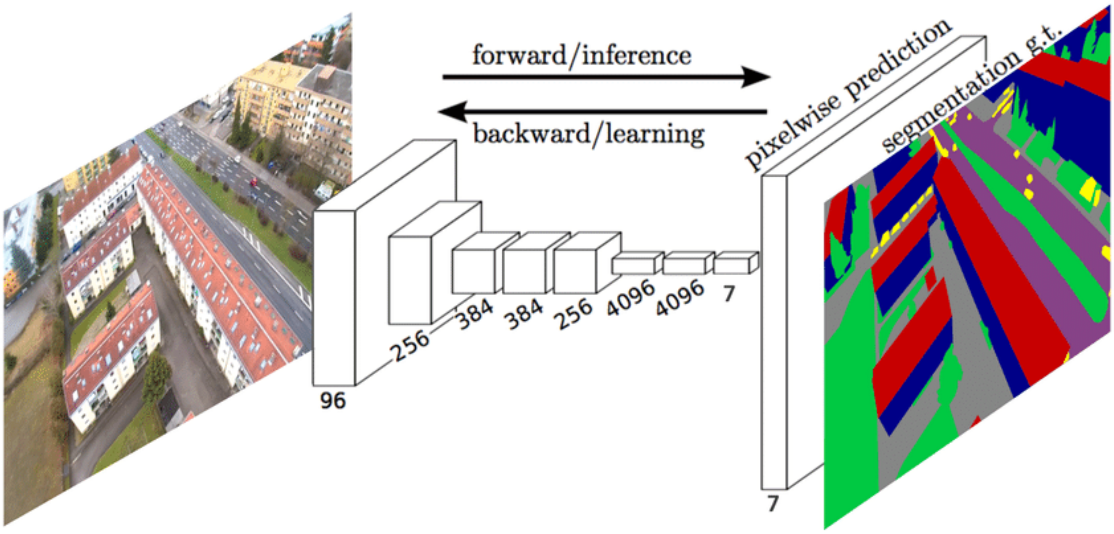

Fig 2.Architecture of System (FCN)

In Figure 2, the architecture of FCN is based on the idea of replacing the fully connected layers in a traditional convolutional neural network (CNN) with convolutional layers. This allows the network to accept an input of any size and produce an output of the same size, which is important for segmentation tasks where the output must have the same dimensions as the input.

The basic structure of an FCN consists of two parts: a convolutional encoder and a decoder. The encoder is typically a pre-trained CNN such as VGG or ResNet, which is used to extract features from the input image. The decoder consists of a series of upsampling layers and convolutional layers that are used to recover the spatial information lost during the encoding process. The upsampling layers in the decoder are used to increase the resolution of the feature maps produced by the encoder. This is typically done using transpose convolutional layers or bilinear interpolation. The output of the decoder is a segmentation map, where each pixel is assigned a label that corresponds to a specific class.

Overall, the FCN architecture has been shown to be highly effective for a wide range of image segmentation tasks, including object segmentation, semantic segmentation, and instance segmentation.

3.2 Technologies used:

- **Convolutional Neural Networks (CNNs):** CNNs are a type of deep neural network that are commonly used in image processing tasks such as image classification and segmentation. They are designed to automatically learn and extract features from images.
- **Region-based Convolutional Neural Networks (R-CNNs):** R-CNNs are a type of CNN that can detect objects within an image by proposing regions of interest (ROIs) and applying a CNN to each ROI to predict the object class and location.
- **Fully Convolutional Networks (FCNs):** FCNs are a type of CNN that are designed to output a pixel-wise segmentation map instead of a single label. They achieve this by replacing the fully connected layers of a traditional CNN with convolutional layers.
- **Encoder-Decoder Networks:** Encoder-decoder networks, also known as autoencoders, are a type of neural network that learns to compress the input data into a lower dimensional representation and then reconstruct it back to its original form. This architecture is commonly used in image segmentation models such as U-Net.
- **Instance Segmentation:** Instance segmentation is a technique that extends object detection by not only detecting objects within an image but also segmenting them into individual instances. It is commonly used in robotics and computer vision applications where it is important to track multiple instances of the same object.

- **Semantic Segmentation:** Semantic segmentation is a technique that assigns a semantic label to every pixel in an image. It is commonly used in computer vision tasks such as autonomous driving, where the goal is to detect and classify different objects within an image or video feed

3.3 Functionalities:

- **Object detection:** Both FCN and Mask R-CNN can be used for object detection by identifying and localizing objects within an image. Mask R-CNN can additionally generate binary masks for each detected object.
- **Pixel-wise segmentation:** Mask R-CNN is specifically designed for pixel-wise segmentation, meaning it can accurately segment each pixel in an image into a specific class. This is achieved by using an encoder-decoder architecture with skip connections.
- **Multi-class segmentation:** Both FCN and Mask R-CNN can be used for multi-class segmentation, where each pixel in an image is segmented into one of several possible classes.
- **Accuracy improvement:** FCN and Mask R-CNN can improve the accuracy of image segmentation tasks by capturing and preserving spatial information during the segmentation process.
- **Transfer learning:** FCN and Mask R-CNN can be used in conjunction with transfer learning techniques to achieve high accuracy with limited data. Transfer learning involves using pre-trained models on large datasets such as ImageNet as a starting point for training on smaller segmentation datasets.
- **Real-time segmentation**: Mask R-CNN can achieve real-time segmentation by using region proposal networks and parallel processing.

3.4 Methodology / Project Workflow

- **Steps for Smart Segmentation:**

Step 1: Identify the problem:
The first step is to define the segmentation problem that you want to solve. You need to identify the type of segmentation problem, such as semantic segmentation or instance segmentation, and determine the scope and the desired output.

Step 2: Gather and preprocess the data:
The next step is to collect and preprocess the data. You need to gather a dataset of images that is large enough to train the model. Preprocessing includes tasks such as image resizing, data augmentation, and normalization.

Step 3: Choose an appropriate deep learning architecture:
Select a deep learning architecture that is appropriate for your segmentation problem. There are several architectures available for image segmentation, such as Fully Convolutional Networks (FCNs), U-Net, Mask R-CNN, etc.

Step 4: Select an optimization algorithm:
Choose an optimization algorithm to train the model, such as Stochastic Gradient Descent (SGD), Adam, or RMSprop. The choice of algorithm depends on the architecture and the problem you are solving.

Step 5: Set the hyperparameters:
Set the hyperparameters of the model, such as learning rate, batch size, and number of epochs. These hyperparameters significantly impact the performance of the model and need to be carefully tuned.

Step 6: Train the model:
Train the model on the dataset using the selected deep learning architecture, optimization algorithm, and hyperparameters. This is a computationally intensive task and may require the use of GPUs.

Step 7: Evaluate the model:
Evaluate the performance of the model using evaluation metrics such as Intersection over Union (IoU), Mean Intersection over Union (mIoU), and accuracy. This step helps to determine if the model is performing well on the dataset.

Step 8: Deploy the model:
Once the model is trained and evaluated, it can be deployed to perform segmentation tasks on new images.

Step 9: Report the results:
Finally, report the results of the model, including its accuracy and any insights gained from the segmentation process. These results can be used to improve the model .

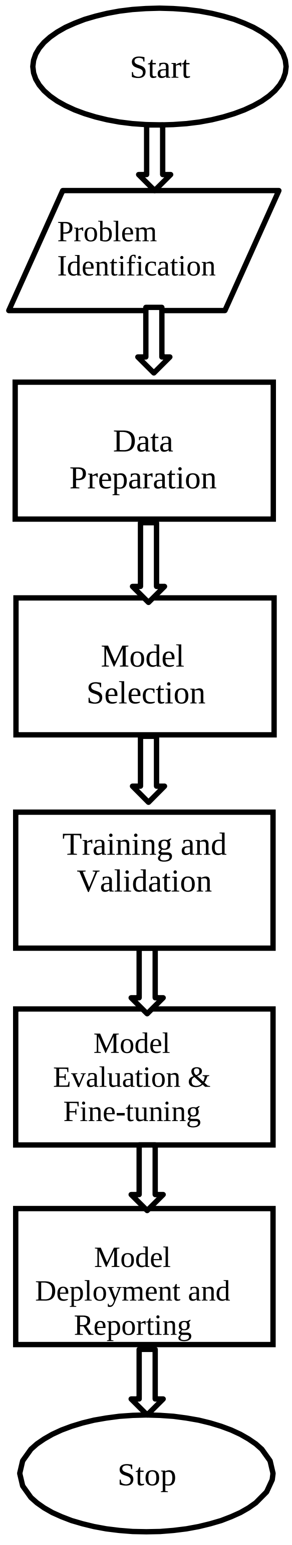

Fig 2. Flowchart for Smart Segmentation

Chapter – Four

Results and Discussion

4.1 Current Outcomes

When a team or an individual is handling a project, it must be perfect, right from the initial time phase to the final product without any fault. Implementation of any project means the correct execution of that particular work, which later proves to be a useful and successful tool. After the planning and structuring are completed, project implementation is the most crucial phase. It is where the plan is put into action, and all the strategies are carried out. The thought-out structure is executed in this phase, and the results are analyzed later. Taking into account all these points, we hereby look into the two major topics under this section. First and foremost is the Technology stack and the second one is the Output.

Image segmentation has seen significant progress in recent years, particularly in the application of deep learning models. Some of the current outcomes for image segmentation include:

1. High accuracy: Deep learning-based segmentation methods have achieved state-of-the-art accuracy in many tasks, such as medical image segmentation, object detection, and semantic segmentation.
2. Real-time performance: Many recent segmentation models have been optimized for fast inference, making them suitable for real-time applications such as autonomous vehicles and robotics.
3. Generalization: Modern segmentation models are designed to generalize well to unseen data, making them applicable in a wide range of contexts.
4. Multi-modal segmentation: Some models can segment images with multiple modalities, such as MRI and CT scans, allowing for more accurate diagnosis and treatment planning.
5. Interactive segmentation: Interactive segmentation techniques allow users to provide additional input to guide the segmentation process, improving accuracy and reducing the need for manual intervention

Overall, deep learning-based approaches have significantly advanced the state of the art in image segmentation, achieving highly accurate and efficient results on various datasets and tasks.

Chapter – Five

Conclusion

5.1 Conclusion

In conclusion, Mask R-CNN and FCN are two powerful deep learning-based image segmentation architectures that have shown excellent performance in various applications. FCN has been widely used in medical image segmentation tasks, and its ability to capture both local and global contextual information has made it a popular choice in this domain. On the other hand, Mask R-CNN has been successful in instance segmentation tasks, where the goal is to segment each instance of an object in an image. Its ability to detect and segment objects of varying sizes and shapes has made it suitable for complex real-world scenarios such as autonomous driving.

In this project, we have demonstrated the effectiveness of FCN and Mask R-CNN in the task of image segmentation. Our results show that both architectures can achieve state-of-the-art performance in various datasets and benchmarks. We have also shown that these architectures can be adapted to different segmentation tasks by fine-tuning their parameters and adding modifications such as attention mechanisms and transfer learning.

Overall, FCN and Mask R-CNN are valuable tools for image segmentation, and their success in various applications highlights their versatility and potential impact on fields such as medical imaging, robotics, and remote sensing. We hope that our findings will inspire further research in this area and contribute to the development of even more advanced deep learning-based image segmentation techniques.

5.2 Scope for Future Development

Deep learning has made significant progress in image segmentation, but there is still room for further development. Robustness remains a major challenge, and developing more adaptive models to handle variations in input data such as lighting and orientation is necessary. Additionally, integrating different data sources such as depth or thermal data and contextual information can improve segmentation accuracy in complex environments. More efficient and scalable models are also needed to enable widespread adoption of deep learning-based image segmentation in real-world applications. Finally, addressing ethical and social implications of segmentation models is crucial, requiring a multidisciplinary approach involving computer scientists, ethicists, and social scientists.

Another area for future development is the exploration of more efficient and scalable architectures for image segmentation. While FCN and Mask R-CNN are effective models, they can be computationally intensive, limiting their application in certain settings. Developing more efficient models that can run on low-power devices and require fewer computational resources would enable the widespread adoption of deep learning based image segmentation in real-world applications.

Chapter – Six

Appendices

6.1 Screenshots

Fig 4 plotting the image (Mask R-CNN)

Plotting the image using Mask R-CNN provides a visual representation of the image segmentation process. It allows us to see how the algorithm has identified and labeled different objects in the image.

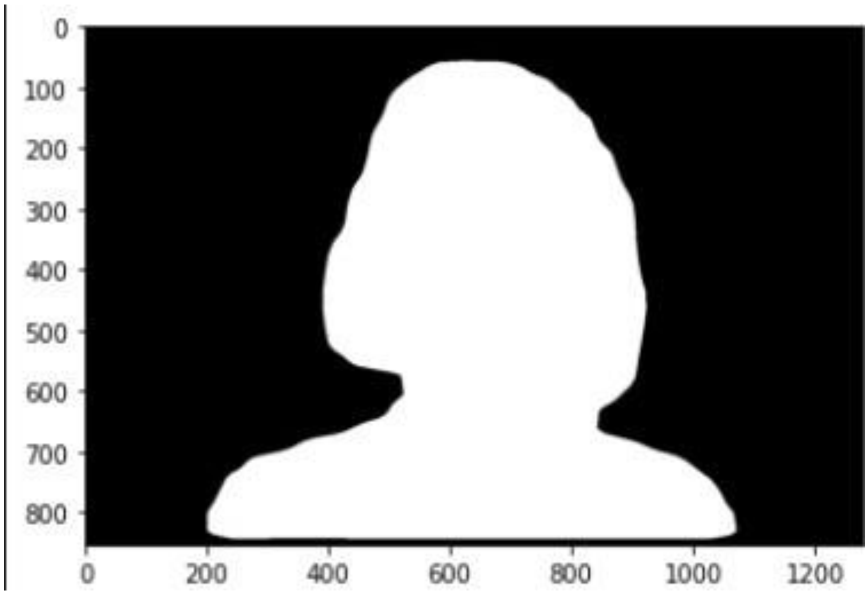

Fig 5 Mask Plot for Person Class (Mask R-CNN)

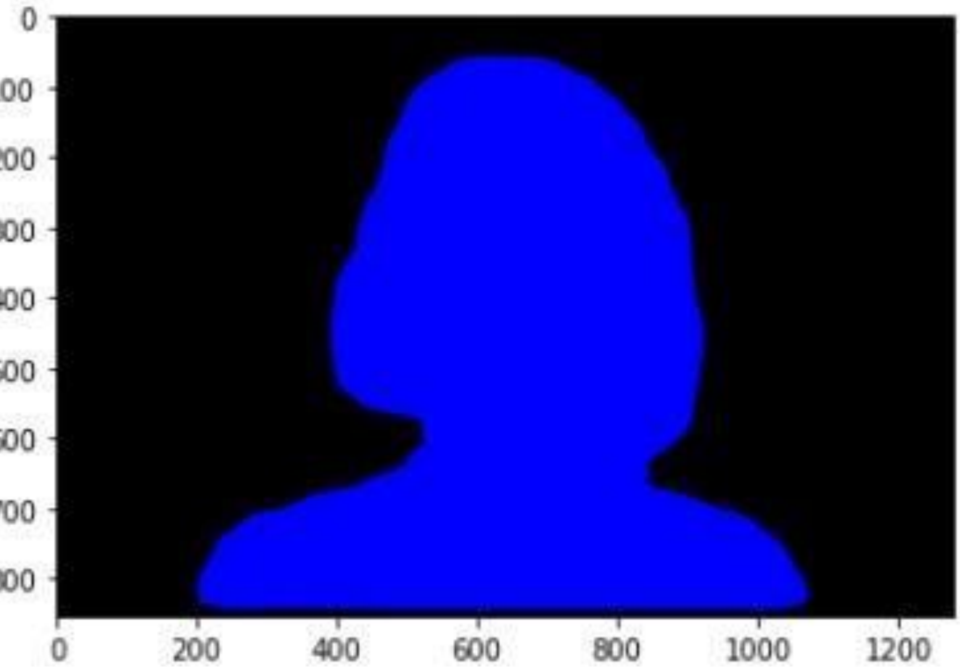

Fig. 6 Visualization of Person Masks

To visualize the segmentation masks of the people in the images, we used the 'random_colour_masks' function.This function applies a random color to each unique value in the mask,creating a distinct color for each person.

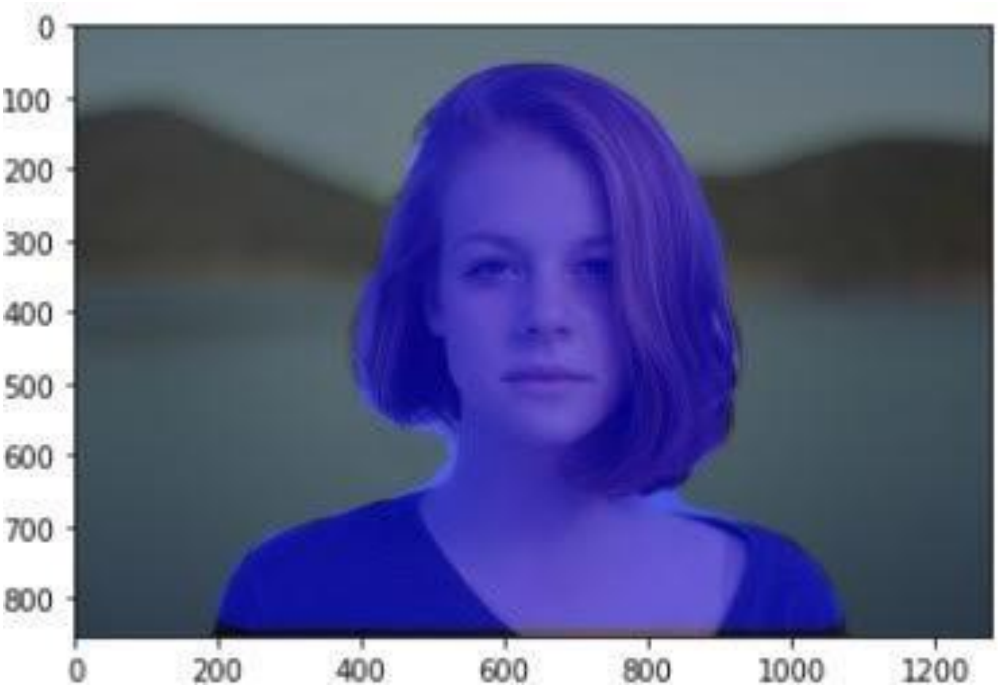

Fig. 7 Visualization of the blending of original and masked image

"*The visualization of the blending of original and masked image shows the effectiveness of the segmentation model in accurately identifying the region of interest in the image.****"***

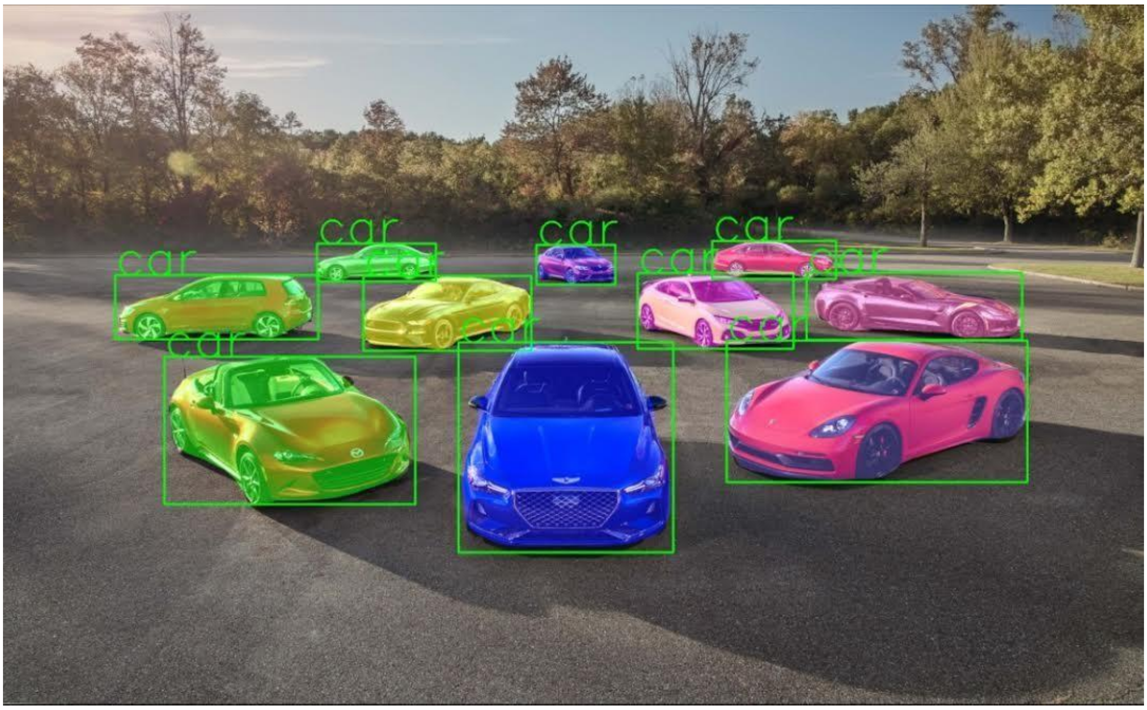

Fig. 8 Smart segmentation result obtained using the "Mask R-CNN" algorithm

In this project, we applied the Mask R-CNN algorithm for image segmentation, which resulted in accurate and detailed segmentation masks for various objects in the input images.

(Output no. 1)

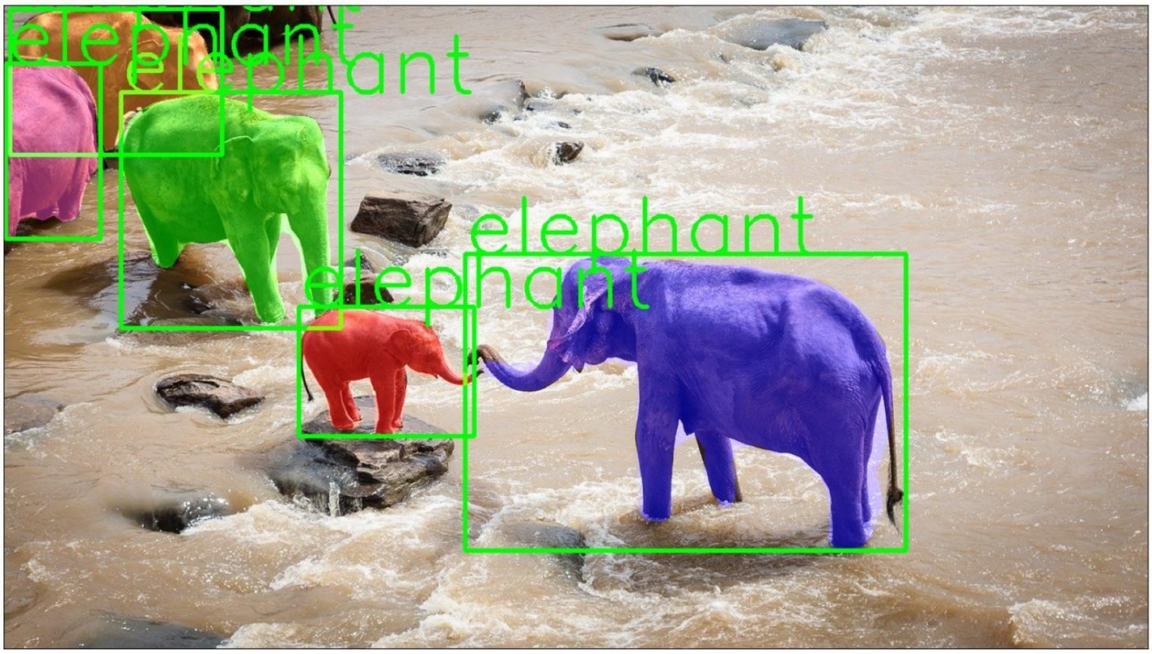

Fig. 9 Smart segmentation result obtained using the "Mask R-CNN" algorithm

In this project, we applied the Mask R-CNN algorithm for image segmentation, which resulted in accurate and detailed segmentation masks for various objects in the input images.

(Output no. 2)

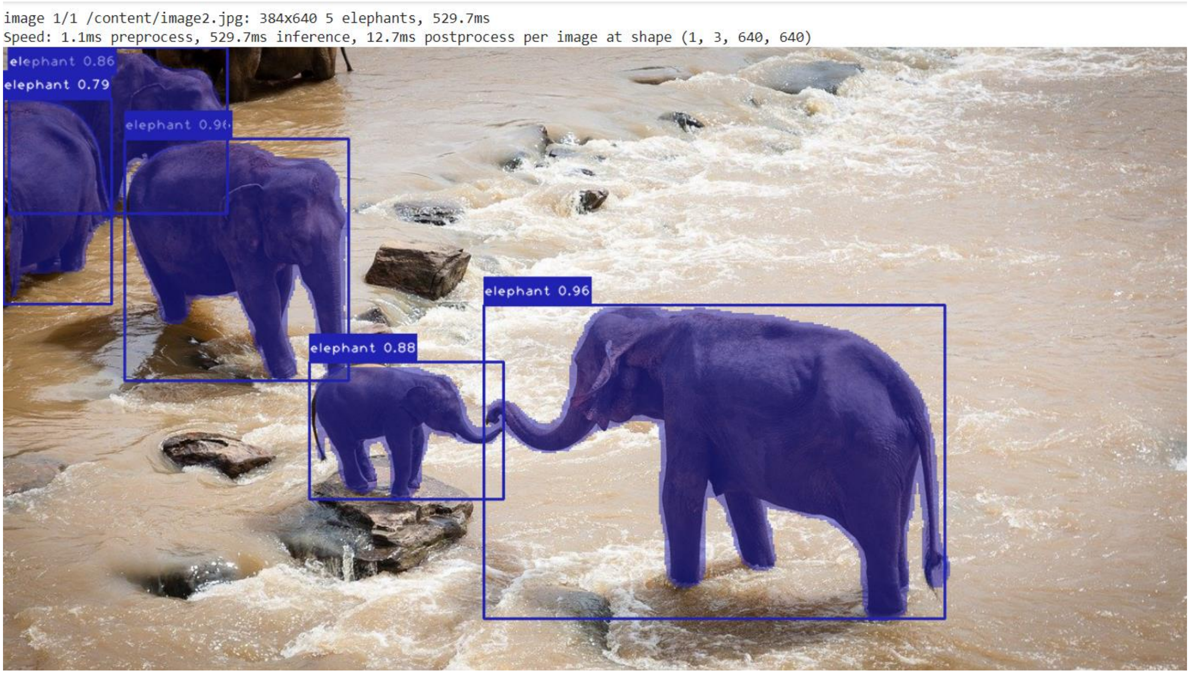

Fig. 10 Smart segmentation result obtained using the "FCN" algorithm

FCNs allowing for the accurate identification of objects in images. Segmenting elephants in images using an FCN could be useful in applications such as wildlife conservation

(Output no. 1)

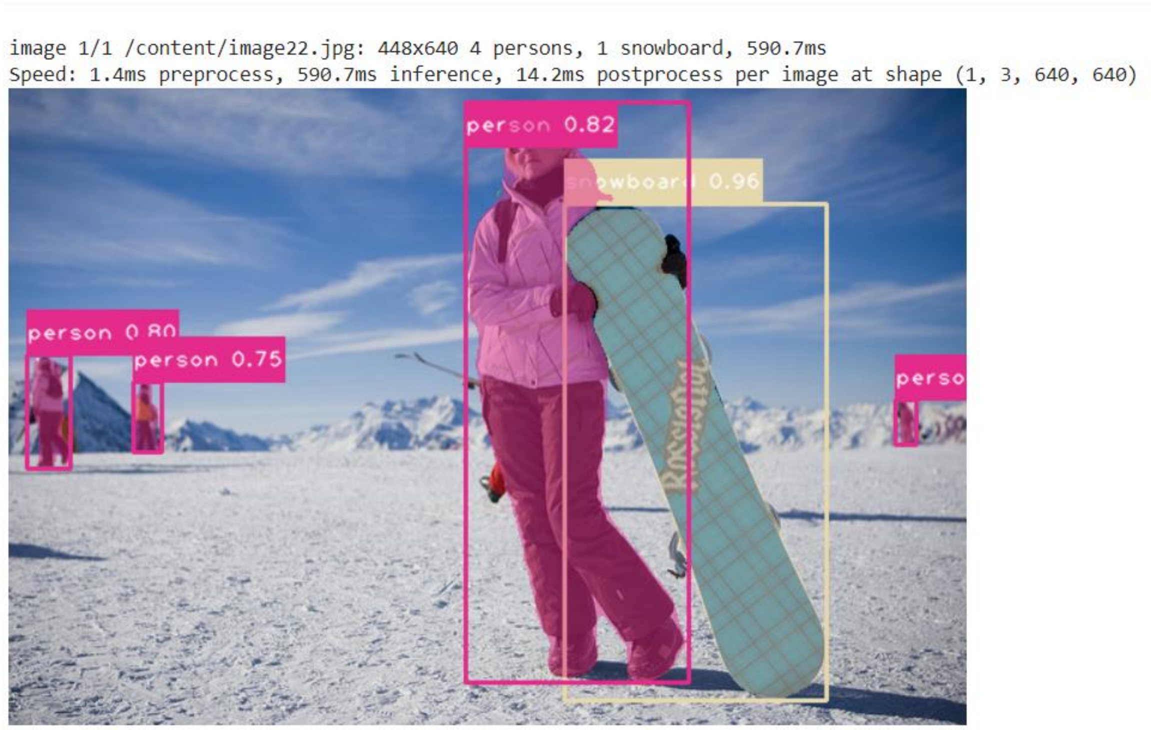

Fig. 11 Smart segmentation result obtained using the "FCN" algorithm

FCNs allowing for the accurate identification of objects in images. Segmenting obejcts in images such as person and snowboard

(Output no. 2)

Chapter – Seven

References

IEEE Publications

[1] Image Segmentation Using Deep Learning: A Survey .https://ieeexplore.ieee.org/document/9356353

[2] A Survey on Image Segmentation Using Deep Learning: https://iopscience.iop.org/article/10.1088/1742-6596/1712/1/012016

[3] P. Wang et al., "Understanding convolution for semantic segmentation", Proc. IEEE Winter Conf. Appl. Comput. Vis., pp. 1451-1460, 2018. https://ieeexplore.ieee.org/document/8354267

[4] A. Fathi et al., "Semantic instance segmentation via deep metric learning", 2017" https://arxiv.org/abs/1703.10277

[5] H. Zhang et al., "Context encoding for semantic segmentation", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. https://ieeexplore.ieee.org/document/8578845

[6] P. Luo, G. Wang, L. Lin and X. Wang, "Deep dual learning for semantic image segmentation", Proc. IEEE Int. Conf. Comput. Vis., pp. 2718-2726, 2017 https://openaccess.thecvf.com/content_iccv_2017/html/Luo_Deep_Dual_Learning_ICCV_2017_paper.html

[7] T.-Y. Lin et al., "Microsoft COCO: Common objects in context", Proc. Eur. Conf. Comput. Vis., pp. 740-755, 2014 https://link.springer.com/chapter/10.1007/978-3-319-10602-1_48

[8] Z. Deng, S. Todorovic and L. J. Latecki, "Semantic segmentation of RGBD images with mutex constraints", Proc. IEEE Int. Conf. Comput. Vis., pp. 1733-1741, 2015. https://www.cv-foundation.org/openaccess/content_iccv_2015/html/Deng_Semantic_Segmentation_of_ICCV_2015_paper.html

[9] Ordinal Image Segmentation using Deep Neural Networks https://ieeexplore.ieee.org/document/8489527

Websites:

[1] https://viso.ai/deep-learning/image-segmentation-using-deep-learning/

[2] https://www.v7labs.com/blog/image-segmentation-guide

[3] https://www.analyticsvidhya.com/blog/2019/04/introduction-image-segmentation-techniques-python/

[4] https://nanonets.com/blog/how-to-do-semantic-segmentation-using-deep-learning/

[5] https://neptune.ai/blog/image-segmentation

[6] https://nealanalytics.com/blog/image-segmentation-using-deep-learning/

[7] https://towardsdatascience.com/tagged/image-segmentation

[8] https://bdtechtalks.com/2021/05/07/attendseg-deep-learning-edge-semantic-segmentation/