Indian Academy of Sciences, Bengaluru
Indian National Science Academy, New Delhi
The National Academy of Sciences India, Prayagraj

SUMMER RESEARCH FELLOWSHIPS — 2024

Format for the four-week Report*,^,@

Name of the candidate	:	Nithish Choubi
Application Registration no.	:	ENGS3700
Date of joining	:	20/05/2024
Name of the guide	:	Prof. Sumohana S. Channappayya
Guide's institution	:	IIT Hyderabad
Place of stay during the tenure of the fellowship	:	✓ Hostel provided by Guide: IIT Hyderabad, Vishveshwaraya Block, 109 Own arrangement Other (Specify) -

Signature of the candidate
Date: 18/06/24

Signature of the guide
Date: 19/6/24

	INSPIRE/KVPY FELLOWSHIP (please fill this box)#		
1.	I am currently a recipient of	INSPIRE FELLOWSHIP	☐ Yes / ☑ No
		KVPY FELLOWSHIP	☐ Yes / ☑ No
		If, YES, fill cols. 2, 3 & 4	
2.	**INSPIRE/KVPY** Fellowship is from – [month]/ – [yr] to – [month]/ – [yr]		
3.	I receive a monthly fellowship of Rs. – from INSPIRE/KVPY towards my living expenses		
4.	I also receive towards contingencies a sum of Rs. – per year		
	I affirm that the information given above is correct.		Signature of the candidate

IMPORTANT NOTES:

*** The four-week report could be between 300 and 350 words.**
^ This format should be the first page of the report and should be stapled with the main report.
 # Mandatory to fill this section, this should be filled and signed by you even if you are not an INSPIRE/KVPY Fellow. Otherwise release of fellowship amount will be withheld.
@ The hard copy of the duly signed report should reach the Academy office within 10 days of completing the first month fellowship. If delayed the fellowship amount will not be paid.

(For office use only; do not fill/tear)

Candidate's name:	Fellowship amount:
Student: Teacher:	Deduction:
Guide's name:	Amount to be paid:
KVPY Fellow: INSPIRE Fellow:	A/c holder's name:
PFMS Unique Code:	
Others	

FOUR WEEK REPORT

TITLE: Exploring Semantic Segmentation
NAME: Nithish Chouti (ENGS3700)
GUIDE NAME: Prof. Sumohana S. Channappayya
INSTITUTE: Indian Institute of Technology, Hyderabad
DATE: 18th June, 2024

The project led by Prof. Sumohana S. Channappayya from the Electrical Engineering Department at the Indian Institute of Technology, Hyderabad, focuses on 'Developing Vision Capabilities for Bi-Pedaled Robot'. The research team, comprising Athira, Ambarish, Akshad, Ashwath Sreeram, and other interns, is working on various subtasks of the project such as object detection, image segmentation, change detection, and crowd counting. This report specifically delves into the task of image segmentation under the expertise of Ashwath Sreeram, as I were curious and interested to learn about image segmentation.

1 What is Image Segmentation?

Image segmentation is a process in computer vision and digital image processing, focusing on dividing an image into segments or groups that share similar characteristics, each labeled accordingly [2]. This process operates digitally, utilizing the pixel representation of an image to cluster these pixels into meaningful segments effectively. More than merely classifying images, image segmentation takes a step further by not only classifying but also localizing various elements within an image [1]. It enhances the classification by delineating the exact boundaries of objects, making it at a more advanced level where it not only identifies but also outlines objects within an image, thereby serving as an advanced form of image classification.

2 Types of Image Segmentation

2.1 Semantic Segmentation

This method labels each pixel of an image with a class label, such as "cat," "dog," or "road." All pixels belonging to the same category are treated as a single entity, without distinguishing between individual instances of that category [1,2]. Figure 2 shows the semantic segmented image of the Figure 1.

2.2 Instance Segmentation

This approach extends semantic segmentation by not only classifying each pixel but also distinguishing between separate instances of the same category [1,2]. For example, in an image with several cats, it identifies and segments each cat individually. Figure 3 shows the instance segmented image of the Figure 1.

2.3 Panoptic Segmentation

This technique combines both semantic and instance segmentation [1,2]. It assigns a class label to each pixel and differentiates between individual instances of objects while also providing a complete understanding of all elements within the image. Figure 4 shows the panoptic segmented image of the Figure 1.

Note: As the project specifically deals with semantic segmentation, I will devel more into the models used for semantic segmentation and explore it in details. Some prerequisite knowledge and concepts about Convolutional Neural Networks [7] and Vision Transformers [5] were required. So I had to read about the details of CNNs and Transformers to get an understanding about the architecture and implementation.

Figure 1: Original Image

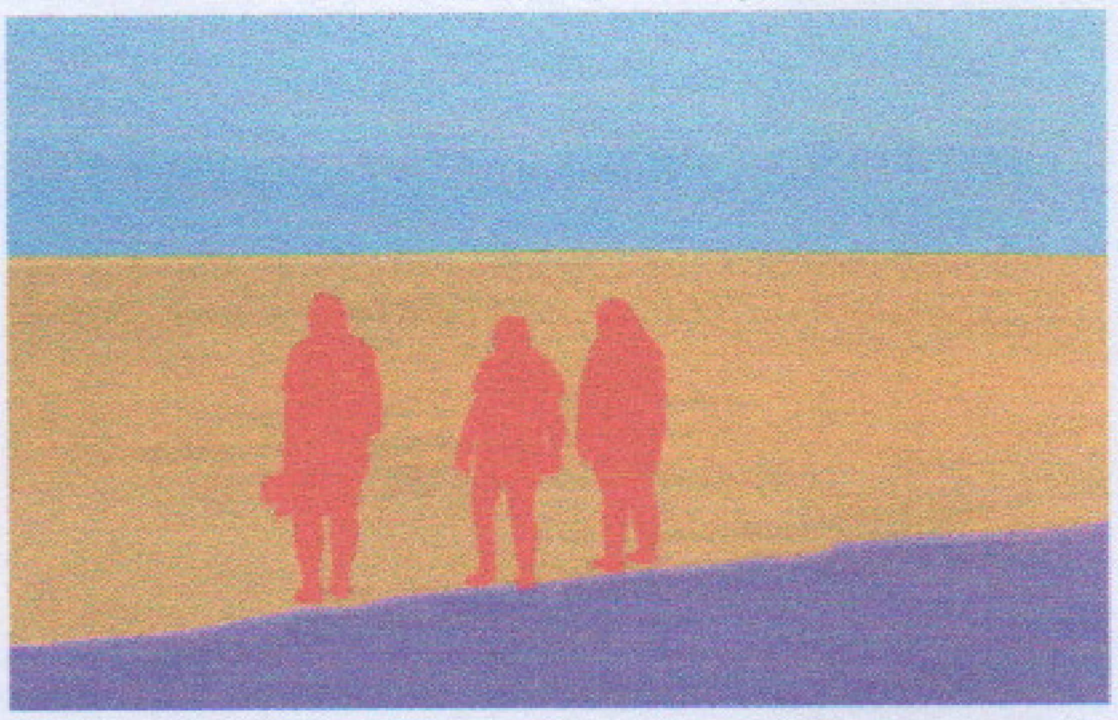


Figure 2: Semantic Segmentation Image

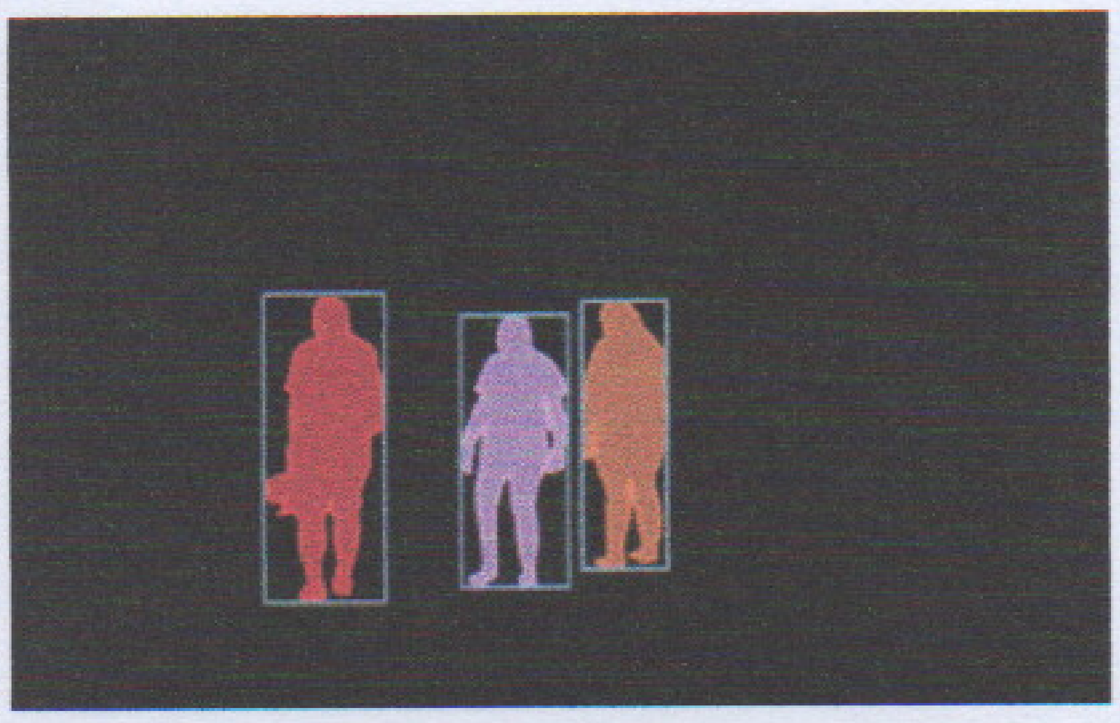


Figure 3: Instance Segmentation Image

Figure 4: Panoptic Segmentation Image

3 Models for Semantic Segmentation

3.1 DeepLabV3+

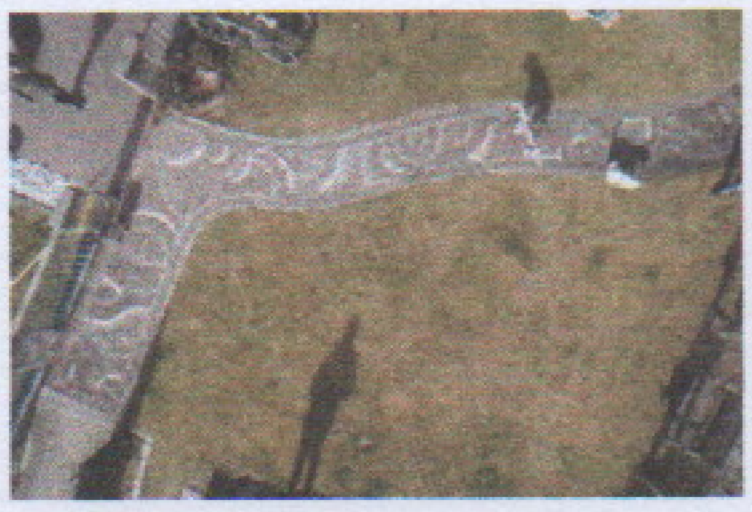
DeepLabv3+ is an advanced image segmentation model developed by Google. It enhances the DeepLabv3 architecture by incorporating an encoder-decoder structure, improving the capture of detailed object boundaries and refining segmentation results [4]. The encoder uses atrous convolution to extract multi-scale contextual information, while the decoder refines the segmentation maps, leading to more accurate and detailed segmentation, especially for complex scenes.

(a) Picture (b) Ground Truth (c) DeepLabV3+ Output

Figure 5: Semantic Segmentation using DeepLabv3+ model [3]

3.2 Segformer

SegFormer is a state-of-the-art image segmentation model that leverages transformers for efficient and effective segmentation [10]. It introduces a hierarchical transformer encoder that captures both local and global features without the need for convolutions. SegFormer is known for its simplicity and scalability, achieving high performance across various benchmarks with lower computational costs compared to traditional convolutional neural network-based approaches.



(a) Picture

(b) Ground Truth

(c) Segformer Output

Figure 6: Semantic Segmentation using Segformer model [3]

4 Current Working Topics

- I have studied the architecture of DeepLabV3+ and Segformer by reviewing their research papers. I am also getting familiar with the code recommended [9] for these models.
- DeepLabV3+ uses various backbone models to enhance feature extraction and improve segmentation accuracy. The project explores multiple backbone models like HRNet, ResNet, Xception, and MobileNetV2 [9]. I have referred the MobileNetV2 [8] backbone and will try to use the MobileNetV3 [6] backbone to compare if the model performs better.
- In continuation to the above, I am studying the architecture of MobileNetV3 to implement it as a backbone for DeepLabV3+. MobileNetV3 employs efficient convolutional operations and incorporates squeeze-and-excitation (SE) blocks to boost performance while maintaining low computational complexity.
- Further as a next step, I will explore about weights quantization to see if the performance improves from the base model.

References

[1] Image segmentation guide. `https://www.v7labs.com/blog/image-segmentation-guide`.

[2] Introduction to image segmentation techniques in python. `https://www.analyticsvidhya.com/blog/2019/04/introduction-image-segmentation-techniques-python/`.

[3] Qing Bu, Wei Wan, and Elizaveta Savitskaya. Comparative analysis of semantic segmentation methods for satellite images segmentation. 2023.

[4] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 801–818, 2018.

[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. 2020.

[6] A. Howard, M. Sandler, G. Chu, L. C. Chen, B. Chen, M. Tan, and H. Adam. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1314–1324, 2019.

[7] K. O'shea and R. Nash. An introduction to convolutional neural networks. 2015.

[8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4510–4520, 2018.

[9] VainF. Deeplabv3plus-pytorch repository. `https://github.com/VainF/DeepLabV3Plus-Pytorch`.

[10] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems*, 34:12077–12090, 2021.