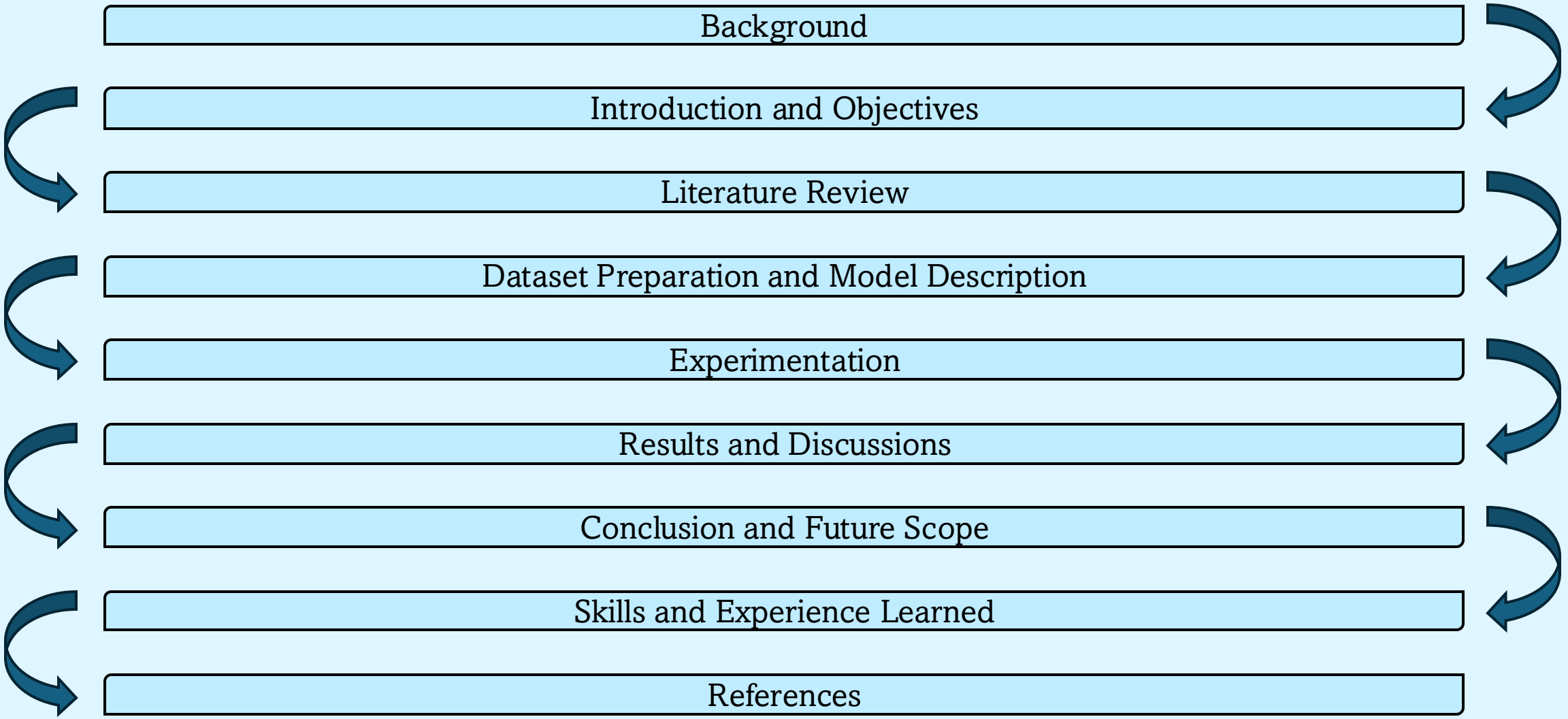
Industrial Training Defense (ME4T001)

AI Research Intern at Ranvox Tech. Pvt. Ltd.

By
Kovid Sharma
21ME01026

School of Mechanical Sciences
Indian Institute of Technology, Bhubaneswar

Content / Outlines



Background

- Ranvox Technology Pvt. Ltd. manufactures Unmanned Vehicles like UAVs and USVs for agriculture, surveillance, security, and research purposes.

- It engages in research pertaining to Unmanned Vehicles, Composite Materials, Sensors, and Artificial Intelligence.

- Drones help in monitoring large areas effectively and reduce the risk to security staff.

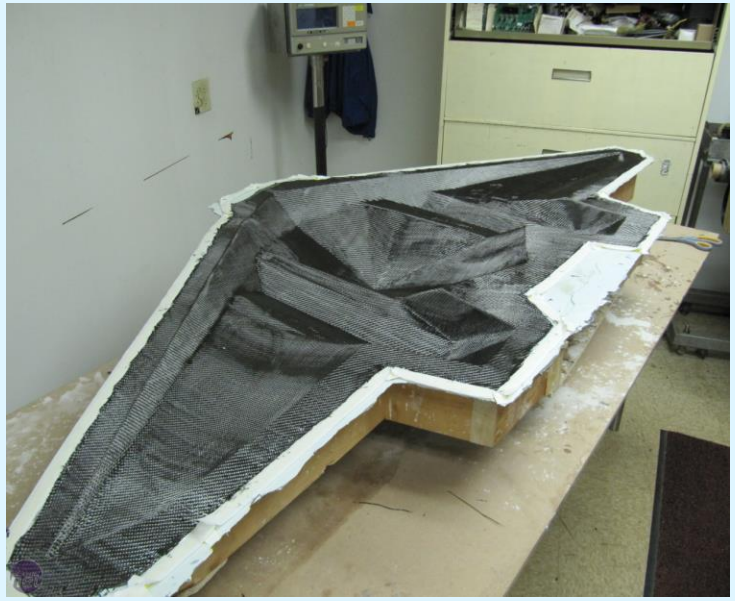


Fig. 1: Moulding Carbon fiber
Image ref. - https://bit-tech.net/guides/carbon-fibre-a-modders-guide/1/

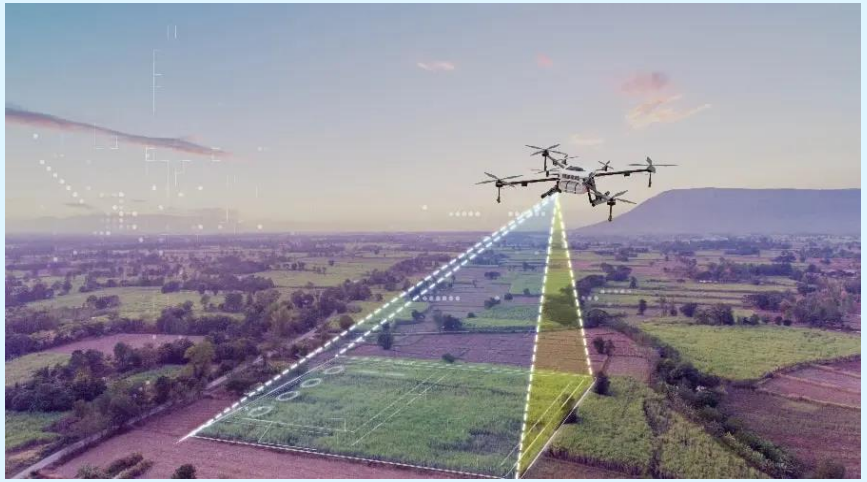


Fig. 2: Drone Mapping
Image ref. - https://www.dronepilotgroundschool.com/drone-mapping-software/

Introduction and Objectives

The following were the objectives of the internship:

1. To create a computer vision model to detect coastline from UAV images.

2. To access the present state of Object Detection in scenarios captured by drones and their real-time implementation on Embedded GPUs.

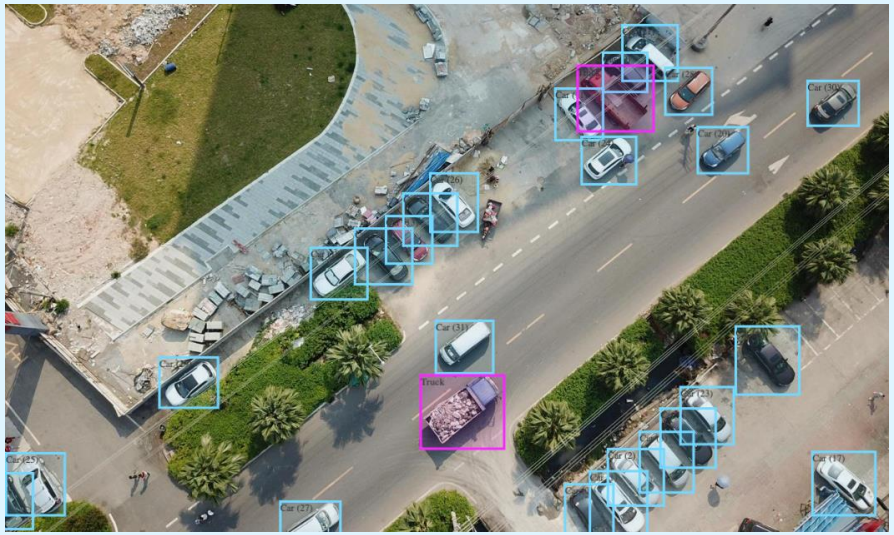


Fig. 3: Drone Object Detection
Source- VisDrone Dataset[1]

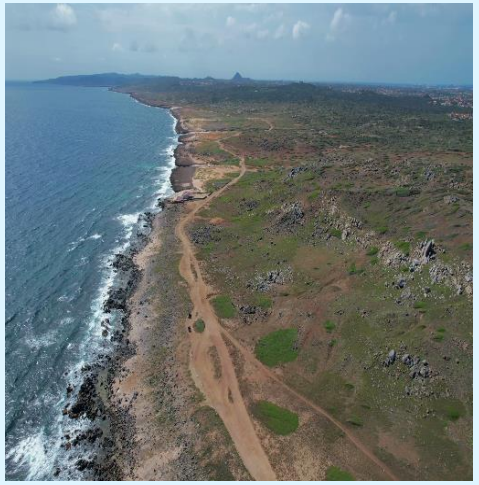


Fig. 4: Drone Image of Coastline

Introduction and Objectives

- Computer Vision enhances the perceptive abilities of drones.

Drone Object Detection

- Object detection in drones ensures accurate identification and categorization.
- Deep Learning based object detectors have shown exceptional performance in both accuracy and efficiency.

Coastline Mapping

- Coastline is an important feature of the coastal morphology.
- Monitoring coastline is crucial in safeguarding the environment[2].
- Required to support the object detection model in determining whether an object was in the sea or on the beach
- Semantic Segmentation can be used to detect the coastline.

Literature Review

- Two types of detectors: Anchor-based and Anchor free
- Anchor-based:
  - Tile a large no. of preset anchors on the image
  - Predict the category and refine the coordinates of these anchors one or several times
  - Types: Single-stage and Two-Stage
  - Two-stage methods: 1st generate region proposals and then they are sent to the second stage to generate the final bounding boxes
  - Single-stage methods: do not generate proposals but still use anchor boxes. Directly classify & regress the anchors to get the final bounding boxes

Literature Review

- Anchor-Based Two-stage examples: RCNN Family[3] (RCNN, Fast RCNN, Faster RCNN, Cascade RCNN), VFNet
- Anchor-based single-stage examples: Single-shot detection(SSD), YOLO[4], RetinaNet, FCOS[5], DETR, EfficientDet
- Anchor-free methods:
  - Directly find objects without preset anchors
  - Transforms task to key point & size estimation
  - Eliminates the Hyper-parameters related to anchors
- Anchor-free examples: CornerNet, ExtremeNet, CenterNet, FoveaBox, YOLOX

Literature Review

- Drones always navigate at different altitudes which causes the object scale to vary violently and burdens the optimization of networks.[6]
- Images contain objects with high density which causes occlusion.
- High speed and low altitude flight -> blur motion on the densely packed objects.
- The image comprises confusing geographic elements because it covers a large area
- Lack of computation power of embedded GPUs.
- Useful Strategies: Data augmentation, Multi-scale testing, Multi-model integration, Utilizing extra classifier.

Dataset Preparation

- Created by sourcing stock videos from the website DroneStock[7]
- Videos were converted to images using OpenCV
- Images were taken every 30 ms and were resized to a resolution of 512 ×512
- labelled using the semantic segmentation workspace of Label Studio[8]
- Binary masks created were exported in PNG format.
- A total of 230 images were labelled

Fig. 5: Image and its binary mask

Dataset Preparation

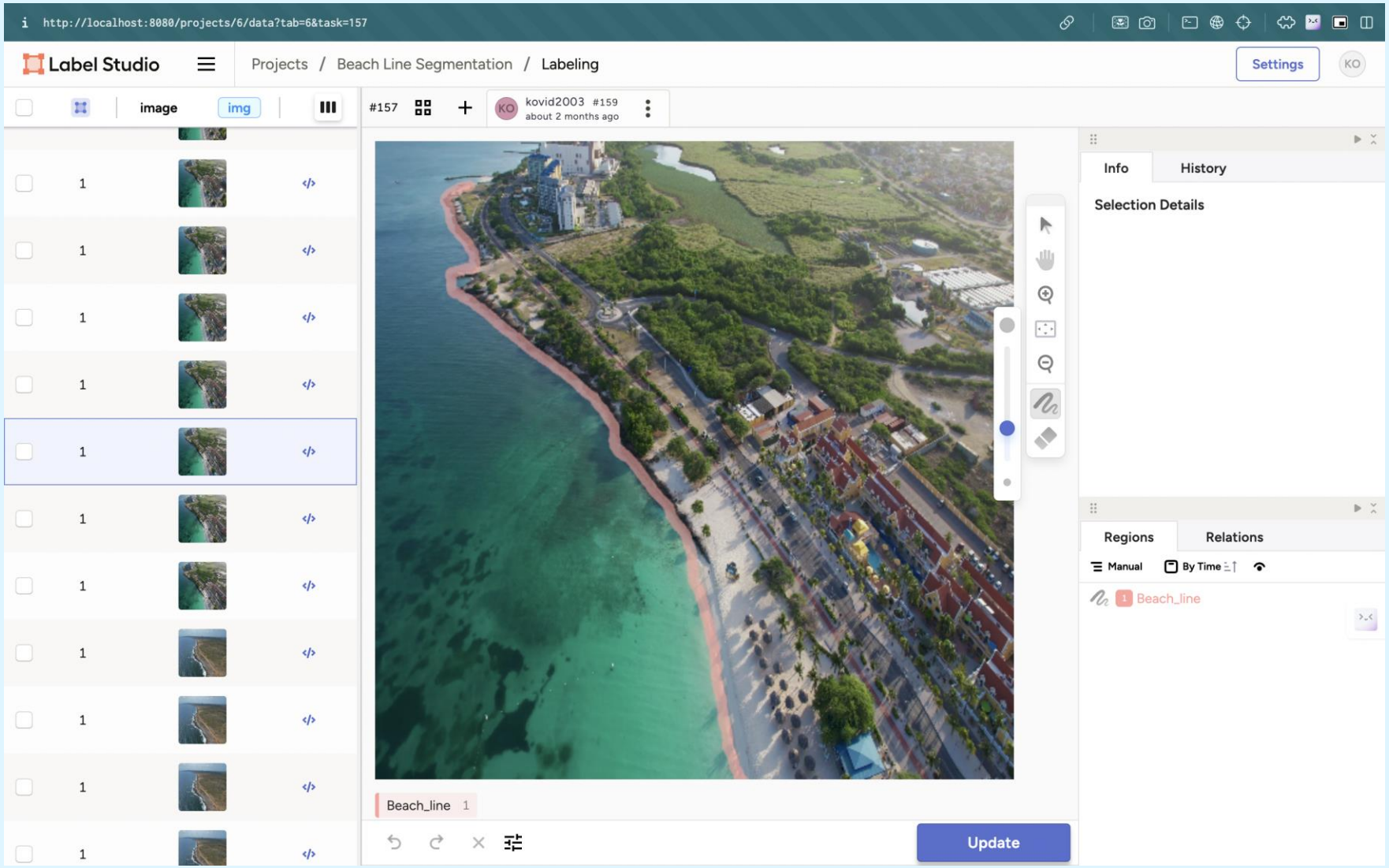


Fig. 6: Semantic Segmentation Workspace in Label Studio

Model Description

- The model used for the segmentation task is a modified U-Net Architecture[9].
- An input size of 512 is used instead of 572.
- Batch Normalization layers are included for faster convergence and training.
- Model was created using the PyTorch library and Python Programming language.
- The model contains 31,037,633 trainable parameters

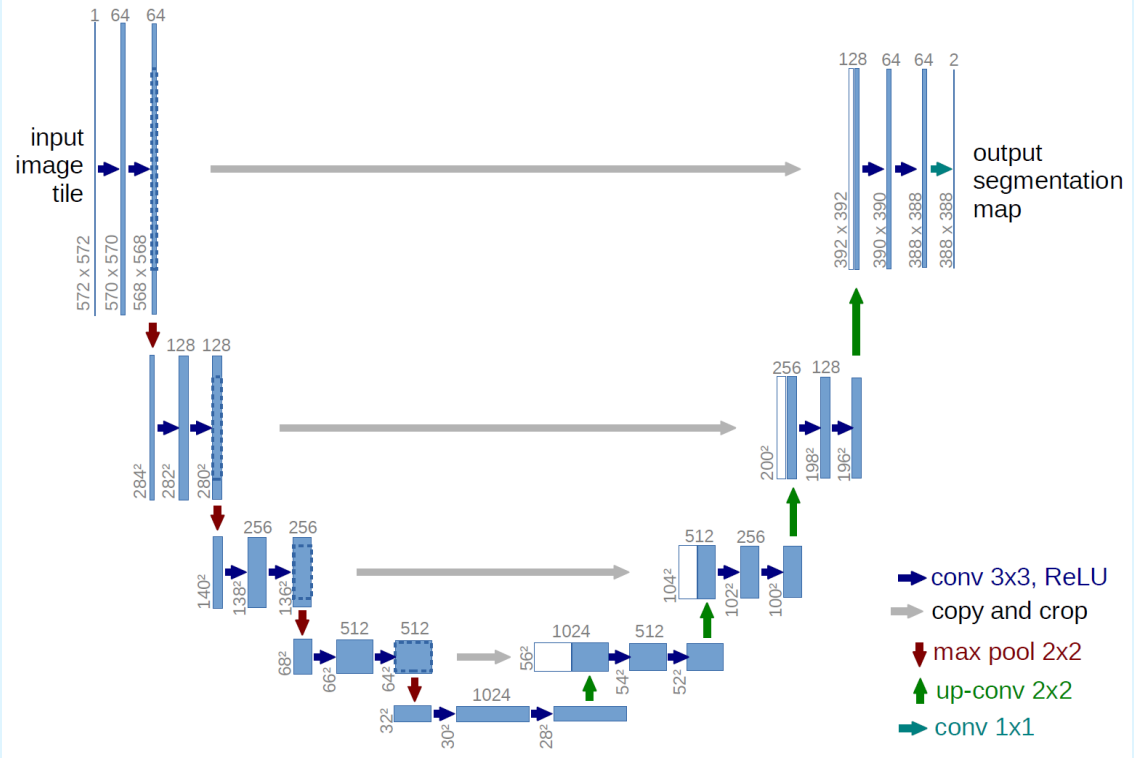


Fig. 7: U-Net Architecture[9]

Experimentation Details

Experimentation Details

- The model was Pre-trained on Binary tree segmentation dataset for 30 Epochs and on the main dataset for 50 Epochs.
- For training, Automatic Mixed Precision was used to improve performance and reduce memory usage.
- An NVIDIA P100 GPU provided by Kaggle was used for training.
- The loss-function used in Binary cross entropy with logits loss.
- Optimizer used is the Adam Optimizer.
- The learning rate schedule for pre-training and training can be seen in Fig. 8a and 8b.
- Data Augmentations were applied using the albumentations library.

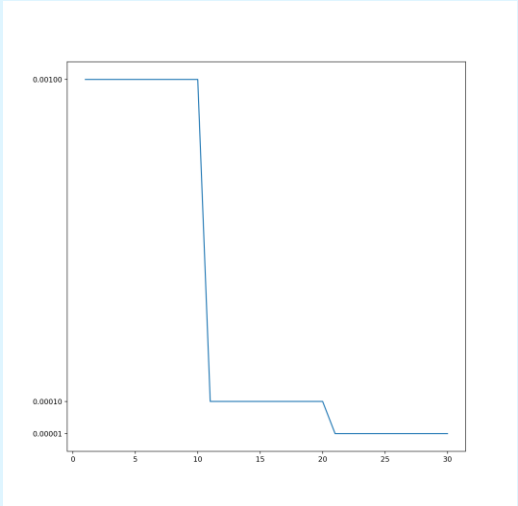


Fig. 8a: Learning Rate Schedule for pre-training

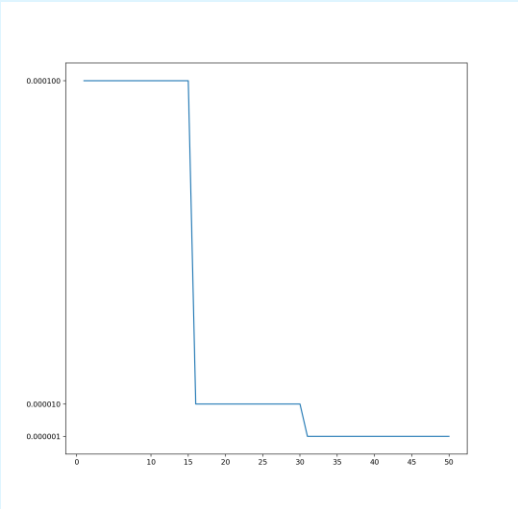


Fig. 8b: Learning Rate Schedule for training

Results and Discussions

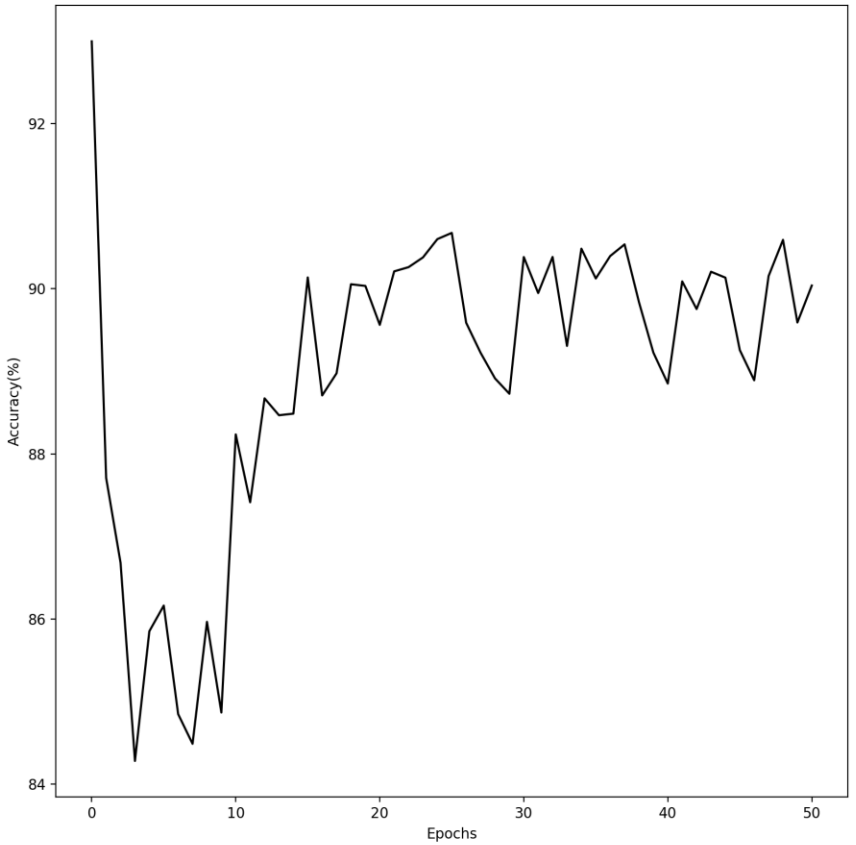


Fig. 9a: Epochs Vs Accuracy

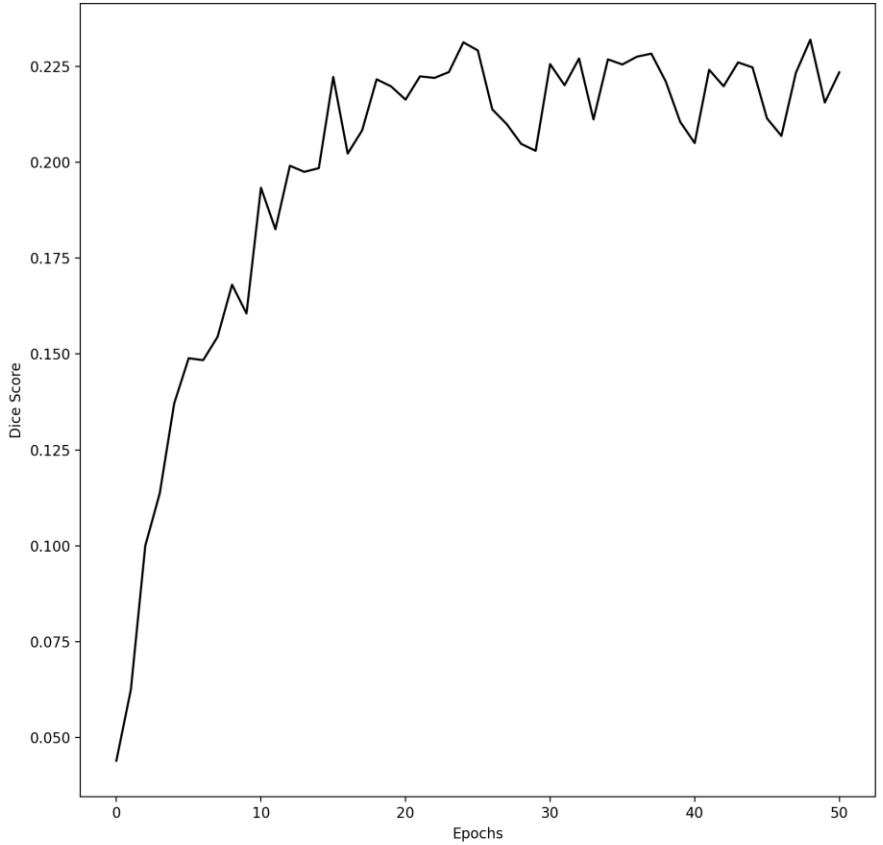


Fig. 9a: Epochs Vs Dice Score

Results and Discussions

- Model achieved 90.0394% accuracy and a dice score of 0.223
- We observed that the accuracy and dice score increases as we reduce the learning rate
- Accuracy is misleading due to unbalanced dataset. Therefore, considering only dice score as a metric.
- We see that initially the accuracy was high as our model was only predicting a blank image.
- Performance of the model can be further increased by preparing a larger dataset and a better labelling strategy.
- Use of weighted binary cross-entropy loss was important, since we had an unbalanced dataset.

Conclusion and Future Scope

- The segmentation performed using U-Net was successful.
- Further experimentation can be done by using modern segmentation models like other models like SegFormer
- Dataset can be further improved by increasing the total images.
- Some pre-processing steps like saliency detection can be explored
- Attention mechanism can be embedded in the model to improve the generalization performance.
- Further, the segmentation features learnt can be used for Object Detection as well by training a detection head.

References

1. https://github.com/VisDrone/VisDrone-Dataset

2. Roshanka Ranasinghe, Assessing climate change impacts on open sandy coasts: A review, Earth-Science Reviews, Volume 160, 2016, Pages 320-332, ISSN 0012-8252, https://doi.org/10.1016/j.earscirev.2016.07.011.

3. R. Girshick, “Fast r-cnn,” in 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448.

4. Joseph Redmon, Santosh Divvala, Ross Girshick, & Ali Farhadi. (2016). You Only Look Once: Unified, Real-Time Object Detection.

5. Zhi Tian et al., "FCOS: Fully Convolutional One-Stage Object Detection," 2019.

6. X. Zhu, S. Lyu, X. Wang and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 2778-2788

7. https://dronestock.com/

8. https://labelstud.io/

9. Olaf Ronneberger, , Philipp Fischer, Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation." (2015).

THANK YOU