Safely-Sail AI

Machine Learning & Computer Vision Based Sailing Safety System

(Clement) Tong Yuk Chun1,\*

1Harrow International School Hong Kong

\*Contact: [clement.yc.tong@gmail.com](mailto:clement.yc.tong@gmail.com)

Abstract— This project presents a machine learning based AI Sailing Safety Monitoring and Analysis system, which is designed to detect the dinghy’s occupants, configurations, and its surroundings and report any developing dangers to lifeguards and sailing centres for early intervention to prevent tragic accidents. There are multiple approaches to this solution but there was a core focus on a vision based instance segmentation model, based on the Yolo-v8n architecture as a pure vision system would allow expanded capabilities beyond using it as just a safety aid.

1. Introduction

Dinghy sailing has been growing in popularity over the years, with a recent spike of interest due to the Covid-19 pandemic. Although there are many measures in place by the sailing community to ensure it is a generally safe sport, sailors are subject to ever changing weather conditions, equipment failure, and poor judgements and high-profile tragic accidents have occurred, mostly as a result of entrapments after capsize, man overboard, and sailors being blown out to sea. In particular, Man-overboard and incidents of sailors being blown out to sea are often only noticed hours or days later when friends and family realize they have failed to return, resulting in a very large search area that often renders SAR (Search and rescue) operations difficult and sometimes unsuccessful.

To help reduce the risk, this project proposes a machine learning based sailing safety system that can alert lifeguards and sailing centres to developing emergency situations so help can be sent before it is too late.

1. Implementation

The idea is to have a small, low cost, solar powered box on the deck of sailing dinghies, similar to cube satellites, or the cameras used by Olympic dinghy sailors. It can be mounted on the mast, the thwart of the boat, or behind the boat. It will house a camera, an embedded processing unit with an AI accelerator, a 4G LTE modem, batteries, a small e-ink screen, and a manual emergency SOS button. A continuous video feed will be fed from the camera and analysed in real time by various AI neural networks. If an emergency is detected, SOS mode will be triggered, and SOS signals will be sent to the relevant authorities. To combat the possibility of lost of signal if the boat capsizes, periodic ‘alive’ signals will be sent to the server every 2 seconds when active and when there is a loss of signal for more than 45 seconds, an alert would be triggered. This timeout period can be extended to avoid unnecessary triggers during normal operations such as when running capsize recovery drills under coach supervision.

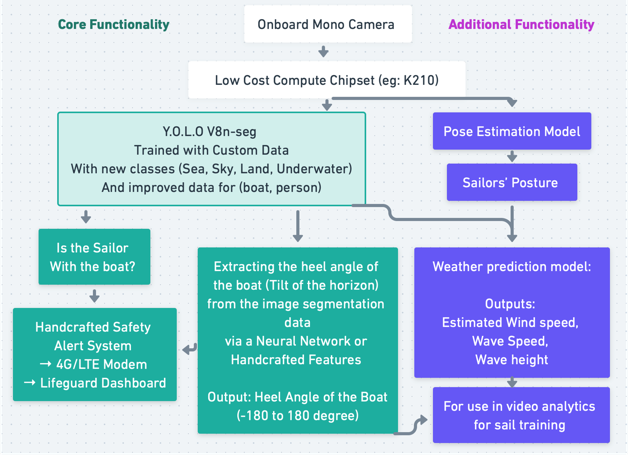
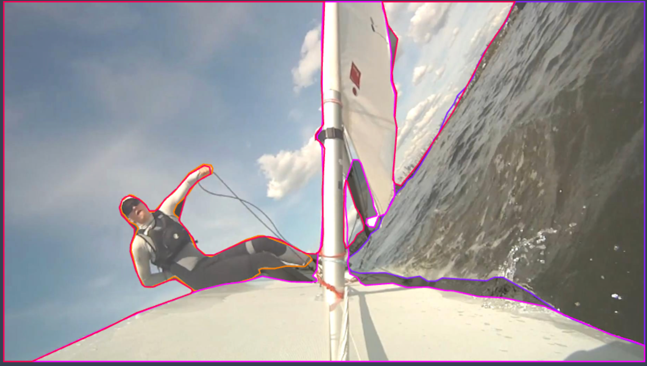


Fig. 1 A diagram showing the intended system architecture.

1. Implementation and analysis

The focus of the implementation was in implementing the vision based machine learning model. The approach chosen was to perform transfer learning on a YOLO-v8n-seg instance segmentation model with pretrained weights and a custom dataset with custom classes. The custom dataset consisted of approximately 60 fully annotated frames, taken from various videos from GoPros onboard sailing dinghies, with varying lighting, weather conditions, sea-state, type of dinghy, number of persons onboard, clothing, etc. Various techniques were used to vary the image from changing the hue, saturation and brightness, to rotating the image to attempt to improve the training results. The resuling mAP50 was 0.681 for the bounding box predictions and 0.499 for the segmentation predictions on the 8 image validation dataset.

A picture containing diagram

Description automatically generated

Fig. 2,3,4,5 Examples of labelled data

The model (yolov8n architecture) achieved the following performances on an M1 CPU (No hardware acceleration used) using an input of size 384x640:

|  |  |
| --- | --- |
| Precision | FPS (approx.) |
| FP32 | 20 |
| FP16 | 20 |

The weights have a size of 6.7MB (Unquantized)

If the model is quantized to INT8 and ran on NPUs/GPUs, it will require less computation time per frame and result in better performance even on low end hardware (eg: K210).

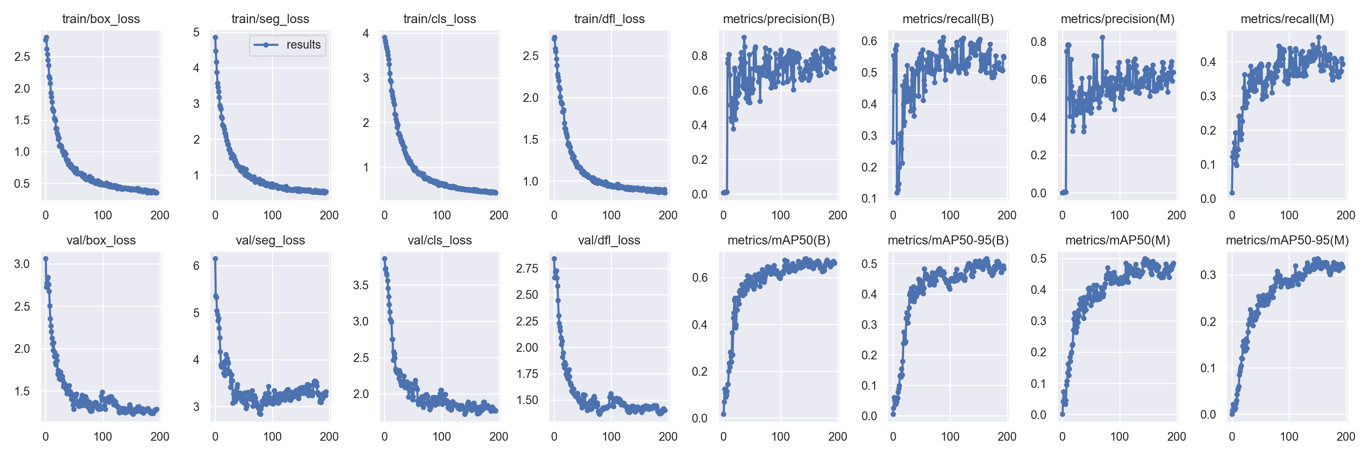


Fig. 2 Training Data from training v2 of the model (Transfer Learning)

1. Conclusion

Although over 1600 frames were actually taken from 11 publicly available videos online, with many more other video sources yet to have used, there was a big challenge in annotating segmentation data as the polygons of the outline of the desired regions (person, sea, sky, boat, etc.) have many points (usually over 20 points for each object) and even though tools like the “smart polygon” by roboflow, a similar tool by v7, and other image annotation tools like labelimg, labelstudio, RectLabel, and more were used, with some attempts with ML assisted labelling with existing YOLO models, it was not of great help as there were no existing models that performs sea-sky segmentation available for use, the 2 most time consuming parts. Hence, labelling each image took around 120 seconds each on average.

The v2 version of the model as shown above used a dataset of approximately 60 labelled images. The v3 dataset has been prepared since 2022-03 and training for the v3 model is underway. The v3 dataset involves 109 images, and over 200 images after image augmentation techniques have been applied. The v3 dataset also includes frames from more video sources, including first hand video sources.

Acknowledgment

Thank you, Dr. Ray C. C. Cheung and his teaching assistants, for the guidance and support during the GEF program.