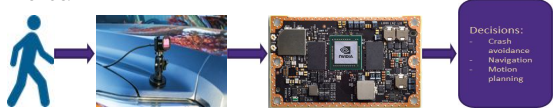


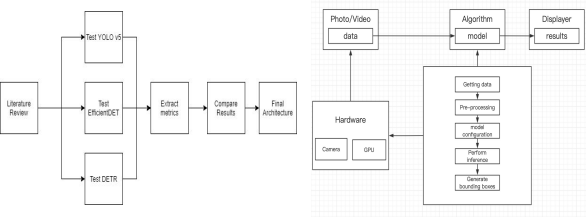
Objectives and Concept

- Create and deploy an algorithm to classify different classes of vulnerable road users (pedestrians, cyclists) from a thermal imaging camera.
- Develop a neural network to both classify and localize classes in real time.
- Deploy and validate efficacy of using thermal image data on an autonomous vehicle.



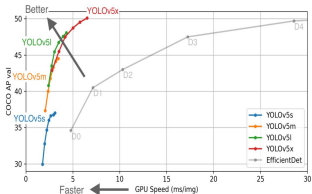
System Implementation

- The three models to compare for designing the final model were YOLO v5, Efficient DET and DETR as speed and accuracy of the model were the priorities.
- The software consists of the object detection model as well as the real-time streaming and inference pipeline.
- The hardware chosen for model deployment were Nvidia Jetson TX2 Xavier NX and Forward Looking infrared (FLIR) camera to capture thermal images.
- Requirements:
 - The system shall identify pedestrians, cyclists, and vehicles
 - The system shall localize classes with a bounding box
 - The system shall be deployed on a hardware accelerator
 - The system should beat FLIR mean average precision (mAP) benchmarks

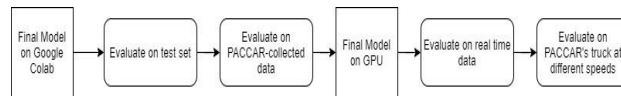


Model Exploration

- The implementations each have different strengths accuracy, efficiency and speed.
- Compared with the results of training on FLIR 128 dataset of these three models, the mAP value of YOLO V5 is the highest, while the DETR value is the lowest.
- The models showed best performance YOLO v5 on cyclists, EfficientDET on cars and DETR on pedestrians..



Final Model Implementation and Deployment



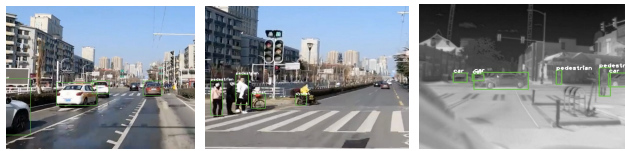
- The two priorities for the model were to have accuracy and speed, in that order. This is not only to maximize performance in a real time setting, but to ensure safety.
- The final model adapts YOLOv5 to be more effective on thermal data with:
 - Filter structure in neural network layers (feature dependent)
 - Hybrid training/transfer learning
 - Hyperparameter tuning
 - Threshold manipulation
- Deployed on NVIDIA Jetson Xavier NX for real time inference.

Results and Conclusions

- The thermal video feed extracts the key features out that are common to all humans regardless of what they are wearing, i.e. body heat. It performs better than RGB.
- Our model beats FLIR's benchmarks across all different classes individually and even when averaged together by 5+%. These metrics were generated by testing on a validation set from the database.
- Although performance is not yet at production level, the results were also robust with the vehicle moving at higher speeds, detecting pedestrians as would be necessary in production and demonstrating the advantage of thermal imaging detection schemes.

Mean Average Precision

Model	People	Bicycles	Car	All
FLIR	0.794	0.580	0.856	0.743
This work	0.84	0.632	0.906	0.793



Future Work, References, and Acknowledgments

- The future work includes further hardware optimization to take towards production and resolving software dependencies to run more effectively in real time.
- Reference:
 - Joseph Redmon. 2015. You Only Look Once: Unified, Real-Time Object Detection.DOI: <https://arxiv.org/abs/1506.02640>
 - EfficientDET:<https://ai.googleblog.com/2020/04/efficientdet-towards-scalable-and.html>
- The team would like to sincerely thank the following individuals for their mentorship throughout the course of the project.
 - Industry Mentor : James Nelson
 - Faculty Mentor : Prof Ming Ting Sun
 - Prof Payman Arabshahi
 - TA : Daniel King

