Hello, my name is Nicholas Bandy. This presentation is for a research proposal on the following topic: Predictive Pedestrian Detection in Autonomous Vehicles: Integrating 3D Human Pose Estimation for Predicting Pedestrian Movements

The focus for this project is on answering the following research question. How can 3D pose estimation be applied to pedestrian detection systems to predict pedestrian movements?

Developing increasingly reliable pedestrian detection systems is important in developing a reliable autonomous vehicle. Wang et. Al 2023 discusses the unpredictable nature of pedestrians, and the challenges in determining where they will move next. Addressing this research question can improve the ability of autonomous vehicles to predict a pedestrians movement. This would help to reduce the number of pedestrian collisions, which Brar and Caulfield (2017) found to be a leading concern for the public. Addressing this would increase the overall reliability and reputation of autonomous vehicles.

A primary objective of this project is to show how a pedestrian detection system can be enhanced to provide better performance through the integration of 3D pose estimation. A predictive model utilizing the results of 3D pose estimation can be developed to give the system an indication of whether or not a pedestrian is planning to cross the road, which will serve as an early indication for the vehicle on whether or not it needs to slow down and prepare to stop. Finally, the goal is to assess the performance of this predictive model in simulation to determine if it would be suitable to implement in real-world scenarios.

Previous attempts to implement pose estimation with autonomous vehicles were made by Zheng et. Al in 2021 and Bauer et. Al in 2023, who both developed a weakly supervised multi modal 3D pose estimation model, which they found promising results in estimating human pose for autonomous vehicles using a combination of both lidar and camera technologies. While both of these models successful in estimating the pose of a pedestrian, they were not further developed to use these pose estimations to make predictions on the pedestrians movement. That is the primary objective of this project, to build on these studies by applying the 3D pose estimation results to predict the pedestrians movement.

The Pedestrian Estimation Dataset from Rasouli et. Al (2019) will be used as the primary dataset for this project. This dataset contains real traffic footage of common traffic scenarios from an onboard camera, with labelled pedestrian intent to either cross the road, or stay on the side they are currently on. This data set is ideal for this project, as this is the exact behaviour we want to use 3D pose estimation to predict.

To obtain 3D pose estimation in real time, MediaPipe BlazePose from Blazarevsky et. Al (2020) will be used. BlazePose has shown the capabilities of providing real time 3D pose estimation that is required for this project, and removes the need to create a new 3D Pose estimation model. This saves us from the unnecessary work of replicating the work of Zheng et. Al 2022 and Bauer et. Al 2023, who have already shown that 3D pose estimation can work with the technologies onboard an autonomous vehicle. BlazePose provides both a full model and a lightweight model that can be useful in ensuring the real-time speeds required for integration into the pedestrian detection system are met. Deploying an effective system for 3D pose estimation in autonomous vehicles faces the challenge of balancing accuracy and speed. Yang et. Al 2022 discusses the trade off between accuracy and speed in these models, which can be addressed by utilizing the lightweight model provided by BlazePose. Figure 1 shows a sample of the key point detection results from BlazePose as published by Blazarevsky et. Al 2020.

By using BlazePose to obtain 3D pose estimation on the Pedestrian estimation dataset, key information such as the angle of bend in each knee, relative position of the feet, and which direction the shoulders, hips, and head are facing can be extracted, as these serve as strong indicators of a persons current and potential movement patterns. The distance between key points can be normalized based on calculating the persons height as done by Fang and Lopez (2018) to provide value in calculating relative distances between key points. This is shown in Figure 2, which is a sample image taken from Fang and Lopez 2018, highlighting how this normalization based on height is done. Using all of the calculated information, a model can be trained on the Pedestrian Estimation Dataset to determine what poses are associated with a pedestrian that wants to step out onto the road.

The model will be evaluated based on the accuracy and precision of predicting whether a pedestrian is going to cross, or stay on their side of the road. Minimizing the number of false negative results, where the model predicts a pedestrian will stay when they actually want to cross is key in achieving the primary objective of reducing the number of pedestrian collisions through early detection. To be ready for real-world implementation, this must not come at the cost of overall accuracy as the model will not be useful if it constantly causes the vehicle to slow for pedestrians that have no intention of crossing the road.

After this project is complete, the final result is a predictive network that will take a pedestrians current pose, calculate key metrics that give indications of a pedestrians movement, and use this information to estimate if they are likely to try and cross the road or if they will stay on their current side.

This project uses an open sourced public dataset that does not contain any personal information to train and test the model. Further exploration into the dataset during the data exploration stage of this project to ensure the dataset contains adequate diversity in pedestrians is important to ensuring there is no bias introduced in the final model. Any additional data gathered for training or testing purposes needs to include the proper permission from the party involved, and should not include any personal or identifying information. It’s also important to ensure proper diversity in any additional collected data, to ensure that the prediction model works on all pedestrians and not just a specific subset or group of people. All data gathered or used in this project must be diverse to ensure the trained model does not gain bias.

The proposed timeline for this project is shown in Figure 3, with 8 months proposed to finalize the project so that it can be ready for presentation In month 9. The first stage of the project is to complete the literature review, which will happen in the first month of the project. The literature review will help to gain a deeper understanding of the research topic and explore all of the related works beyond the scope of this proposal. The next step is to explore and prepare the datasets, which should last around one month in month 2 of this project. As the planned datasets are open sourced and publicly available it will be easy to acquire, allowing for more time to explore the dataset during this period. If the literature review or data exploration leads to the conclusion that more data is required, It should be collected and explored during this period. Next development begins on the predictive model, which is expected to last two months and take place in months 3 and 4 of the project. Initial calculations for the key determinants from the pose estimation results will be established, and an initial model for making the predictions will be trained and tested. After the initial model is complete, the focus will be on model refinement. This should last around 2 months during months 5 and 6 of the project. Modifying hyperparameters, calculations for key determinants, and ensuring the model performs its best in terms of both speed and accuracy occurs during this period. Once the model has been finalized, a two month period during months 7 and 8 will be used to ensure that the research is fully documented, and develop a report which analyzes the final results. Finally, the last month in month 9 of the project is used to present and share the results with my colleagues.

Thank you for listening to my presentation, I look forward to further research and discussions on this topic.

References:

Bauer, P., Bouazizi, A., Kressel, U., Flohr, F. (2023) "Weakly Supervised Multi-Modal 3D Human Body Pose Estimation for Autonomous Driving," Available From https://ieeexplore-ieee-org.uniessexlib.idm.oclc.org/document/10186575

Bavarevsky, V., Grishchenko, I., Raveedran, K., Zhu, T., Zhang, F., Grundmann, M. (2020). BlazePose:On-device Real-time Body Pose tracking. Available From: https://arxiv.org/abs/2006.10204

Brar, J.S. and Caulfield, B.. (2017). Impact of autonomous vehicles on pedestrians' safety. Available from https://doi.org/10.1109/itsc.2017.8317963.

Fang, Z., Lopez, A. (2018) Is the Pedestrian going to Cross? Answering by 2D Pose Estimation. Available From: https://doi.org/10.1109/ivs.2018.8500413.

Rasouli, A., Kotseruba, I., Kunic, T., Tsotsos, J. (2019). PIE: A Large-Scale Dataset and Models for Pedestrian Intention Estimation and Trajectory Prediction. International Conference on Computer Vision. Available From: https://data.nvision2.eecs.yorku.ca/PIE\_dataset/

Wang, H., Jin, L., He, Y., Huo, Z., Wang, G., Sun, X. (2023). Detector–Tracker Integration Framework for Autonomous Vehicles Pedestrian Tracking. Remote Sensing, 15 (8), 2088. Available from https://doi.org/10.3390/rs15082088.

Yang, S., Chen, Z., Ma, X., Zong, X., Feng, Z. (2022). Real-time high-precision pedestrian tracking: a detection–tracking–correction strategy based on improved SSD and Cascade R-CNN. Journal of Real-Time Image Processing, 19 (2), 287–302. Available from https://doi.org/10.1007/s11554-021-01183-y.

Zheng, J., Shi, X., Gorban, A., Mao, J., Yang, S., Qi, C, Liu, T., Chari, V., Cornman, A., Zhou, Y., Li, C., Anguelov, D. (2021) Multi-modal 3D Human Pose Estimation with 2D Weak Supervision in Autonomous Driving. Available From: https://arxiv.org/abs/2112.12141