85-412 Project Report
Gustavo Silvera

Building A Driving Model From Recorded Data In Simulation

Driving is a complex activity where safety is highly dependent on driver attention and awareness. It is no surprise then that driver inattention continues to be a major cause of accidents since the early twentieth century [9]. As fully autonomous vehicles are still out of reach for the foreseeable future, it is important to study the internal state of the human driver to optimize driver safety and efficiency features. Driver monitoring technologies that use implicit cues like driver eye gaze could improve safety and effectiveness by inferring situational awareness (SA) and driver intention to estimate driver behaviour and build a driving model based on human experience and behaviour. For this project, I will be leveraging a system built for monitoring external and internal driving behaviour to build and train a model that can use internal and external driving cues to estimate vehicle inputs.

To further motivate the project idea, an accurate real-time driving model could be used to enhance safety features, reduce direction confusion and distraction, and provide more effective driver assistance and guidance. By knowing what the human driver's intentions should be given the internal and external situation, the model could correct for faulty inputs and aid in risk avoidance, resulting in improved safety/convenience. This system could provide improved vehicle operation for drivers with difficulty making inputs themselves, such as those with motor impairments. Additionally, a driving model could be used to improve driving education by acting as a teacher and correcting mistakes for student drivers. There is no shortage of modern uses that a human-centric driving model can be applied to.

The objective of this project is to create a cognitively plausible driving model for my synthetic simulated environment by training the primary input mechanisms against recorded human driver data. Contrary to previous works [7, 15, 16] that trained via simulation in ACT-R [1] alone, this project aims to utilize recorded data from a human participant to train from. This has several benefits, notably that the subsymbolic logic for determining exact magnitudes of the analog driving inputs is a learned parameter from the participant data, hence no programmer input is used to bias the results. Additionally, the model may be able to capture subtle driving features from the data that programmers would have difficulty implementing, such as how driver eye gaze relates to intention. Ideally, the final model will be able to predict the steering, throttle, and brake inputs to the vehicle from the simulation just as the human did.

This project will use the simulated environment from CARLA [10], a 3D open source driving simulator for autonomous driving research, but modified for human-in-the-loop behavioural and interactions research. The modified simulator is known as DReyeVR [5] and is designed for inserting human drivers into the CARLA simulator in virtual reality (VR) and running experiments with the human participant. This system allows for a significant amount of real-time simulation data collection, including eye data from the eye tracker built-in to the VR headset [11].

The eye-tracker module in DReyeVR contains an interface to access several key variables such as (left, right, & combined) eye gaze & origin vectors, pupil position, pupil diameter, eye openness, and associated timestamps. The eye trackers are located inside the HTC Vive Eye Pro VR headset [11] so they will be accurate and non disruptive to the overall user experience, minimizing extraneous cognitive load [8]. This

is a useful signal as it allows researchers to investigate relationships between driving behaviour and driver eye data. There has previously been work on eye gaze and how it relates to mental workload during cognitive tasks [8], but this work specializes in driving behaviour similar to [9, 13].

In sum, the collected data includes external simulator data, such as ego-vehicle position, velocity, orientation, other actors, etc. and internal sensor data such as the participants eye data, head position/orientation, and control inputs.

Prior to training the driving model, I held several hypotheses relating the physiological eye data and expected driving inputs that I hoped to see represented with the model. For steering, I believed there would be a strong correlation between magnitude (and direction) and lateral eye gaze and head pose (ie. “looking” direction). This comes from natural intuition that drivers typically look in the direction they intend to steer prior to performing the maneuver. For throttle, I would assume that a forward facing looking direction would be correlated with higher velocities (more throttle), since humans typically drive in a longitudinal fashion. For braking, I also thought there could be some correlation with pupil diameter, since applying the brakes to decelerate could be related to lower cognitive load as the vehicle is slowing down and the driver can spend more time learning about the environment. I lean on pupil diameter as a signal for cognitive load as demonstrated in prior works [3, 4, 5] that the two are often linked.

For the overall driving model, my design entails three distinct neural networks each with the purpose of computing steering, throttle, and braking values independently. I decided on using standard neural networks for these models (with some ReLU’s) to take advantage of recent temporal dependencies with the recorded input time-series

data. Another option could have been to use Long-Short-Term-Memory (LSTM) modules [13] which are known to utilize long-term temporal dependencies [13] but this technique provides unnecessary flexibility with increased complexity. Since the standard neural networks sufficiently capture nearby temporal correlations [13] and were feasible to train on my laptop, I elected to go this route.

Another important function of this project is to explore the logic behind the symbolic-subsymbolic representation of the interaction between simultaneous driving inputs. Specifically, if we consider the three inputs (steering, throttle, and brake) as symbols that the model is aware of (encoded by the programmer), then it also needs a high-level mechanism to define the combination of the three symbols that follows standard driving behaviour. For instance, while the model learns the subsymbolic representations of each input independently via training, the programmer includes logic for “driving rules” that would be hard to learn without enough data. As an example, the model might not realize that the throttle and braking inputs must be clamped between 0 (pedal is fully released) and 1 (pedal is fully depressed), hence negative and large numbers are invalid. Additionally, as we humans know from our training and physical disposition, pressing the brakes and throttle simultaneously is not only difficult to do, but also damages our vehicle, wastes energy, and conflicts with our driving education [14]. Since it is unlikely this logic will be learned during the training process, it suffices to manually add this logic to the output pass once the subsymbolic components (input magnitudes) are computed.

85-412 Project Report
Gustavo Silvera

In order to begin training the model I needed to collect driving data from a human participant in the simulator. This data came from a user study where the participant (their name was anonymized to “Jacob”) was told to follow three separate routes with various driving situations such as 4-way intersections, highway keeping, and suburban travel. Once the data on the three routes was collected, a total of ~10 minutes of data at ~30hz was combined. For these recordings, only data with 100% sensor validity was kept, enforcing that both eyes were open. Finally, there was some light smoothing to the noisy data since the eye trackers demonstrated some variance during normal operation. Unfortunately, the data was not captured at a constant tick-rate since it depended on the variable frame rate of the simulator. However, I found that typically the data’s capture-rate was stable enough for analysis.

After performing the necessary implementation (using Pytorch) to build the model and training mechanisms, the individual models were built as follows:

Model	**Parameters**
Steering	- Network layers: 1→64→128→256→256→256→1 - Loss Function: Mean Squared Error - Number of Epochs: 50 - Learning Rate: 0.001 - Optimizer: Adam
Throttle	- Network layers: 1→128→256→ReLU→256→1 - Loss Function: L1 Loss - Number of Epochs: 50 - Learning Rate: 0.001 - Optimizer: Adagrad
Brake	- Network layers: 1→128→256→256→256→ReLU→256→1 - Loss Function: L1 Loss - Number of Epochs: 50 - Learning Rate: 0.01 - Optimizer: Adagrad

Most of the model parameters were obtained via trial-and-error experimentation to find a decent parametrization that resulted in decreased loss over epochs. Each of the individual input models were given the same set of parameters (external and internal sensor data) as well as the predicted input values of the other two inputs. This mimics a “memory” feature where historical model predictions from a previous iteration are used as inference parameters for the current iteration, reflecting the short-term temporal dependencies mentioned earlier.

It is worthwhile to note that unlike the steering data, the throttle and brake (Figure 1) data is generally more discontinuous and “spiky”. This leads to difficulties when performing regression as outliers and anomalies have a large effect on the model. To combat this I found that using the L1Loss (Mean absolute error) loss criteria function was more effective than L2Loss (Mean squared error) since L1Loss is more robust to outliers. This reduced the training time and enabled convergence at <50 epochs.

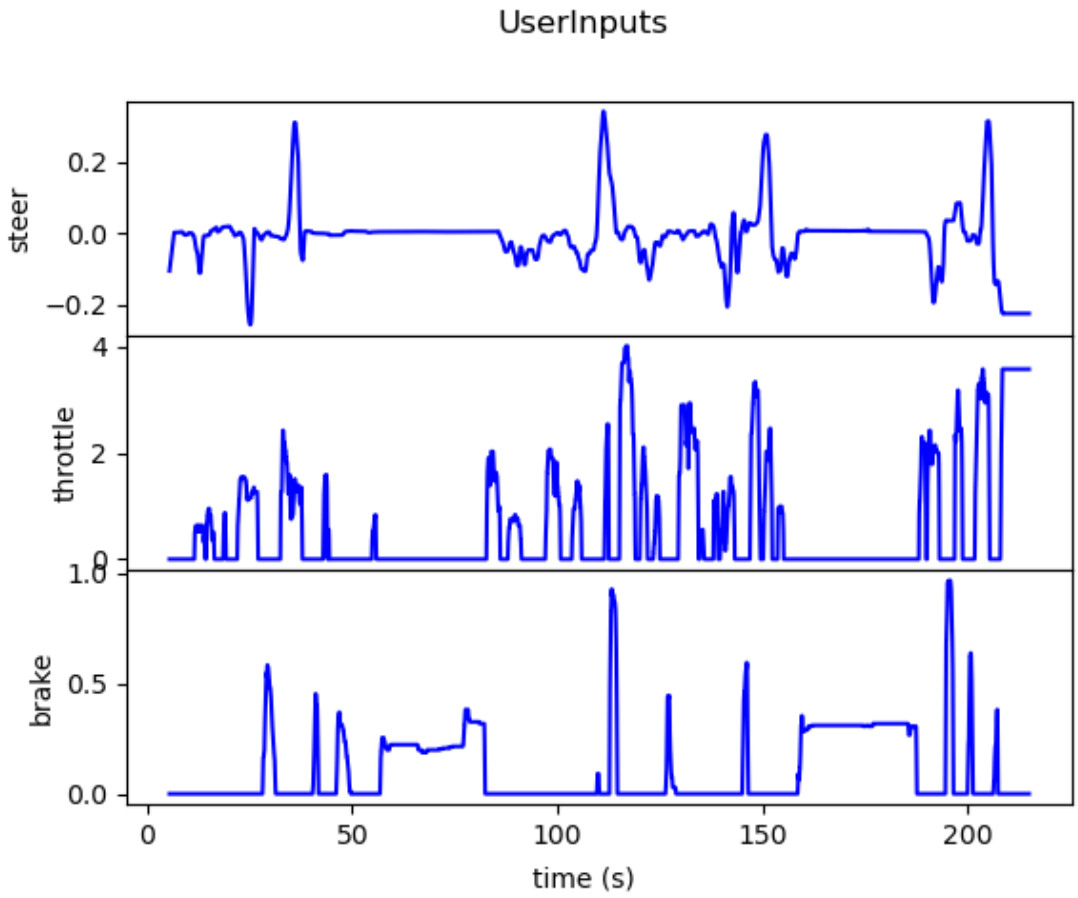

Figure 1: Demonstration Of Steering, Throttle, And Brake Inputs From Jacob54.txt

Once the training for the three models was completed:

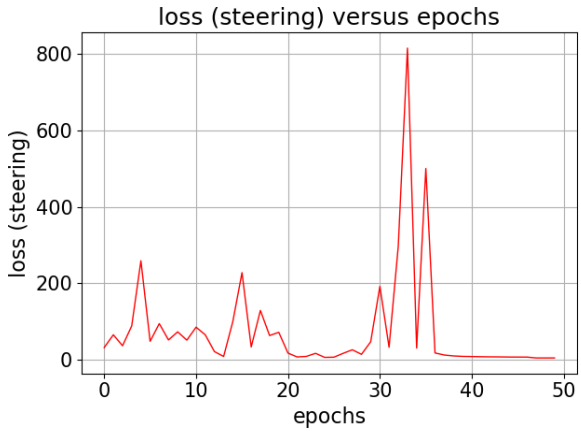

Figure 2.1: Loss Per Epoch For Steering Model Training

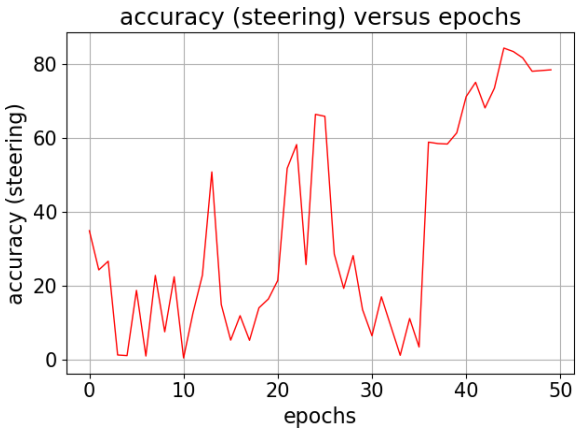

Figure 2.2: Accuracy Per Epoch For Steering Model Training

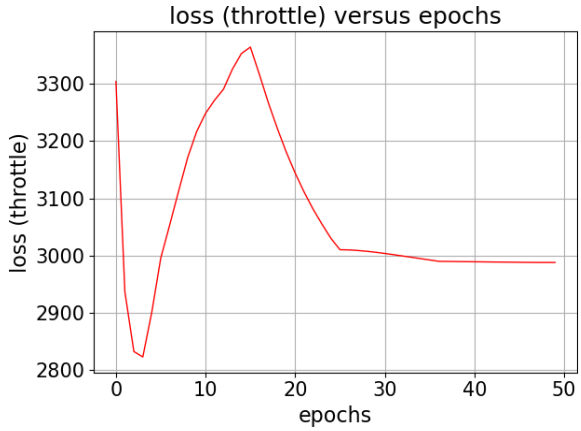

Figure 3.1: Loss Per Epoch For Throttle Model Training

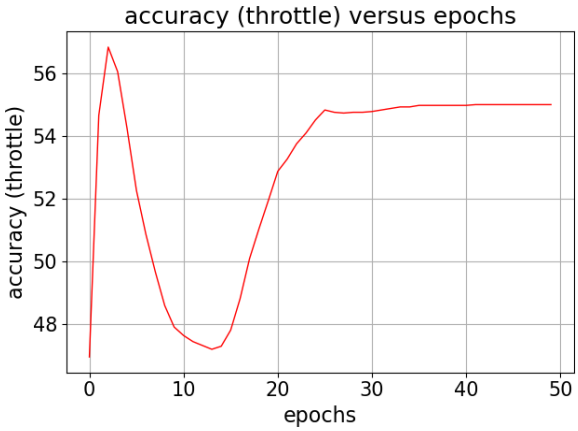

Figure 3.2: Accuracy Per Epoch For Throttle Model Training

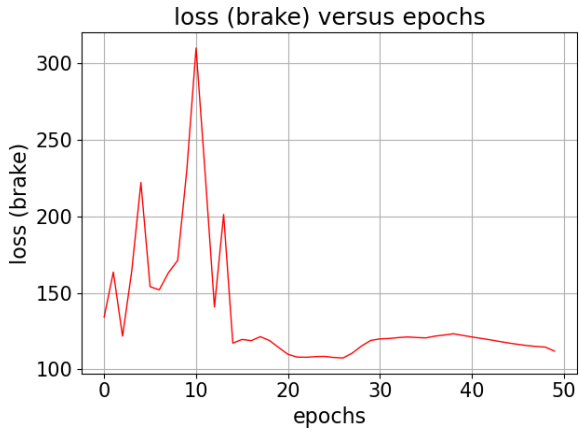

Figure 4.1: Loss Per Epoch For Brake Model Training

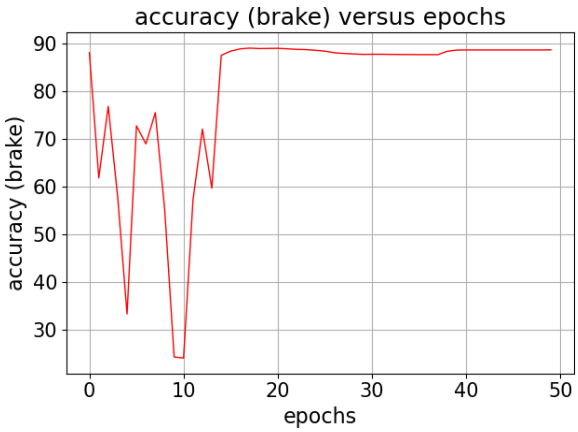

Figure 4.2: Accuracy Per Epoch For Brake Model Training

With sufficient epochs, the training for each of the models resulted in overall steady loss decrease and accuracy increase or stagnation.

Additionally, by utilizing some integrated gradient techniques [12] for feature importance, the following average feature importance plots were generated:

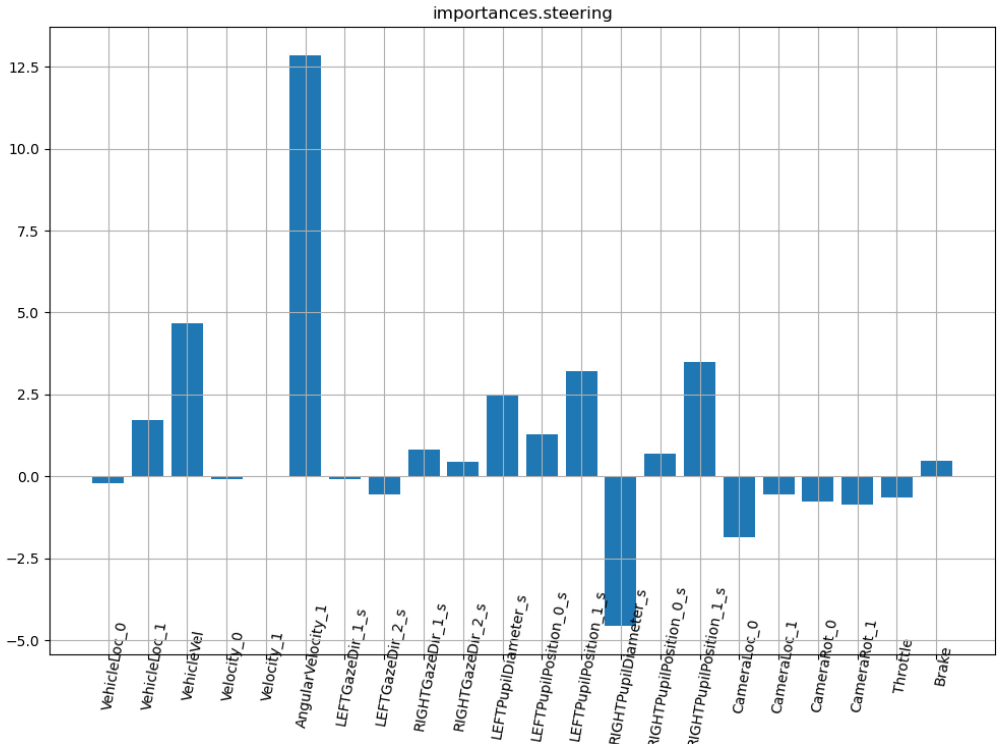

Figure 5: Average Feature Importances For Steering Model

The feature importance plot (Figure 5) for the steering model highlights how there is a significant positive correlation with left and right lateral pupil position as predicted. For the internal data features (human data) these have the largest magnitude, whereas for the external data, the largest correlations come from vehicle velocity and angular velocity. This is expected as externally, the steering input is primarily what drives the

angular acceleration of a standard vehicle. It is also interesting to see how the right pupil diameter has a strong negative correlation with steering input, suggesting a matching negative correlation to cognitive load and steering magnitude.

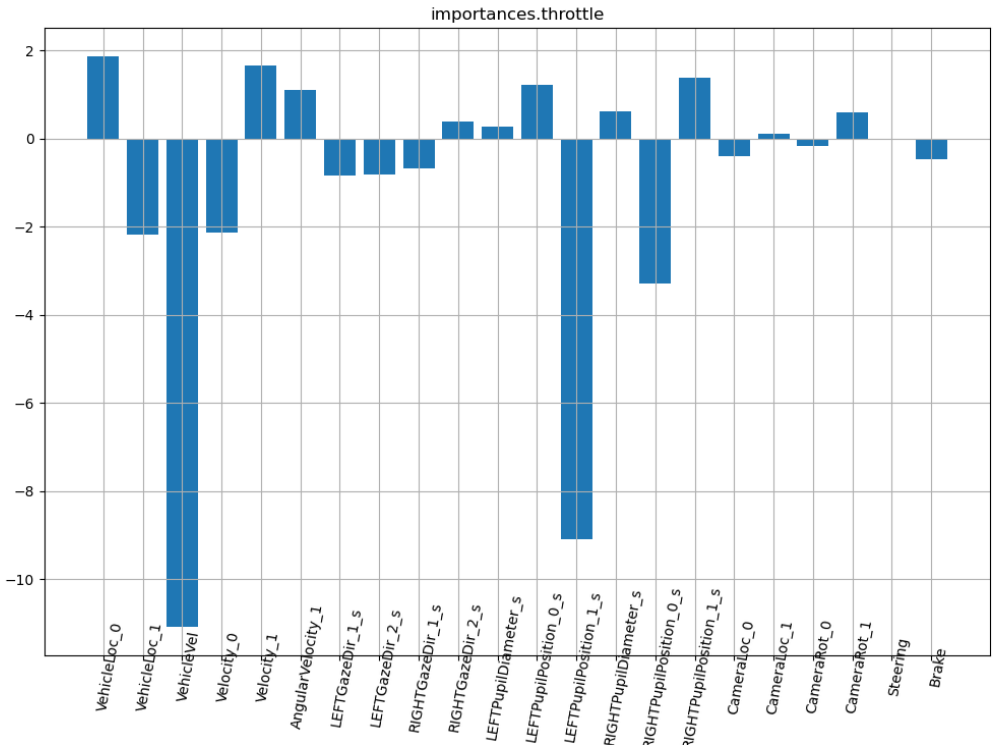

Figure 6: Average Feature Importances For Throttle Model

The feature importance plot (Figure 6) for the throttle model highlights how there is a significant negative correlation with the vehicle velocity. This can be explained by intuition as typically when driving at high speeds, the throttle is not pressed very far since there is no need to further accelerate. On the other hand, acceleration (high throttle) is necessary when operating at low speeds in order to begin driving. It is interesting to see that there is a negative correlation with left lateral pupil position but

not right lateral pupil position, this could potentially suggest a lack of sufficient training or data. Overall, there also seems to be low correlations with the other gaze features, suggesting that those feature importances are not necessarily skewed in one particular direction and can easily bounce back and forth.

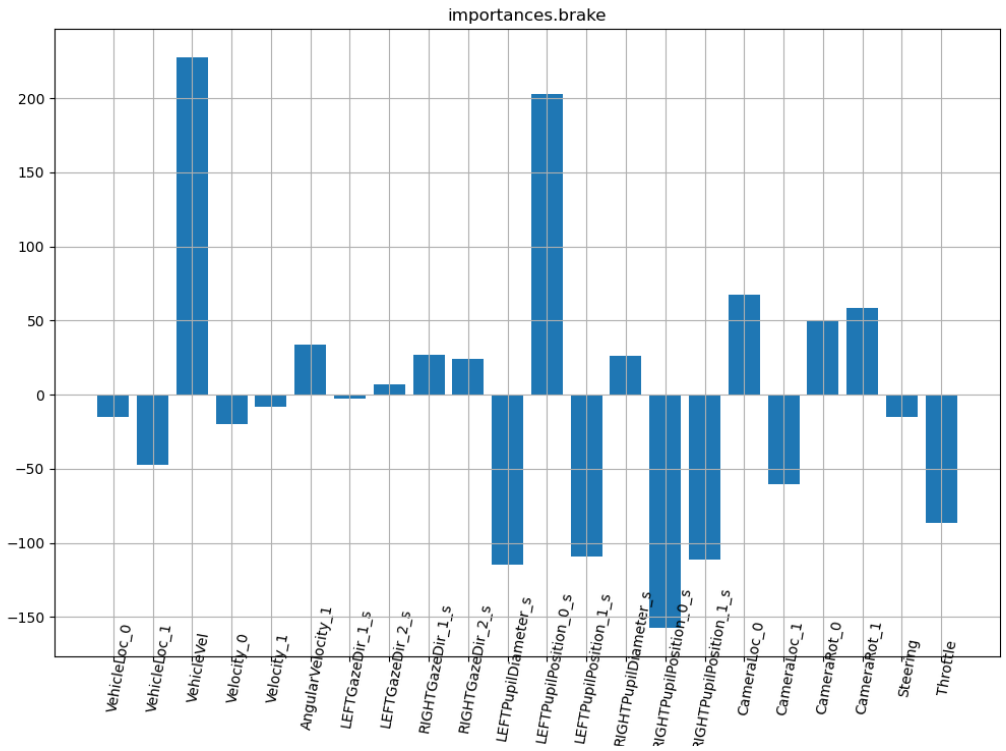

Figure 7: Average Feature Importances For Brake Model

The feature importance plot (Figure 7) for the brake model highlights how vehicle velocity and vertical pupil position both have a positive correlation with the brake input. Mirroring the throttle model case, when operating at a high velocity drivers are more prone to apply the brakes at a greater magnitude, which explains the large positive correlation with velocity. However, the internal driver features describe a

relationship where the lateral pupil positions demonstrate a strong negative correlation and the vertical pupil positions demonstrate a mixed positive and negative correlation. I believe this could be explained by the model recognizing that drivers often look around in rapid movements (saccades) when braking to fill gaps in their situational awareness.

Finally, when combining the resulting predictions from the three models as part of the overall DrivingModel, the following plots are generated:

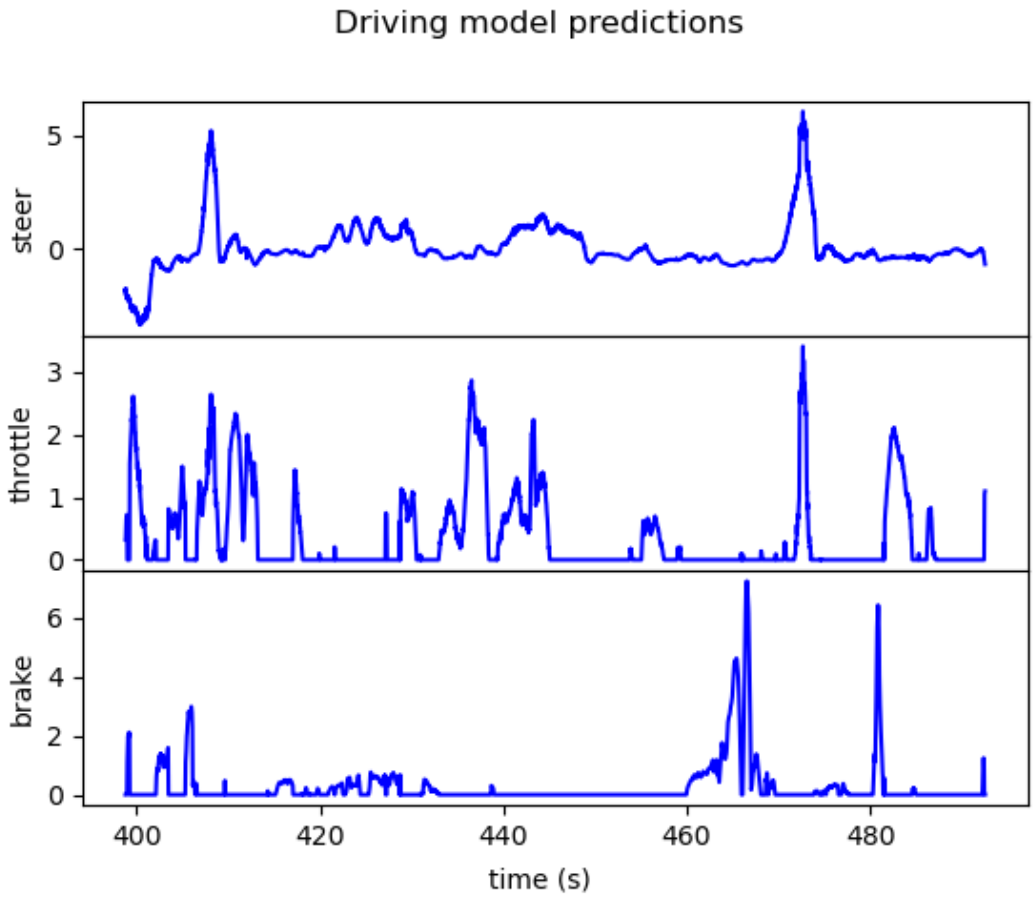

Figure 8: Final Driving Model Predictions For Test Data

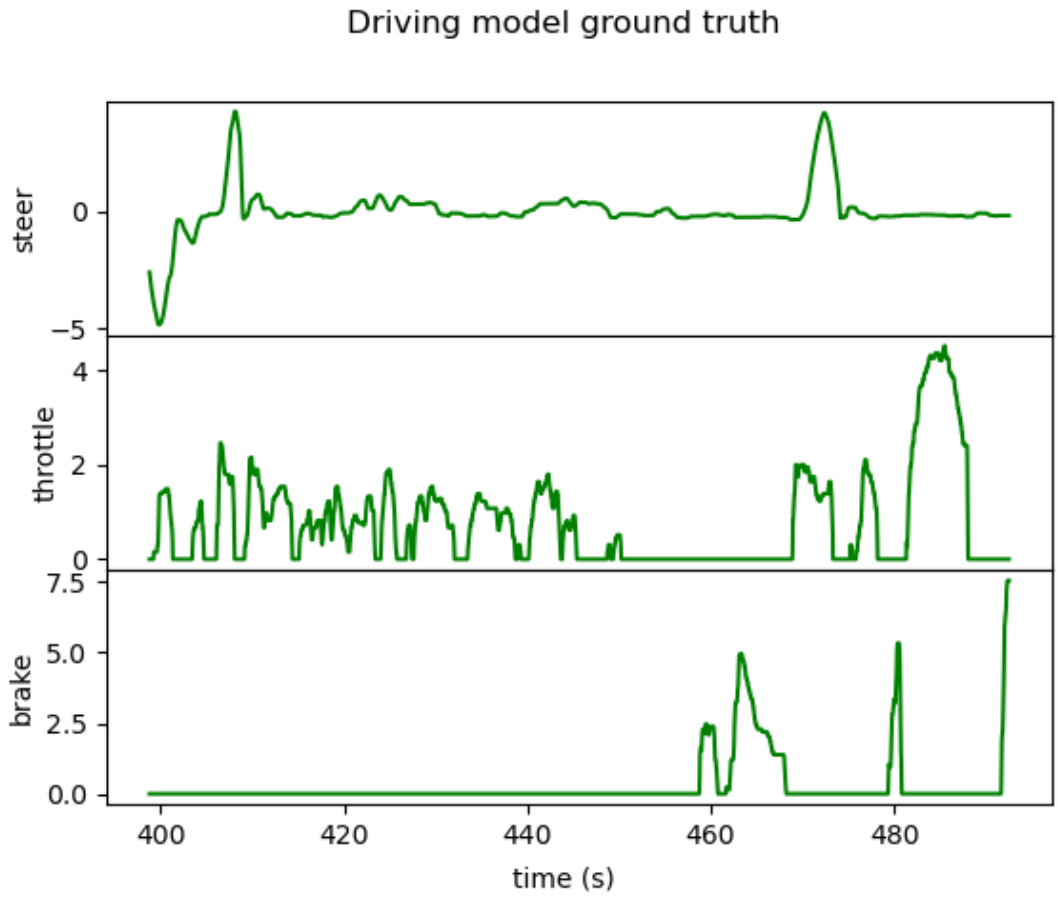

Figure 9: Actual Recorded Driving Input Ground Truth For Test Data

This plot showcases the final results of the model on previously unseen (testing) data and demonstrates the predictive potential given a small amount of training. While certainly imperfect, the overall model has shown potential for learning and utilizing human behaviour, both internally (eye data, head pose) and externally (vehicle state), to train its individual subsymbolic components. Comparing the predictions from the ground truth, the steering model closely resembles the actual inputs, and the throttle and brake models are less accurate but still often match the peaks from the data. Additionally, notice that the magnitudes for all three inputs roughly match those of the ground truth,

which are relatively normalized so they are not necessarily clamped between [0, 1]. Furthermore, the throttle and brake predictions never produce negative values and are never simultaneously active, as defined by the symbolic logic described earlier.

There is exciting potential for future work in this space, especially with more time and focus on the data capture. Due to limitations in time and scheduling, I was only able to gather data from a single participant (“Jacob”), however this project could be extended to train on data from multiple participants. This could present some interesting findings about how different people drive and their unique driving cues. Additionally, it would be interesting to see an implementation of explicit vehicle dynamics in the model so it wouldn’t need to learn this behaviour subsymbolically. Implementing this might speed up the learning process by leaning on more symbolic logic and focusing the intrinsic learning on the driving input. Finally, developing a stronger correlation between eye data and the driver’s internal state could reveal underlying features of the driver’s intentions, situational awareness, cognitive load, and potentially more. There is a wide range of applicability that a future project building off this one could contribute to the larger field of cognitive modeling.

REFERENCES:

[1] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of mind. Psychological Review, 111(4), 1036-1060.

[2] Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. Psychological review, 115(1), 101–130. https://doi.org/10.1037/0033-295X.115.1.101

[3] Ioanna Katidioti and Jelmer P. Borst and Douwe J. Bierens de Haan and Tamara Pepping and Marieke K. van Vugt and Niels A. Taatgen, “Interrupted by Your Pupil: An Interruption Management System Based on Pupil Dilation”, doi:10.1080/10447318.2016.1198525, 2016

[4] Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2016. Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. In Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16). Association for Computing Machinery, New York, NY, USA, 78–85. DOI:https://doi.org/10.1145/2993901.2993908

[5] Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. Science, 132, 349–350.

[6] Silvera, Gustavo, Abhijat Biswas, and Henny Admoni. "DReyeVR: Democratizing driving simulation in virtual reality for behavioural & interaction research." arXiv preprint arXiv:2201.01931 (2022).

[7] Salvucci DD. Modeling Driver Behavior in a Cognitive Architecture. Human Factors. 2006;48(2):362-380. doi:10.1518/001872006777724417

[8] Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2016. Measuring Cognitive Load using Eye Tracking Technology in Visual Computing. In Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization (BELIV '16). Association for Computing Machinery, New York, NY, USA, 78–85. DOI:https://doi.org/10.1145/2993901.2993908

[9] I. Kotseruba and J. K. Tsotsos, “Behavioral research and practical applications of drivers’ attention,” arXiv:2104.05677, 2021

[10] Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., & Koltun, V. (2017, October). CARLA: An open urban driving simulator. In Conference on robot learning (pp. 1-16). PMLR.

[11] HTC. 2019. HTC Vive Pro Eye. https://www.vive.com/us/product/vive-pro-eye/overview/

[12] Captum.ai. 2022. Captum · Model Interpretability for PyTorch. [online] Available at: <https://captum.ai/tutorials/Titanic_Basic_Interpret>.

[13] A. Palazzi, D. Abati, s. Calderara, F. Solera and R. Cucchiara, "Predicting the Driver's Focus of Attention: The DR(eye)VE Project," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 7, pp. 1720-1733, 1 July 2019, doi: 10.1109/TPAMI.2018.2845370.

[14] Driversed.com. 2022. Brakes. [online] Available at: <https://driversed.com/driving-information/the-vehicle/brakes/>.

[15] Salvucci, D. D., & Gray, R. (2004). A two-point visual control model of steering. Perception, 33(10), 1233–1248. https://doi.org/10.1068/p5343

[16] Haring, Kerstin & Watanabe, Katsumi & Ragni, Marco & Konieczny, Lars. (2013). The Use of ACT-R to Develop an Attention Model for Simple Driving Tasks. Psychology Research, ISSN 2159-5543. 3. 189-198.