2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Human-Aligned Longitudinal Control for Occluded Pedestrian Crossing With Visual Attention

Vinal Asodia[1], Zhenhua Feng[2], and Saber Fallah[1]

Abstract— **Reinforcement Learning (RL) has been widely used to create generalizable autonomous vehicles. However, they rely on fixed reward functions that struggle to balance values like safety and efficiency. How can autonomous vehicles balance different driving objectives and human values in a constantly changing environment? To bridge this gap, we propose an adaptive reward function that utilizes visual attention maps to detect pedestrians in the driving scene and dynamically switch between prioritizing safety or efficiency depending on the current observation. The visual attention map is used to provide spatial attention to the RL agent to boost the training efficiency of the pipeline. We evaluate the pipeline against variants of an occluded pedestrian crossing scenario in the CARLA Urban Driving simulator. Specifically, the proposed pipeline is compared against a modular setup that combines the well-established object detection model, YOLO, with a Proximal Policy Optimization (PPO) agent. The results indicate that the proposed approach can compete with the modular setup while yielding greater training efficiency. The trajectories collected with the approach confirm the effectiveness of the proposed adaptive reward function.**

I. Introduction

Autonomous Vehicles (AVs) represent a technological advancement with the potential to offer numerous benefits, such as increased safety by reducing human errors and greater efficiency in traffic flow. However, for the adoption of AVs, their actions must be aligned with human values such as the safety of other road users, driving efficiency, and ride smoothness of vehicle control. One such method to instill human values into AV systems is Reinforcement Learning (RL) which is a widely used machine learning paradigm for creating generalizable AV systems and has shown promising results [1][2][3]. Within RL, a reward function is crafted to guide the system to complete a given task and if designed in a certain manner, it can be an avenue to provide human values to the AV system.

Most of the existing RL methods use fixed reward functions, which comprise several components to capture different human values and the total reward equates to the summation of these components. Recent advances include components to ensure passenger safety and efficiency of vehicle control [4][5][6][7], as well as passenger comfort [8]. Additional prior research [9] proposed a decentralized general reward function to balance the levels of egotistic and altruistic behaviors, which allows AVs to effectively coexist with other road users. This is a common goal within the research community, as there are ongoing discussions on how AVs can coexist with human drivers and the social-ethical considerations that need to be made. Studies such as [10][11] offer important insights into the current methodologies of making AV systems human-aligned and the accompanying benefits such as building public trust and overall safer systems. However, a re-emerging issue arises with the fixed reward functions, as they often rely on balancing objectives that can conflict with each other. A key example would be safety and efficiency. If the AV prioritizes safety over efficiency, it can lead to overly conservative driving. Conversely, if there is an emphasis on vehicle efficiency over safety, it can lead to reckless driving.

To circumvent the issues with fixed reward functions, Thornton et.al, proposed the concept of adaptive reward functions, which dynamically adjust the importance of each component in the reward function depending on the last observation of the driving scene [12][13]. In their work, they crafted adaptive reward functions to navigate through pedestrian crossings and incorporate the human values of safety, legality, and mobility. If the AV does not observe any pedestrians, the AV will prioritize mobility over safety. However, if a pedestrian is in the AV's path then the AV will switch its priorities toward safety and legality. In these studies, LiDAR and object detection modules have been widely used to detect pedestrians.

In this paper, we propose a solution involving visual attention maps, which have been used extensively in the past to increase understanding of complex scenes [14][15]. Given an image $\mathbf{I} \in \mathbb{R}^{C \times W \times H}$, where C is the number of color channels, W and H are the width and height of the image, the attention map $\mathbf{M} \in \mathbb{R}^{W' \times H'}$ can be generated, where each value m_{ij} of the attention map represents an attention score (typically between $[0, 1]$) for the corresponding pixel of the image $\mathbf{I}$. An attention map can be generated in a supervised or unsupervised manner. The main advantage of producing attention maps in a supervised manner over unsupervised methods is that there is direct control over what objects are salient. For unsupervised methods like self-attention, the model identifies salient objects based on the input data, which can result in regions highlighted with no clear rationale.

To apply the attention scores to the features of an image, typically cross-wise multiplication is performed between the feature embeddings and the attention map [16]. This technique is referred to as spatial attention [17], which is a subcategory of attention mechanisms that imitates how humans

[1]V. Asodia and S. Fallah are with the Department of Mechanical Engineering Sciences, University of Surrey, Guildford, GU2 7XH UK. {va00191, s.fallah}@surrey.ac.uk

[2]Z. Feng is with the School of Computer Science and Electronic Engineering, University of Surrey, Guildford, GU2 7XH UK z.feng@surrey.ac.uk

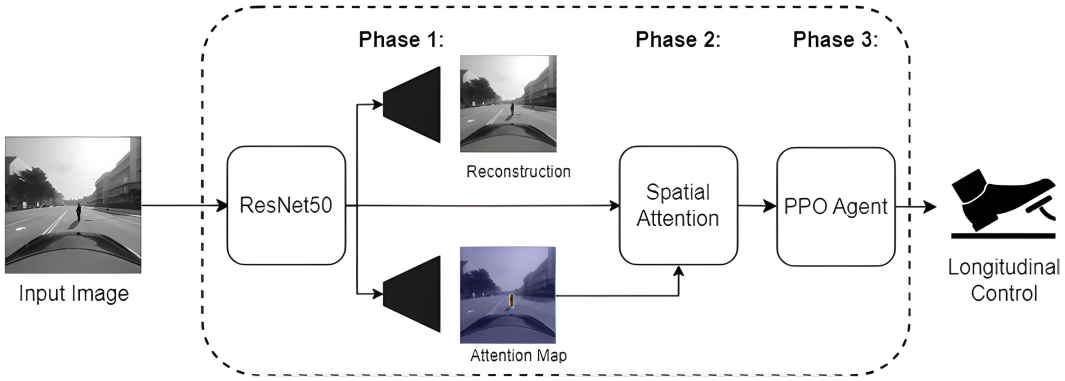

Fig. 1. An overview of the proposed pipeline, which has 3 phases. Phase 1 takes in the image observation and passes it to the encoder-decoder model, which produces a visual attention map, highlighting the pedestrian in the scene and a reconstruction of the original image. Phase 2 involves using the visual attention map to provide spatial attention to the feature embeddings of the input image, which are then used by the PPO agent to output the ego vehicle's longitudinal control in Phase 3.

focus on specific objects or features of their surroundings when completing a given task. By highlighting the features of salient objects in the input image, the model's decision-making is less influenced by irrelevant features, potentially increasing the model's convergence rate.

There is a large volume of published work that has incorporated attention into AV systems to boost performance and offer transparency in several driving tasks. Recent examples include driving through un-signalized intersections [18], highway safety planning [19], and navigating through 5-lane junctions [20]. The key problem with existing attention solutions lies in the interpretability of the model, specifically understanding the semantic reasoning behind the weighting of different image regions. In previous studies [20][21][22], when examining the resulting visual attention maps from the respective pipelines, regions of the image that are not pertinent to the task at hand have been highlighted without offering a clear rationale for why those regions have been emphasized. This is a significant issue as these highlighted regions may harm the model's decision-making and in the event of a collision, the absence of a clear rationale for these regions complicates the task of diagnosing the system. To address this issue, we propose to guide the model on what to focus on by creating visual attention map labels, highlighting key objects in the scene like pedestrians, and training the model to generate the attention maps in a supervised manner. This form of supervised attention allows stakeholders to provide reasoned focus to the model and minimize instances where irrelevant areas are highlighted. To output the longitudinal vehicle control, we use the well-established RL algorithm, Proximal Policy Optimization (PPO) [23].

To summarize, the key contributions of the proposed method include:

- We propose an end-to-end pipeline that generates visual attention maps in a supervised manner, which allows us to have greater control over what the system should focus on when driving. The attention maps capture the presence of a variety of objects in a single, human-readable format.
- We present a method to boost the performance of vision-based AV systems, by using the visual attention map to provide spatial attention to the PPO agent.
- We formulate an adaptive, human-aligned reward function for the PPO agent using the attention map to identify key objects in real-time.

II. Methodology

A. Visual Attention Map Generation

The pipeline is illustrated in Figure 1 and is split into 3 phases. To generate the visual attention map we have chosen to pre-train an encoder-decoder model in a supervised manner. First, we pass a grayscale image observation (dimensions $1 \times 224 \times 224$) to a ResNet50 encoder [24], which outputs the image's feature embedding. The feature embedding is a vector that captures the features of the original image. Afterwards, the features are passed to two separate decoder streams. The first stream acts as a traditional autoencoder and outputs the reconstruction of the original image, from the feature embedding. The purpose of the first stream is to learn an accurate set of feature embedding. After network training, this stream is removed as it is no longer needed for inference. The second stream generates the visual attention map from the feature embedding. The visual attention map has a dimensionality of $1 \times 56 \times 56$ and contains values between $[0, 1]$. Areas containing a pedestrian will have values nearing 1, while regions without a pedestrian will have values closer to 0. The benefit of using a supervised attention map is that it enables the developer to explicitly control what the system should focus on for a given task. Additionally, the supervised attention maps can capture the information of multiple objects in one human-readable format. Section II-C demonstrates how this task-related attention can be used to formulate an adaptive reward function for the RL agent.

Both decoder streams consist of a series of deconvolutional layers and the training details will be given in Section III-C.

B. Spatial Attention

Continuing onto phase 2 of the pipeline, Figure 2 illustrates the steps needed to provide spatial attention to

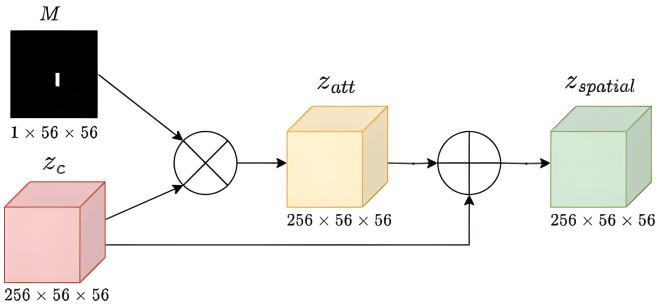

Fig. 2. Phase 2 involves using the attention map to apply spatial attention to the feature map of the input image in a two-step process: 1) a cross-wise product between the attention map $\mathbf{M}$ and latent feature map $\mathbf{Z}_c$, which produces a set of intermediate features $\mathbf{Z}_{att}$ and 2) an addition operation that produces the final features with spatial attention $\mathbf{Z}_{spatial}$.

the original image feature map using the attention map $\mathbf{M}$. Specifically, we apply spatial attention to the original image feature map, $\mathbf{Z}_c \in \mathbb{R}^{256\times56\times56}$, which is the output of the penultimate layer of the encoder. First, we perform an element-wise product between the visual attention map and latent feature map $\mathbf{Z}_c$, which produces an intermediate feature map $\mathbf{Z}_{att} \in \mathbb{R}^{256\times56\times56}$. This is followed by an element-wise addition step between the intermediate features $\mathbf{Z}_{att}$ and the original features $\mathbf{Z}_c$, to give the final set of features with added spatial attention $\mathbf{Z}_{spatial} \in \mathbb{R}^{256\times56\times56}$. These features are then flattened and sent to the PPO agent for longitudinal control.

C. Reinforcement Learning Agent

The final phase of the pipeline is the PPO agent, which will perform the ego vehicle's longitudinal control. We have opted to use the reliable implementation from StableBaselines3 [25], as we can easily combine the PPO agent with the earlier phases of the pipeline. PPO uses an Actor-Critic style architecture, consisting of a policy network that outputs the action from the learned policy π and a value network that outputs an estimate of the expected return of a state, which is used in calculating the advantage function. Both policy and value networks are made up of a series of linear layers. The following sections will outline how we formulated the RL problem.

1) State Space: The observation space consists of $1 \times 224 \times 224$ grayscale images from a camera placed on the ego vehicle's dashboard. We selected a grayscale format for the image to lessen the computational burden of the system.

2) Action Space: For the system to perform smooth and efficient longitudinal control, the PPO agent will output 1 continuous action between the range $[-1, 1]$ to control the speed of the ego vehicle. If the action is greater than 0, the agent will adjust the ego vehicle's throttle. Conversely, if the action is less than 0, the agent will exert the equivalent level of braking. The continuous action output will allow the agent to smoothly regulate the ego vehicle's speed.

3) Reward Function: We adapt the reward function from [12], which takes the safety, legality, efficiency, and smoothness of the ego vehicle's longitudinal control into account. The reward function is split into 3 components (Eq. 1, Eq. 2, and Eq. 3), with the first component accounting for safety and legality:

$$g_{safe}(x_t, u_t) = -(\zeta \frac{v_t^2}{d_t + \epsilon} + \eta \mathbf{1}(d_1 = 0))\mathbf{1}(c_t), \quad (1)$$

where $\zeta > 0$ is a weight on the penalty applied to the ego vehicle if it drives too fast in the presence of a pedestrian, v_t is the ego vehicle's speed, d_t is the distance between the ego vehicle and the pedestrian, $\epsilon > 0$ is a constant to offset the denominator, $\eta > 0$ is a penalty to guide the ego vehicle to stop when the pedestrian is in its path and c_t is a Boolean value that denotes if the pedestrian is in the ego vehicle's path (more details on c_t is given in Section II-D). This term ensures that the ego vehicle brakes if the pedestrian is crossing the road.

The second component ensures efficiency in the ego vehicle's control and is formulated as:

$$g_{efficient}(x_t, u_t) = \lambda v_t \mathbf{1}(\neg c_t), \quad (2)$$

where $\lambda > 0$ is a weight on the reward term that incentivizes the ego vehicle to move at a greater speed when there is no pedestrian in its path. To ensure the ego vehicle's longitudinal control is smooth and comfortable for the passengers, the final component penalizes the agent for large acceleration values, which would result in jerky motions. The smoothness component is defined as:

$$g_{smooth}(x_t, u_t) = -\xi(v_t - v_{t+1})^2 = -\xi(a_t \Delta t)^2, \quad (3)$$

The selection of this reward function was motivated by its ability to consider a wide array of factors, ranging from legal and safety considerations to the effectiveness and smoothness of ego vehicle control.

D. Attention Map based Object Identification

The safety and efficiency components of the reward function rely on c_t, which is a Boolean flag representing the presence of a pedestrian in the ego vehicle's view. The authors of [12] used a LiDAR sensor to determine c_t, whereas we propose to use the attention map generated by the pipeline to identify the presence of the pedestrian in real time. At timestep t, if there is a detection in the attention map above a certain threshold area, then c_t is set to 1, signaling the proximity of a pedestrian close enough to warrant the ego vehicle reducing its speed. If there is no detection or the detected pedestrian is below the threshold area, then the pedestrian is either not crossing or they are too far away to justify slowing down, resulting in c_t being set to 0. This is a major improvement over the original approach [12], as LiDAR is an expensive sensor that only captures the shape of an object, whereas the attention map can be combined with the camera image to capture the colors and textures of each key object. This provides a more human-readable solution that can be easily upscaled to incorporate more types of objects and scenarios.

III. Experimental Setup

To evaluate the pipeline, we use the CARLA Urban Driving simulation environment [26], which provides a set of dynamic agents and sensors.

A. Occluded Pedestrian Crossing Scenario

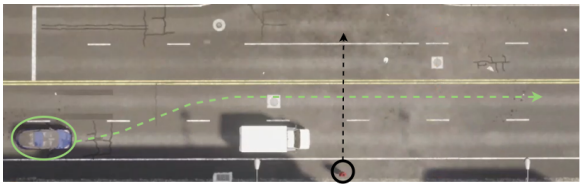

Fig. 3. Occluded pedestrian crossing scenario setup within CARLA. The ego vehicle is spawned at one end of the road and must overtake the large parked van and yield for the occluded crossing pedestrian, highlighted in black, to successfully reach the goal point at the other end of the road.

Figure 3 shows the setup in CARLA where the ego vehicle starts at one end of a multi-lane road, needing to navigate past a large stationary vehicle to reach the opposite end. Beyond the parked vehicle, an occluded pedestrian will move at a speed of $4km/h$ and momentarily stop in the lane occupied by the ego vehicle. The RL agent's goal is to safely and efficiently guide the ego vehicle to the end of the road by adjusting its longitudinal speed, while a PID controller outputs the lateral control needed to overtake the stationary vehicle.

To test the robustness of the approach, three variations of the scenario will be evaluated: one with occlusion but no traffic, another with both occlusion and traffic, and a third with a moving occlusion and traffic. An additional set of results will be collected where the large stationary vehicle is replaced with a smaller vehicle to partially obscure the pedestrian.

The ego vehicle is equipped with an RGB camera on its dashboard to capture the visual state and a collision sensor to detect any collisions with the parked vehicle or pedestrian.

B. Dataset

To train the encoder and attention map generator, a dataset of $130k$ visual attention map labels was collected in CARLA. To obtain the visual attention map labels, we first used CARLA to draw 3D box annotations (i.e. bounding boxes) around the pedestrian in the simulator. Then we built a projection matrix using the ego vehicle's camera's intrinsic parameters such as the focal length and dimensions, to translate the 3D corner points of the pedestrian bounding box to 2D corner points on the image plane of the camera. To focus only on pedestrians in front of the camera, we computed the dot product between the ego vehicle's forward vector and the vector leading to the bounding box. With these 2D corner points, we constructed the visual attention map labels by initializing an array of 0s and setting the elements within the bounding box to 1. To ensure the model could generalize well, for each sample we uniformly randomized the latitude and longitude of both the ego vehicle and the pedestrian. Additionally, we randomized the ego vehicle's rotation and the pedestrian's altitude.

C. Training Details

To train the encoder to generate the visual attention maps, we used the Mean Squared Error loss function with the Adam optimizer and a step learning rate scheduler that reduced the learning rate by a magnitude of 10 every 33 epoch. Additionally, we randomly performed the horizontal image flip transformation to half of the samples within each batch during training to increase the encoder's understanding of the scene. The encoder was set to train for 100 epochs. We utilized early stopping with a patience value of 3 epochs to prevent overfitting, and training was stopped after 63 epochs.

For training the PPO agent, we froze the parameters of the encoder and removed the decoder head responsible for reconstructing the original image as it is no longer needed. The PPO agent was trained for $200k$ timesteps and we opted to keep the hyper-parameters from StableBaselines3 as they had proven to be effective.

All training was conducted on an NVidia GeForce RTX 2080ti graphics card, with the encoder and PPO agent taking approximately 20 and 2.5 hours to train, respectively, totaling 22.5 hours of training time.

D. Model Evaluation

We compared the proposed pipeline with a modular approach, "Adaptive YOLO+PPO," that combines ScaledYOLOv4 [27] for pedestrian detection with the PPO agent and uses the proposed adaptive reward function. This comparison aimed to assess whether our attention mechanism could rival a well-established perception model. The YOLO model exclusively detected pedestrians for a fair comparison. We also evaluated the impact of the adaptive feature of the reward function by comparing it to a variant of the proposed pipeline without the c_t boolean flag, referred to as the "PPO" baseline.

The evaluation of the adaptive reward function focused on safety, efficiency, and smoothness components. Safety was assessed through collision tests across the scenario variants stated in Section III-A, including different levels of pedestrian occlusion (i.e. partial and full occlusion). We adjusted the level of occlusion to investigate how different degrees of pedestrian visibility affect the pipeline's performance. The efficiency and smoothness components of the adaptive reward function were evaluated by collecting trajectories for each scenario variant using the proposed approach.

An ablation study was conducted to highlight the significance of various pipeline components. Mean training reward per timestep was used as the performance metric across four setups: the proposed pipeline outlined in Figure 1, the pipeline without the final cross-wise addition step in Figure 2, a setup with an additional decoder head to reduce the visual attention map, which is passed to the PPO agent, and a setup without the attention mechanism, directly passing the image's feature embeddings to the PPO agent.

IV. Results and Discussion

To evaluate the performance of the pipeline, we measured the success rate, collision rate, and average stopping distance

TABLE I

COLLISION TEST RESULTS FROM RUNNING EACH SETUP FOR 100 EPISODES IN THREE VARIANTS OF THE OCCLUDED PEDESTRIAN CROSSING SCENARIO, WHERE S = SUCCESSFUL, C = COLLISION, SD = STOPPING DISTANCE, P = PARTIAL OCCLUSION, AND F = FULL OCCLUSION.

	PPO						**Adaptive YOLO + PPO**						**Proposed Approach**					
	S (%)		**C (%)**		**SD (m)**		**S (%)**		**C (%)**		**SD (m)**		**S (%)**		**C (%)**		**SD (m)**	
Scenario	**P**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**	**F**
Occlusion & No Traffic	2	1	98	99	4.70	**4.59**	94	91	6	9	3.94	3.72	**99**	**98**	**1**	**2**	**5.00**	4.56
Occlusion & Traffic	1	1	99	99	2.92	2.69	92	90	8	10	3.88	3.48	**97**	**95**	**3**	**5**	**5.12**	**4.55**
Moving Occlusion & Traffic	1	0	99	100	2.18	N/A	91	86	9	14	4.04	3.81	**96**	**93**	**4**	**7**	**5.24**	**4.71**

for each setup for 100 episodes in variants of the occluded pedestrian crossing scenario. The primary takeaway from Table I is that the proposed approach has a higher success rate than the Adaptive YOLO+PPO baseline. However, more importantly, neither setup was able to achieve a 100% success rate in any of the scenario variants. For all setups, the collision rate increases as the scenario becomes more complex, and the average stopping distance for the proposed approach is approximately a vehicle's length, which highlights the effectiveness of the safety component of the adaptive reward function (Eq. 1). The Adaptive YOLO+PPO approach exhibited a shorter average stopping distance, indicating that analyzing the attention map to determine c_t is more effective than relying on YOLO's bounding boxes. However, the delayed pedestrian detection by YOLO could be attributed to the lower image quality of the simulation platform compared to real-life scenarios. The results for the PPO setup follow the same trend for each scenario variant. This is due to the PPO setup trying to give equal weight to both safety and efficiency has resulted in a "tug-of-war" within the agent. This caused the majority of episodes to end in a collision. By comparing the results with partial and full occlusion, it is evident that having a partial view of the pedestrian allows both the Adaptive YOLO+PPO and the proposed approach to give way to the pedestrian more often. Unfortunately, there are still rare occasions where both setups collide with the pedestrian.

These results highlight a limitation of the pipeline: in busy driving scenes, where pedestrians are visible in one instance and occluded in the next, using a single frame loses the historical context of the pedestrian's position. To address this limitation, the pipeline could be extended to incorporate stacked frames to capture temporal information, which would provide historical context about the pedestrian's presence. Additionally, it is vital to provide the AV system with a global context by capturing the behaviors and intent of other road users. This enhancement would aid the AV in navigating through dense traffic, where its vision may be limited. Furthermore, the adaptive reward function could be refined by adjusting the variable c_t to represent the probability of a pedestrian entering the vehicle's path, taking into account both the temporal and global context of the current scene, rather than limiting it to binary values of 0 or 1. Implementing these adjustments would enable the AV to navigate dense traffic areas more safely and efficiently.

A. Trajectory Analysis

To illustrate the impact of each component of the adaptive reward function, we have collected the speed, acceleration, and reward per second for 1 trajectory in each scenario variant using the proposed approach (shown in Figure 4). Taking a look at the first variant, in the beginning, there is no pedestrian present ($c_t = 0$) and the vehicle aims to maximize the efficiency component of the reward function (Eq. 2) by accelerating to its maximum speed. From $t = 1$ to $t = 6$, the vehicle maintains its maximum speed of $6ms^{-1}$, resulting in its acceleration oscillating around $0ms^{-2}$ and the agent accumulating a maximum reward per second of 2. Shortly after, a pedestrian is detected in the vehicle's path (setting c_t to 1), which switches the efficiency component of the reward function to the safety component (Eq. 1). Here the vehicle is penalized for driving at high speeds toward the pedestrian, resulting in a high deceleration at $t = 6.5$, which nets the agent a negative spike in reward from the smooth component (Eq. 3). The vehicle remains stationary until the pedestrian leaves its path and c_t is set back to 0, activating the efficiency component and causing the vehicle to drive off. The trajectories for the remaining rows of Figure 4 follow a similar trend, however, the smoothness of control seems to deteriorate as the scenario becomes more difficult. This is supported by the trajectory for the moving occlusion variant, where there are multiple minor applications of the throttle and brake during times when the pedestrian is not detected. This is indicative that the smoothness component is adequate for simple scenarios, but needs improvement or finetuning to handle busier driving scenes.

B. Ablation Study

To conclude the evaluation of the pipeline, we present the results from the ablation study to highlight the importance of each stage in the pipeline. Figure 5 depicts the training curves for each setup outlined in Section III-D and the full proposed pipeline yielded the highest mean reward at the end of training. The most interesting aspect of this graph is the drop in performance when the final cross-wise addition step is removed. We hypothesize that the cross-wise multiplication step outlined in Section II-B overly suppresses regions where

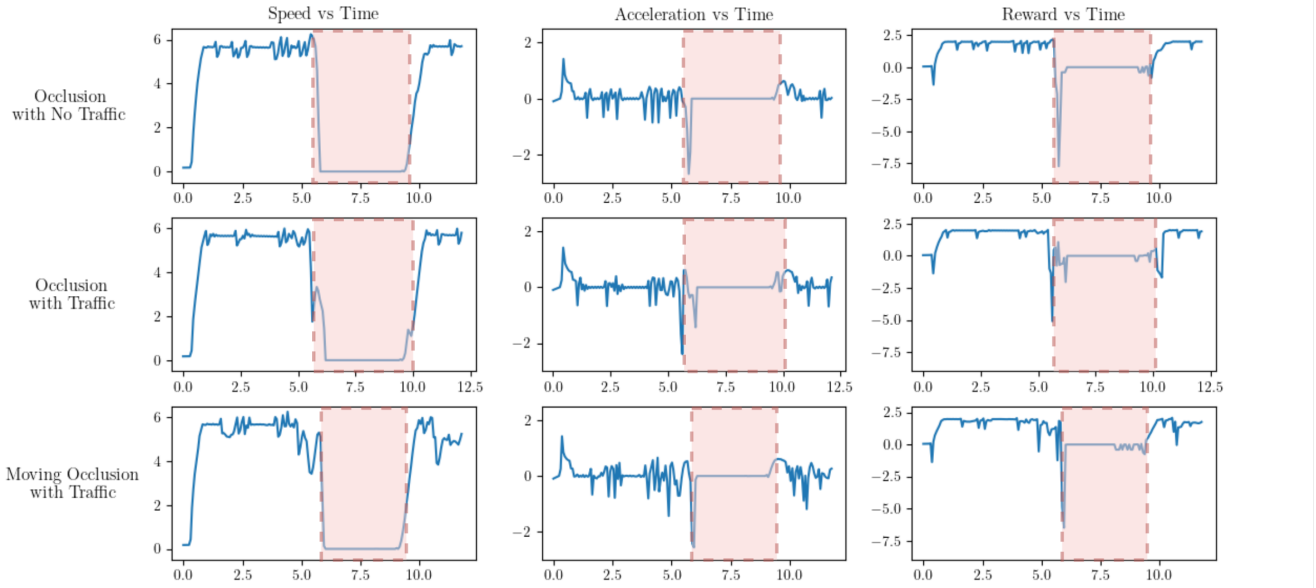

Fig. 4. Grid of plots depicting the speed, acceleration, and reward per second for a trajectory collected in each scenario variant using the proposed approach. Each row represents a scenario variant: occlusion with no traffic, occlusion with traffic, and moving occlusion with traffic. The areas highlighted in red represent situations where $c_t = 1$ and a pedestrian is detected in the vehicle's path.

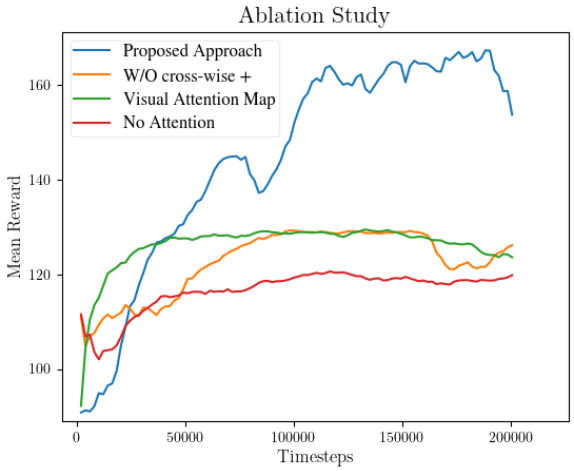

Fig. 5. Ablation Study results.

the attention values are near 0 and that results in useful information being lost. This notion is further supported by the sub-optimal performance of the "Visual Attention Map" setup and highlights the need to improve the visual attention maps generated by the system. This could be achieved by refining the attention map labels and applying a Gaussian filter to smooth areas of greater importance, which would add attention to areas surrounding key objects. Finally, it is evident that removing all attention mechanisms leads to the largest decrease in performance and emphasizes the system's struggle in learning from the raw feature embeddings.

V. Conclusion

In this work, we proposed an end-to-end pipeline that generates visual attention maps for two main purposes: to apply spatial attention to the system and to identify critical objects like pedestrians for an adaptive, human-aligned reward function that we propose. This approach not only aligns with key human values like safety, legality, and mobility but also provides greater transparency into the system's decision-making process than existing models. We demonstrated the pipeline's effectiveness in an occluded pedestrian crossing scenario in CARLA and used a modular baseline of YOLO with PPO to evaluate the pipeline. The results indicated that the proposed pipeline could compete with the modular setup of YOLO with PPO and the adaptive feature of the reward function is effective at balancing conflicting objectives (i.e. safety and efficiency). However, neither setup was able to achieve a 100% success rate in dense and complex traffic settings. Directions for future work are motivated by the limitations of the current pipeline. Possible improvements involve enhancing the pipeline by incorporating stacked frames to capture temporal information and monitoring the behaviors of other road users to capture global context. This deeper understanding of the driving scene can be used to refine the adaptive reward function to represent the probability of pedestrian interactions in a nuanced manner.

References

[1] Yildirim, Mustafa, Sajjad Mozaffari, Luc McCutcheon, Mehrdad Dianati, Alireza Tamaddoni-Nezhad, and Saber Fallah. "Prediction based decision making for autonomous highway driving." In 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), pp. 138-145. IEEE, 2022.

[2] Li, Quanyi, Zhenghao Peng, and Bolei Zhou. "Efficient learning of safe driving policy via human-ai copilot optimization." arXiv preprint arXiv:2202.10341 (2022).
[3] Zhang, Zhejun, Alexander Liniger, Dengxin Dai, Fisher Yu, and Luc Van Gool. "End-to-end urban driving by imitating a reinforcement learning coach." In Proceedings of the IEEE/CVF international conference on computer vision, pp. 15222-15232. 2021.
[4] Muzahid, Abu Jafar Md, Syafiq Fauzi Kamarulzaman, Md Arafatur Rahman, and Ali H. Alenezi. "Deep reinforcement learning-based driving strategy for avoidance of chain collisions and its safety efficiency analysis in autonomous vehicles." IEEE Access 10 (2022): 43303-43319.
[5] Sun, Jingbo, Xing Fang, and Qichao Zhang. "Reinforcement Learning Driving Strategy based on Auxiliary Task for Multi-Scenarios Autonomous Driving." In 2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS), pp. 1337-1342. IEEE, 2023.
[6] Elallid, Badr Ben, Hamza El Alaoui, and Nabil Benamar. "Deep Reinforcement Learning for Autonomous Vehicle Intersection Navigation." arXiv preprint arXiv:2310.08595 (2023).
[7] Cui, Jianxun, Boyuan Zhao, and Mingcheng Qu. "An Integrated Lateral and Longitudinal Decision-Making Model for Autonomous Driving Based on Deep Reinforcement Learning." Journal of Advanced Transportation 2023 (2023).
[8] Zhang, Mei, Kai Chen, and Jinhui Zhu. "An efficient planning method based on deep reinforcement learning with hybrid actions for autonomous driving on highway." International Journal of Machine Learning and Cybernetics 14, no. 10 (2023): 3483-3499.
[9] Selvaraj, Dinesh Cyril, Shailesh Hegde, Nicola Amati, Francesco Deflorio, and Carla Fabiana Chiasserini. "A Deep Reinforcement Learning Approach for Efficient, Safe and Comfortable Driving." Applied Sciences 13, no. 9 (2023): 5272.
[10] Umbrello, Steven, and Roman V. Yampolskiy. "Designing AI for explainability and verifiability: a value sensitive design approach to avoid artificial stupidity in autonomous vehicles." International Journal of Social Robotics 14, no. 2 (2022): 313-322.
[11] Atakishiyev, Shahin, Mohammad Salameh, Hengshuai Yao, and Randy Goebel. "Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions." arXiv preprint arXiv:2112.11561 (2021).
[12] Thornton, Sarah M., Francis E. Lewis, Vivian Zhang, Mykel J. Kochenderfer, and J. Christian Gerdes. "Value sensitive design for autonomous vehicle motion planning." In 2018 IEEE intelligent vehicles symposium (IV), pp. 1157-1162. IEEE, 2018.
[13] Thornton, Sarah M., Benjamin Limonchik, Francis E. Lewis, Mykel J. Kochenderfer, and J. Christian Gerdes. "Toward closing the loop on human values." IEEE Transactions on Intelligent Vehicles 4, no. 3 (2019): 437-446.
[14] Zhao, Yinuo, Kun Wu, Zhiyuan Xu, Zhengping Che, Qi Lu, Jian Tang, and Chi Harold Liu. "Cadre: A cascade deep reinforcement learning framework for vision-based autonomous urban driving." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 3, pp. 3481-3489. 2022.
[15] Lateef, Fahad, Mohamed Kas, and Yassine Ruichek. "Saliency heatmap as visual attention for autonomous driving using generative adversarial network (GAN)." IEEE Transactions on Intelligent Transportation Systems 23, no. 6 (2021): 5360-5373.
[16] Wörmann, Julian, Daniel Bogdoll, Etienne Bührle, Han Chen, Evaristus Fuh Chuo, Kostadin Cvejoski, Ludger van Elst et al. "Knowledge augmented machine learning with applications in autonomous driving: A survey." arXiv preprint arXiv:2205.04712 (2022).
[17] Guo, Meng-Hao, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R. Martin, Ming-Ming Cheng, and Shi-Min Hu. "Attention mechanisms in computer vision: A survey." Computational visual media 8, no. 3 (2022): 331-368.
[18] Seong, Hyunki, Chanyoung Jung, Seungwook Lee, and David Hyunchul Shim. "Learning to drive at unsignalized intersections using attention-based deep reinforcement learning." In 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), pp. 559-566. IEEE, 2021.
[19] Chen, Guoxi, Ya Zhang, and Xinde Li. "Attention-based Highway Safety Planner for Autonomous Driving via Deep Reinforcement Learning." IEEE Transactions on Vehicular Technology (2023).
[20] Fu, Wen, Yanjie Li, Zhaohui Ye, and Qi Liu. "Decision Making for Autonomous Driving Via Multimodal Transformer and Deep Reinforcement Learning." In 2022 IEEE International Conference on Real-time Computing and Robotics (RCAR), pp. 481-486. IEEE, 2022.
[21] Kim, Jinkyu, and John Canny. "Interpretable learning for self-driving cars by visualizing causal attention." In Proceedings of the IEEE international conference on computer vision, pp. 2942-2950. 2017.
[22] Kim, Jinkyu, Teruhisa Misu, Yi-Ting Chen, Ashish Tawari, and John Canny. "Grounding human-to-vehicle advice for self-driving vehicles." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10591-10599. 2019.
[23] Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal policy optimization algorithms." arXiv preprint arXiv:1707.06347 (2017).
[24] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
[25] Raffin, Antonin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. "Stable-baselines3: Reliable reinforcement learning implementations." The Journal of Machine Learning Research 22, no. 1 (2021): 12348-12355.
[26] Dosovitskiy, Alexey, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. "CARLA: An open urban driving simulator." In Conference on robot learning, pp. 1-16. PMLR, 2017.
[27] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "Scaled-yolov4: Scaling cross stage partial network." In Proceedings of the IEEE/cvf conference on computer vision and pattern recognition, pp. 13029-13038. 2021.