# Mitigating Causal Confusion in Driving Agents via Gaze Supervision

**Abhijat Biswas**
Carnegie Mellon University
abhijat@cmu.edu

**Badal Arun Pardhi**
Carnegie Mellon University
bpardhi@andrew.cmu.edu

**Caleb Chuck**
The University of Texas at Austin
calebc@cs.utexas.edu

**Jarrett Holtz**
Bosch
jaholtz@utexas.edu

**Scott Niekum**
The University of Texas at Austin
sniekum@cs.utexas.edu

**Henny Admoni**
Caernegie Mellon University
henny@cmu.edu

**Alessandro Allievi**
Bosch
alessandro.allievi@us.bosch.com

**Abstract:** Imitation Learning (IL) algorithms such as behavior cloning (BC) do not explicitly encode the underlying causal structure of the task being learnt. This often leads to mis-attribution about the relative importance of scene elements towards the occurrence of a corresponding action, a phenomenon termed *Causal Confusion* or *Causal Misattribution*. Causal confusion is made worse in highly complex scenarios such as urban driving where the agent has access to a large amount of information per time-step (visual data, sensor data, odometry, etc.).

Our key idea is that while driving, human drivers naturally exhibit an easily obtained, continuous signal that is highly correlated with causal elements of the state space: eye gaze. We collect human driver demonstrations in a CARLA-based VR driving simulator, DReyeVR, allowing us to capture eye gaze in the same simulation environment as other training data commonly used in prior work. Further, we propose a contrastive-learning method to use gaze-based supervision to mitigate causal confusion in urban driving IL agents — exploiting the relative importance of gazed-at and not-gazed-at scene elements for driving decision making. We present preliminary quantitative results that suggest the promise of gaze-based supervision in improving the driving performance of IL agents.

**Keywords:** Causal confusion; causal misattribution; Eye gaze; Urban Driving; Imitation Learning; Behavior Cloning

## 1 Introduction

Imitation learning is a popular method for learning urban driving policies due to its ease of implementation and de-coupling of the data collection/action step and the training step by allowing offline learning of control, among other factors.

In the original paper identifying casual confusion [1], the authors use the example of the realistic driving setting to illustrate this phenomenon where, counter-intuitively, access to more information

Aligning Robot Representations with Humans Workshop @ 6th Conference on Robot Learning (CoRL 2022), Auckland, New Zealand.

yields poorer task performance by the imitation learning agent. In the aforementioned example, an IL agent learns from demonstration images from inside the cab of a vehicle with and without a brake light indicator on the dash. In the case where a brake light is present and always on when the brake is applied, the agent may learn to brake only when the brake light indicator is on. This is an undesirable misattribution of cause and effect.

Several additional works exist in which recurrent/history-based imitation models perform worse than their counterparts without access to this historical information. For instance in [2], imitation learning policies are trained with and without "history" information about the trajectory of the car in the past. The model with history has better performance on held-out demonstration data, but much worse performance when actually deployed which is an indication that causal confusion is occurring. Another example is in [3] —"*In particular, we identify a typical failure mode due to a subtle dataset bias: the inertia problem. When the ego vehicle is stopped (e.g., at a red traffic light), the probability it stays static is indeed overwhelming in the training data. This creates a spurious correlation between low speed and no acceleration, inducing excessive stopping and difficult restarting in the imitative policy.*"

## 2 Causal confusion in urban driving

Figure 1: Salience map generated by using blur-based saliency [4]. Clockwise starting from top-left: *left-camera* image with overlaid salience; *center-camera* image with overlaid salience; *right-camera* image with overlaid salience; next *waypoint* (in heatmap form) with overlaid salience; next *waypoint* (in heatmap form) overlaid on center image; quartered salience image to show scale of relative salience across inputs (top-left corresponds to *left*, top right to *center*, bottom right to *waypoint*, bottom left to *right*) This scene depicts the instant a vehicle comes to a stop after which it fails to restart. NOTE: the bulk of the salience weight as shown by the quartered image is at the base of the traffic light in the *center* image and the *waypoint*.

In particular, we consider the popular and well-established Learning by Cheating (LBC) [5] model for autonomous urban driving in CARLA. This method uses a 2-step approach where a teacher model is first trained with access to ground truth, overhead-view semantic segmentation maps around the ego-vehicle (approaching perfect perception). Then, this agent is used as an oracle to train a sensorimotor agent which only has access to RGB (*left/center/right* views) sensor data as well as a high-level command from a global plan. We use the latest author-provided code and model [6] which is a slight deviation from the original paper [5]. Of particular note, the LBC sensorimotor model takes a ten channel image as input where 9 channels correspond to three RGB images (*left, center, right*) and the last is a heatmap with the only "hot" region being a Gaussian distribution centered at the next waypoint in the frame of the *center* camera.

As one may expect, the LBC model also shows symptoms of suffering causal confusion. Some qualitative descriptions from the authors can be found in [7]: "*I believe that the network has some issues with starting/stopping*". These problems seem to occur especially in the absence of surrounding vehicles which may be used as causal cues for accelerating out of a stop.

We especially notice traffic light infractions where the LBC agent either does not stop for a red light or fails to restart after stopping at a red light. We also notice cases where the agent stops at a red light but restarts when opposing traffic is moving, even though the red light has not changed.

### 2.1 Saliency-based causal confusion diagnosis

To investigate the relative importance of regions of the input state space in making decisions, we used saliency methods to investigate the decision making process of the LBC model. Specifically, we used a modified version of the blur-based saliency method by Greydanus *et al.* [4]. This saliency method is network agnostic and works by blurring different regions of the given visual input and measuring the difference in output with the original input. This method reasons that regions which, when blurred, cause the greatest difference in output are the most salient.

Using this blur-based saliency measure, we are able to generate saliency maps for the LBC method such as in Fig 1.

## 3 Method

### 3.1 Gaze data collection


(a) Physical setup with participant driver in driving pose, alongside experimenter's setup monitoring the simulation.

(b) First person DReyeVR simulator perspective during the same episode with eye reticle (red crosshair) denoting eye gaze on in-world navigational sign that gives drivers route direction. The crosshair is for illustration only (not shown in VR).

Figure 2: Example experimental setup during gaze data collection

Human demonstration data was collected in the DReyeVR simulator [8], a modified version of the CARLA simulator to enable human driving in VR. DReyeVR also enables the collection of driver eye gaze as they use the simulator. Drivers were tasked with completing a navigational sign following task (see Fig. 2b) and their driving actions (steering, throttle, brake) as well as eye gaze movements were recorded.

Eye gaze was collected at the simulator rate, about 50Hz. Eye gaze can be a noisy and high frequency signal and so, we performed pre-processing in the following manner. First, driver eye gaze movements were classified into low-velocity fixations and high-velocity saccades using I-BMM, an off the shelf gaze event classifier [9]. Then, saccades were discarded (during these, drivers are moving their eyes between fixation points and cannot pay attention to the point of regard). Finally, fixations were aggregated into attention maps by initializing a Gaussian distribution centered at each fixation point and aggregating these across a window of gaze history. This eye gaze was obtained in the form of 3D gaze coordinates in the virtual world, allowing us to project the gaze point-of-regard to virtual cameras in the world (such as the *left, center, right* images taken in as input by LBC).

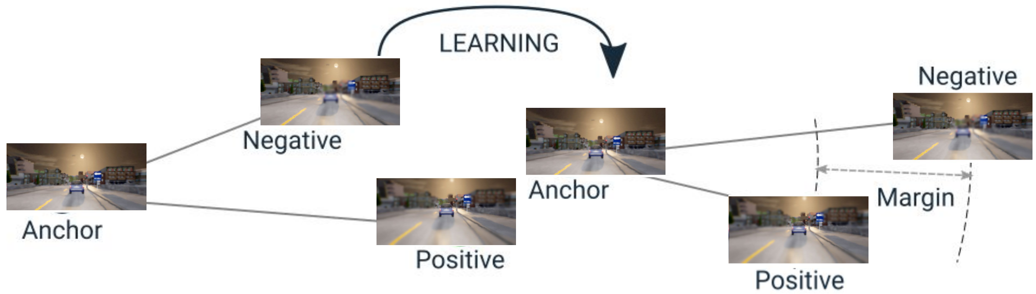


Figure 3: Gaze-based supervision via triplet loss. Input data points are represented with center image but all three images are correspondingly blurred. See Fig 4 for the full input triplet in greater detail.

Figure 4: Example triplet in more detail with gaze-contingent blur applied. In order (top-bottom): Anchor inputs (*left, center, right* image - no blur), Negative image (blur in attention regions), Positive image (blur in non-attention regions)

We collected data from $N = 7$ drivers, all of whom had held a US driver's license for more than one calendar year. Each participant drove 5 routes, with the first being for acclimatization to the VR simulator (this data was not used). However, 3 participants were unable to complete all 4 routes due to motion-sickness in the simulator and 4 routes had to be discarded due to improper data recording. In total, we have 17 routes with about 4 minutes of driving data each or about 70 minutes of data total. This is much lower than the auto-generated data used to train the LBC models, which is upwards of 300 minutes.

### 3.2 Gaze-based supervision via contrastive loss

As Fig. 1 shows, a large portion of the blur-based salience lies on the base of the traffic light – i.e. when blur is applied to this region, this causes the largest change in the LBC model's predicted output.

Our simple idea to provide gaze supervision comes from mitigating this misplaced salience. We use a triplet loss and gaze based salience as shown in Fig. 3. In our formulation, an unblurred set of input images (*left, center, right*, waypoint) constitutes the anchor data point. The negative input is constructed by applying Gaussian blur (same parameters as [4]) to important scene locations (indicated by attention maps) in the same set of images. The corresponding positive point has the same blur applied to the unimportant scene regions (complement of attention maps). The reasoning for this formulation is as follows: the most important regions for decision making for actions lie in the gazed-at regions (as indicated by attention maps) and the non-gazed at regions do not contain information that would change the driving decision. The triplet loss minimizes the distance of the anchor and the positive point while maximizing the distance of the anchor and the negative point. Hence, our loss enforces that visual inputs that are blurred in locations that are unimportant to driving should lead to a smaller change in network output than the same blur applied in important regions. An example triplet is shown in Fig. 4.

### 3.3 Fine-tuning details

As explained in the paper [5], LBC training takes place in two steps: first, by learning a privileged agent that learns to drive with perfect sensor information and then, by using it to supervise a sensorimotor that learns to ”see” via RGB images. In this work, we focus primarily on mitigating causal confusion in the sensorimotor model since that is the one that learns the task with sensory inputs (and greater potential for causal confusion) and because it is the final deployed model. We also focus on fine-tuning the sensorimotor model rather than training from scratch since the amount of data with gaze-supervision is much lower than the auto-generated driving data.

In this work, we show results of fine-tuning the sensorimotor agent using either the driving supervision loss used by the LBC authors [6] (LBC) or via the gaze based triplet loss (Triplet).

In models that use the LBC loss, the choice of privileged model to use as the teacher to the sensorimotor model during training is an important area of consideration. This teacher model could either be trained using the same data as the corresponding sensorimotor model (self-trained) or we could simply use the best performing privileged agent model released by the authors of LBC (LBC best). We show preliminary results using both types of teacher models in Table 1.

### 3.4 Quantitative evaluation

To evaluate our fine-tuned models, we used the validation set from the CARLA leaderboard benchmark [10]. This set contains 26 driving routes spread over 3 virtual towns, of which 1 is unseen in the training data. We use the $DrivingScore$ metric from the Carla leaderboard which is calculated as the average of $RouteCompletionPercentage \times InfractionScore$ per route. Here $RouteCompletionPercentage$ is the percentage of the route completed by the driving agent and $InfractionScore$ is a number in $[0, 1]$ that encapsulates the number of infractions committed by the driving agent. $InfractionScore$ starts at 1 for each route and is progressively decreased per infraction (we refer readers to [10] for details). Hence, the maximum achievable $DrivingScore$ would be 100.

| Model Type | Sensorimotor weights | Teacher model | Training data | Loss used | DS (↑) |
|---|---|---|---|---|---|
| pre-trained LBC [5] | self-train | LBC best | RBE | LBC | 16.4 |
| human data only (scratch) | self-train | self-train | DRVR | LBC | 1.68 |
| human data only | LBC best | LBC best | DRVR | LBC | 4.93 |
| human + expert mix (scratch) | self-train | self-train | RBE + DRVR | LBC | 7.05 |
| human + expert mix | LBC best | LBC best | RBE + DRVR | LBC | 20.28 |
| gaze-based triplet only | LBC best | N/A | DRVR | Triplet | 5.41 |

Table 1: Preliminary experimental results on the Carla AD leaderboard val set. Abbreviation guide: LBC best best teacher model provided by LBC authors [5]; RBE - demonstrations from rule based expert; DRVR - demonstrations from human drivers in the DReyeVR simulator

## 4 Discussion & Future work

From our preliminary results, the first noticeable trend is driving performance degradation due to fine-tuning on solely the human demonstrator driving data (DRVR). This may be due to the much smaller size of the DRVR data compared to that auto-generated by the rule-based expert (RBE). However, using mixed RBE + DRVR training data using the LBC loss does give much better driving performance. In the fine-tuning case, this even outperforms the pre-trained LBC model.

Promisingly, finetuning the sensorimotor model using triplet loss leads to better performance than both training from scratch and finetuning using the LBC driving loss. However, this is ongoing work and several paradigms need to be explored before being able to make useful conclusions from this line of research inquiry.

First, the RBE data does not contain associated gaze data, hence we cannot use RBE+DRVR data directly with the gaze-based triplet loss. However, we do plan to investigate the efficacy of mixed RBE+DRVR data in a custom training regime where we also mix the LBC and Triplet losses. In this paradigm, the LBC supervision loss would be calculated using all of the data but the triplet loss would only be calculated for samples which have associated gaze data with them.

Further, in the construction of triplets for gaze supervision, deleting objects that are not gazed at is a more direct way of enforcing their absence in the positive sample than simply blurring them out. This deletion could be done at the simulator level while generating training data, or via partial convolutions to block out certain image regions as in [11]. We would like to explore more explicit construction of triplets of this manner, in the future.

Finally, in addition to quantitative evaluation via driving scores, we plan to qualitatively evaluate the mitigation of causal confusion by investigating the stop/start issue described in Sec. 2.

### Acknowledgments

This work was funded in part by the National Science Foundation (IIS-1900821), the Tang Family AI Innovation Fund, the Link Foundation, and Bosch. This work has partly taken place in the Personal Autonomous Robotics Lab (PeARL) at The University of Texas at Austin. PeARL research is supported in part by the NSF (IIS-1749204, IIS-1925082), AFOSR (FA9550-20-1-0077), and ARO (78372-CS, W911NF-19-2-0333).

## References

[1] P. De Haan, D. Jayaraman, and S. Levine. Causal confusion in imitation learning. *Advances in Neural Information Processing Systems*, 32, 2019.

[2] D. Wang, C. Devin, Q.-Z. Cai, P. Krähenbühl, and T. Darrell. Monocular plan view networks for autonomous driving. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2876–2883. IEEE, 2019.

[3] F. Codevilla, E. Santana, A. M. López, and A. Gaidon. Exploring the limitations of behavior cloning for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9329–9338, 2019.

[4] S. Greydanus, A. Koul, J. Dodge, and A. Fern. Visualizing and understanding atari agents. In *International conference on machine learning*, pages 1792–1801. PMLR, 2018.

[5] D. Chen, B. Zhou, V. Koltun, and P. Krähenbühl. Learning by cheating. In *Conference on Robot Learning*, pages 66–75. PMLR, 2020.

[6] Learning-by-cheating carla challenge 2020 submission, . Accessed Sept 20, 2022. https://github.com/bradyz/2020_CARLA_challenge/.

[7] Causal confusion in learning-by-cheating github issue, . Accessed Sept 20, 2022. https://github.com/bradyz/2020_CARLA_challenge/issues/16.

[8] G. Silvera, A. Biswas, and H. Admoni. Dreyevr: Democratizing virtual reality driving simulation for behavioural & interaction research. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '22, page 639–643. IEEE Press, 2022.

[9] E. Tafaj, G. Kasneci, W. Rosenstiel, and M. Bogdan. Bayesian online clustering of eye movement data. In *Proceedings of the symposium on eye tracking research and applications*, pages 285–288, 2012.

[10] Carla leaderboard. Accessed Sept 20, 2022. https://github.com/carla-simulator/leaderboard.

[11] C. Li, S. H. Chan, and Y.-T. Chen. Who make drivers stop? towards driver-centric risk assessment: Risk object identification via causal inference. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10711–10718. IEEE, 2020.