Supplemental Material to: Advisable Learning for Self-driving Vehicles by Internalizing Observation-to-Action Rules

Jinkyu Kim, Suhong Moon, Anna Rohrbach, Trevor Darrell, and John Canny
EECS, University of California, Berkeley
{jinkyu.kim, suhong.moon, anna.rohrbach, trevordarrell, canny}@berkeley.edu

Content

This supplementary material provides details on our evaluation on the simulated environment called CARLA [3] (Section 1) and our human evaluations (Section 2). We also provide the dataset details (Section 3).

1. Evaluation on Simulated Environment

Our driving model is based on the work by NVIDIA [1] and Codevilla *et al*. [2] where they successfully deployed a ConvNet to drive in real-world scenarios. Our model generally outperforms prior work in control prediction. However, to further evaluate the model, we migrate our model from the offline training to a simulated environment, called CARLA [3]. We observe that the use of semantic segmentation as the internal representation of visual scenes is helpful in transferring between real-world and simulated setting. We first train our model on the BDD-X dataset [4] and evaluate in the CARLA simulator (version: 0.9.6). Note, that we use a PID controller to perform lateral and longitudinal control in the simulator from our control commands output. We also use the Robot Operating System (ROS) for the message passing of segmentation, detection, control nodes. We consider the following four typical driving scenarios: (a) Stopping at red traffic lights, (b) Stopping at red traffic lights in a heavy rain, (c) Stopping at a stop road marking, and (d) Stopping for a jaywalker, see Figure 1 (a)–(d). We then provide the driving model with the following advice: "the light is red", for scenarios (a) and (b) above, "there is a stop sign", for scenario (c), and "there is a pedestrian crossing", for scenario (d).

2. Human Evaluation

In this human evaluation, users are observing a driving model under the following three conditions.

- **Case 1** (non-explainable model): User only observes the car's behavior.

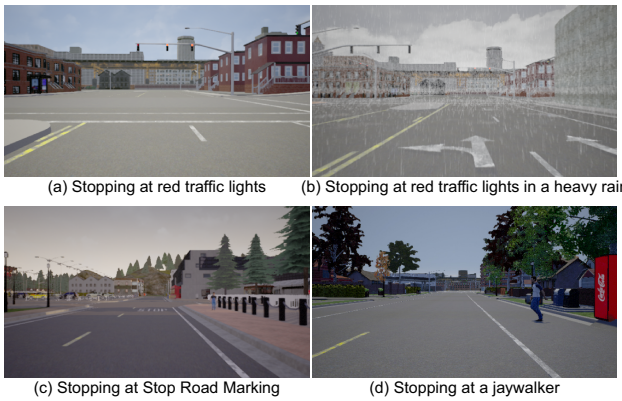


Figure 1: Four driving scenarios where we run our driving model in the CARLA [3] simulator.

- **Case 2** (explainable w/ attention and textual explanations): User observes the model's behavior along with the pixel-level attention and textual explanations.
- **Case 3** (explainable w/ human-to-vehicle advice): User observes the model's behavior, attention and textual explanations, before and after providing advice.

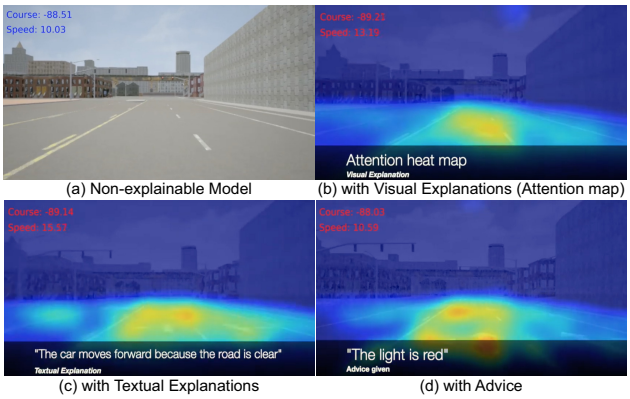
We illustrate each case in Figure 2 (a)–(d).

User Pool. The 20 human evaluators were recruited online. We required them to have (i) English language proficiency, (ii) familiarity with the US driving rules, and (iii) minimal driving experience. 10 responses were collected for each case. We split these human evaluators equally into two groups **A** and **B**. Group **A** observed Case 1 above, while Group **B** observed Case 2 and 3.

Questionnaire. The following questions were used to measure how the users trust the system.

(Q.1) Can you briefly describe why the system has failed?

(Q.2) How confident are you about it?



(a) Non-explainable Model (b) with Visual Explanations (Attention map)

(c) with Textual Explanations (d) with Advice

Figure 2: Screenshots of how the users observe the car's behavior with (a) a non-explainable model, (b-c) an explainable model with visual attention and textual explanations, and (d) an explainable model with human-to-vehicle advice.

(Q.3) How much do you trust this driving system?

Note that Q.1 is a open-ended question, where the users are allowed to response in open text format, while for Q.2 and Q.3 the users provide ratings on the Likert scale from 1 to 5. Figure 3 shows the questionnaire we used.

3. Dataset Details

We used BDD-X dataset [5], which is composed of over 77 hours of driving within 6,984 videos. The videos are taken in diverse driving conditions, e.g. day/night, highway/city/countryside, summer/winter etc. On an average of 40 seconds, each video contains around 3-4 actions, e.g. stopping, speeding up, slowing down, turning right etc., all of which are annotated with a description and an explanation. BDD-X dataset contains over $26K$ activities in over $8.4M$ frames. We used the same training/validation/test splits as provided by [5]. Following Xu *et al*. [6], we filtered out data for the following cases: (i) invalid course and speed measurements, (ii) invalid timestamps, and (iii) missing log measurements. Finally, we used the training/validation/test splits, containing 5116, 642, and 629 videos, respectively.

As we explained in our main paper, our vehicle controller predicts a future trajectory $\mathcal{P} = [p_{t,\Delta}, p_{t,2\Delta}, \ldots, p_{t,N\Delta}]$ along with speed $\hat{v}_t$. Each point $p_{t,j\Delta}$ for $j = \{1, 2, \ldots, N\}$ is characterized by its future longitudinal and latitudinal location after the time $j\Delta$. We estimate such a future trajectory from IMU sensor measurements (i.e. vehicle's speed and course). To this end, we project future agent motion onto the current facing direction (i.e. course) and compute relative poses. For sensor logs that are not synchronized with the time-stamps of video data, we use the (linearly) interpolated measurements.

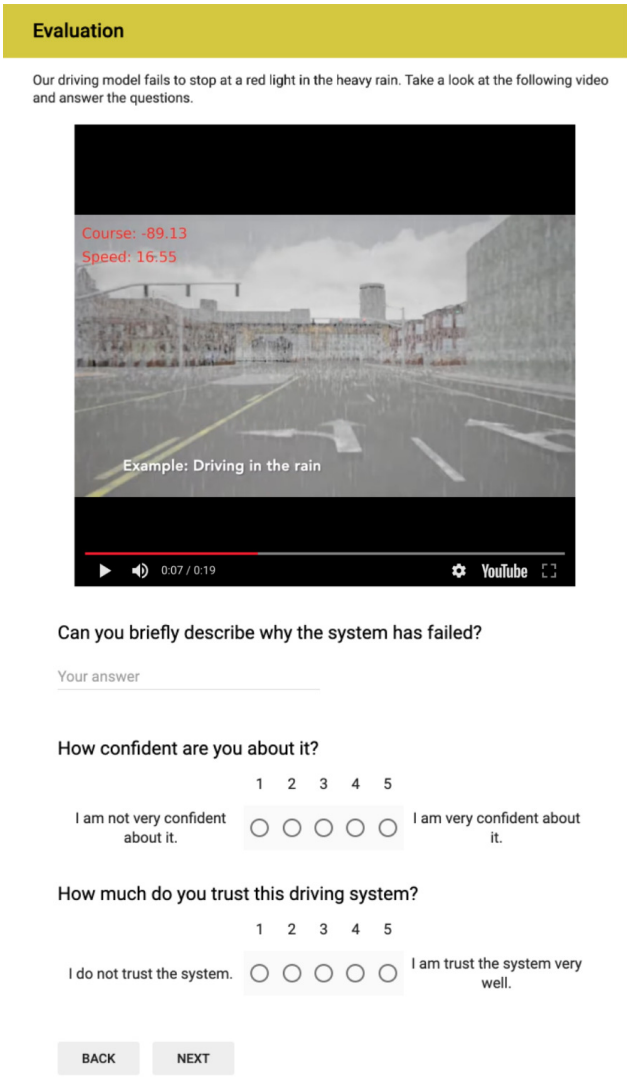


Figure 3: Our interface for the human evaluation.

References

[1] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, et al. End to end learning for self-driving cars. *CoRR abs/1604.07316*, 2016. 1

[2] Felipe Codevilla, Matthias Miiller, Antonio López, Vladlen Koltun, and Alexey Dosovitskiy. End-to-end driving via conditional imitation learning. In *ICRA*, pages 1–9. IEEE, 2018. 1

[3] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. Carla: An open urban driving simulator. *CoRL*, 2017. 1

[4] Jinkyu Kim, Terihusa Misu, Yi-Ting Chen, Ashish Tawari, and John Canny. Grounding human-to-vehicle advice for self-driving vehicles. *CVPR*, 2019. 1

[5] Jinkyu Kim, Anna Rohrbach, Trevor Darrell, John Canny, and Zeynep Akata. Textual explanations for self-driving vehicles. In *ECCV*, 2018. 2

[6] Huazhe Xu, Yang Gao, Fisher Yu, and Trevor Darrell. End-to-end learning of driving models from large-scale video datasets. In *CVPR*, 2017. 2