Supplementary Document: Look Both Ways: Self-Supervising Driver Gaze Estimation and Road Scene Saliency

Isaac Kasahara[1], Simon Stent[2], and Hyun Soo Park[1]

[1] University of Minnesota, USA, {kasah011, hspark}@umn.edu
[2] Toyota Research Institute, Cambridge, MA, USA, simon.stent@tri.global

Here we provide additional results and details of our method that may be of use to the reader. We refer interested readers to also check out our code and dataset at https://github.com/Kasai2020/look_both_ways.

1 Additional Qualitative Results

In Fig. 6 of the main paper, we included qualitative results of drivers who were driving naturally without wearing gaze tracking glasses. In Fig. 1 below, we include results of the same drivers with and without the glasses for further qualitative assessment.

From qualitative inspection, we believe that our method was able to generalize reasonably well on new subjects both with and without gaze tracking glasses. Unfortunately, due to the lack of ground truth data without the glasses, we cannot provide quantitative results for the drivers when not wearing glasses.

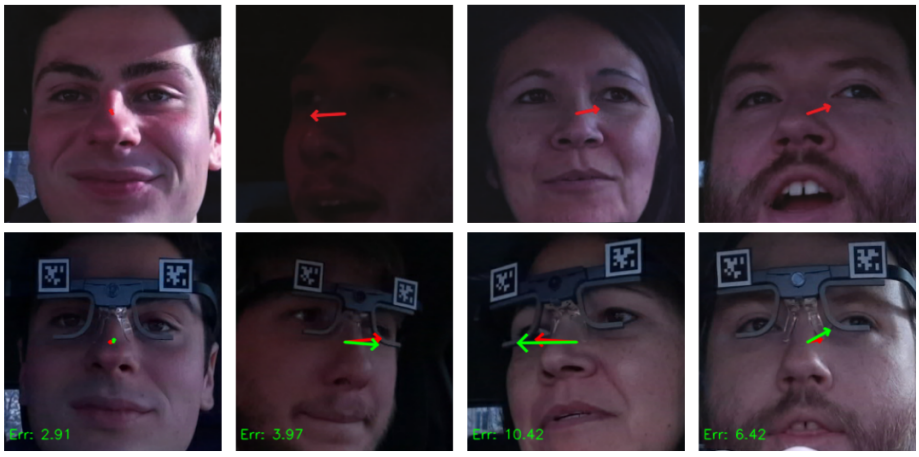

Fig. 1. Qualitative results of the subjects with and without gaze tracking glasses glasses.

2 Additional Experiment Results

In Table 2 of the main paper, we separated our dataset into supervised/self/test sections by driver, meaning the test split contained data from subjects not seen in the training data. This helps to assess model generalization to new drivers. Here, we further explored using a self-supervised split and test split that contained the same drivers, but a different driving session with unseen data. This scenario may be useful in situations where the model is pre-trained on other drivers, and then refined on a particular subject without the use of ground-truth data to allow for the model to be specific to each driver. Table 1 below shows our model performance using this alternative data split. We see an expected increase in

Method	Self	Test
Supervised-only	9.0	9.6
Ours	8.5	9.1

Table 1. Performance on a new data split in MAE_g.

angle accuracy on the self-supervised data split, but also now a larger increase in angle accuracy on the test data split as well.

3 Dataset Synchronization

The inward facing face camera and the outward facing stereo cameras were synchronized using a hardware trigger provided by the manufacturer. The gaze tracking glasses were synchronized by showing a stop-watch to both the camera systems at the start of the recording and synchronizing manually afterwards. Due to occasional frame drops between the two camera systems over time, we estimate synchronization error of up to 200ms in theory. We recognize this may affect the accuracy of the ground truth labels during rapid eye movement of the drivers.

4 Calibration Between 3D and Color Camera for 3D eye center

In order to project the 2D eye location into 3D space for our dataset, additional calibration was needed between the inward facing RGB and inward facing depth camera. We used a built-in function from the Azure Kinect's SDK for this purpose (`k4a_transformation_depth_image_to_color_camera()`). This transformed the depth image into the color camera coordinate system. Then, we could proceed with using the color camera's relative location along with the

transformed depth image to project the 2D eye location into 3D. Our 2D eye centers were obtained using OpenPose facial recognition that obtains keypoints of the driver's face. If OpenPose had above a 0.5 confidence value for the eye center, it was used to project from 2D into 3D relative to the inward facing camera.