Supplementary Material for De-confounded Gaze Estimation

Ziyang Liang[1,2], Yiwei Bao[1], and Feng Lu[1,2*]

[1] State Key Laboratory of VR Technology and Systems, School of CSE, Beihang University, Beijing, China
[2] Peng Cheng Laboratory, Shenzhen, China
{liangziyang, baoyiwei, lufeng}@buaa.edu.cn

1 Analysis of The Separated Features

To further analyze the performance of the FSM and the semantic information learned by the three features: head pose feature $\boldsymbol{H}$, gaze feature $\boldsymbol{G}$, and gaze-irrelevant feature $\boldsymbol{I}$, we look into the outputs of the three auxiliary tasks themselves. Their output errors/accuracy are 6.61°, 3.21° and 100% on $\mathcal{D}_E$, and 12.72°, 6.42° and 100% on $\mathcal{D}_G$, respectively, which can be considered quite good. This partially demonstrates the validity of our learnt features of $\boldsymbol{H}$, $\boldsymbol{G}$ and $\boldsymbol{I}$.

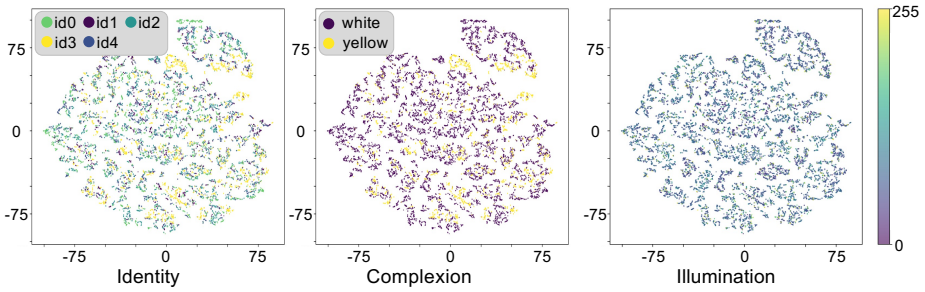

Fig. 1: T-SNE visualizations of gaze-irrelevant feature $\boldsymbol{I}$ on $\mathcal{D}_E$. From left to right, the coloring is based on the person's identity, complexion, and the image's brightness in the sample.

Besides the above analysis, we visualized the embeddings of gaze-irrelevant feature $\boldsymbol{I}$ using T-SNE to verify that $\boldsymbol{I}$ did not collapse into a meaningless value, as shown in Fig. 1.

2 Further Ablation study on Gaze-irrelevant Feature

To further verify that the performance improvement of FSCI is not due to the introduction of additional head pose loss and to validate the effectiveness of the

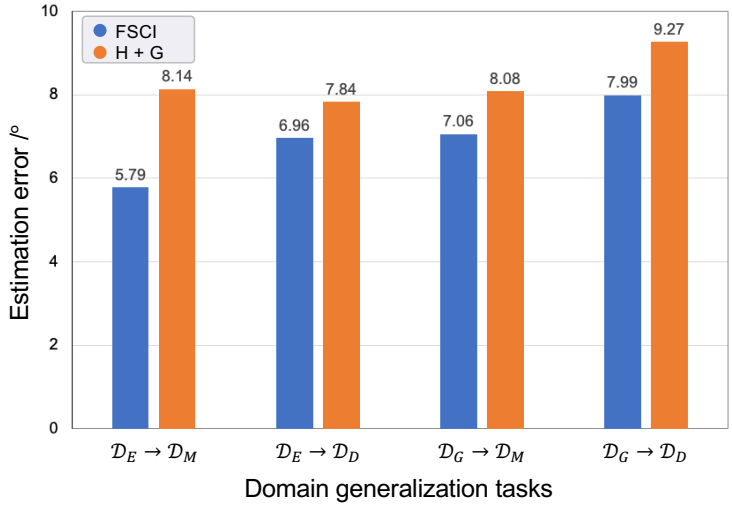

Fig. 2: Further ablation study on gaze-irrelevant feature $\boldsymbol{I}$. $\boldsymbol{H} + \boldsymbol{G}$ represents using only $\boldsymbol{H}$ and $\boldsymbol{G}$ to predict gaze direction $\boldsymbol{g}$.

causal intervention module, we use the features $\boldsymbol{H}$ and $\boldsymbol{G}$ extracted by FSCI to predict the gaze direction $\boldsymbol{g}$ solely, yielding the results of [$\mathcal{D}_E \rightarrow \mathcal{D}_D$: 7.84°, $\mathcal{D}_E \rightarrow \mathcal{D}_M$: 8.14°, $\mathcal{D}_G \rightarrow \mathcal{D}_D$: 9.27°, $\mathcal{D}_G \rightarrow \mathcal{D}_M$: 8.08°]. As shown in Fig. 2, those errors are $\sim 1°$ lager than our final results. This proves the effectiveness for the CIM as above mentioned, and more importantly, shows that the performance improvement of FSCI is not mostly due to the extra head pose auxiliary loss.