EVA-GCN: Head Pose Estimation Based on Graph Convolutional Networks (*Supplementary Material*)

Miao Xin*
Institute of Automation (CASIA)
Chinese Academy of Sciences
Beijing, China
miao.xin@ia.ac.cn

Shentong Mo†
Carnegie Mellon University
Pittsburgh, United States
shentonm@andrew.cmu.edu

Yuanze Lin†
Beihang Univerity
Beijing, China
linyuanze@buaa.edu.cn

1. Landmark locations

The landmark labels in this work follow the classical 68-point landmark label strategy [2, 1]. The landmark number is shown in Figure 1(a). The final selected landmarks (Figure 1(b)) are No. 1, 3, 8, 13, 15, 17, 21, 22, 26, 31, 33, 35, 36, 39, 42, 45, 48, 51 and 54.

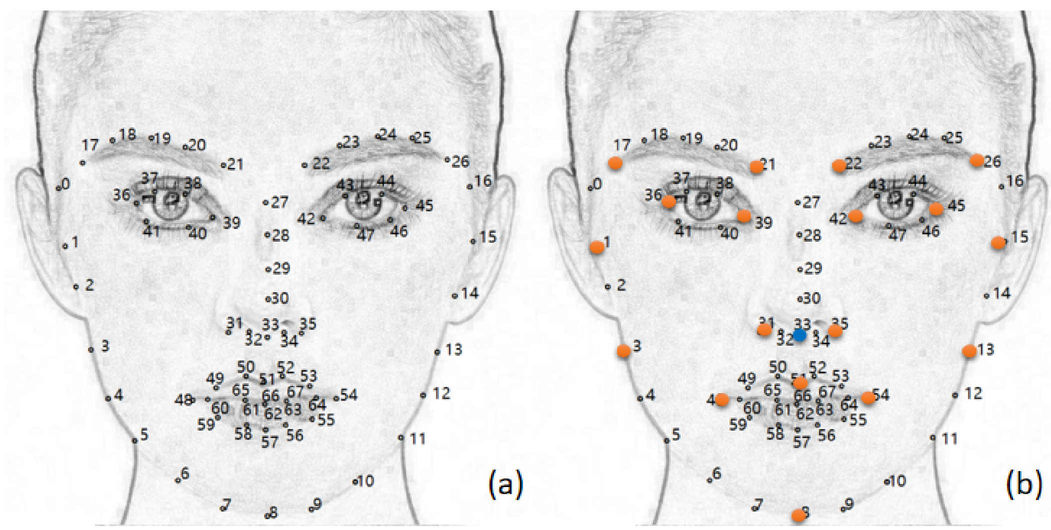

Figure 1. (a) Landmark label strategy and (b) the selected 19 landmarks.

The high-resolution figures *w.r.t.* the facial landmarks' saliency and stableness are shown in Figure 2 and Figure 3. The boxes in the figure denote different facial parts of faces. The comparison is within one facial part.

2. An implement for low-power devices

We test the network on an edge computing device with an ARM CPU (2×A72 + 4×A53 cores, 1.8GHz, Figure 4). To better utilize the computing resource, we implement an end-to-end network with CNN-based landmark detection and EVA-GCN-based head pose estimation. The CNN part in this model is an EfficientNet-B0 [4]. It regresses 19 heatmaps which locate the 19 landmarks. These location vectors form the input of the EVA-GCN using the same approach as stated in [3]. The EfficientNet is articulated to the EVA-GCN. This end-to-end model achieves 45 FPS on the edge computing device, and the MAE is 4.89 on AFLW2000 dataset [5].

*Corresponding author.
†Shentong Mo and Yuanze Lin were interns at CASIA.

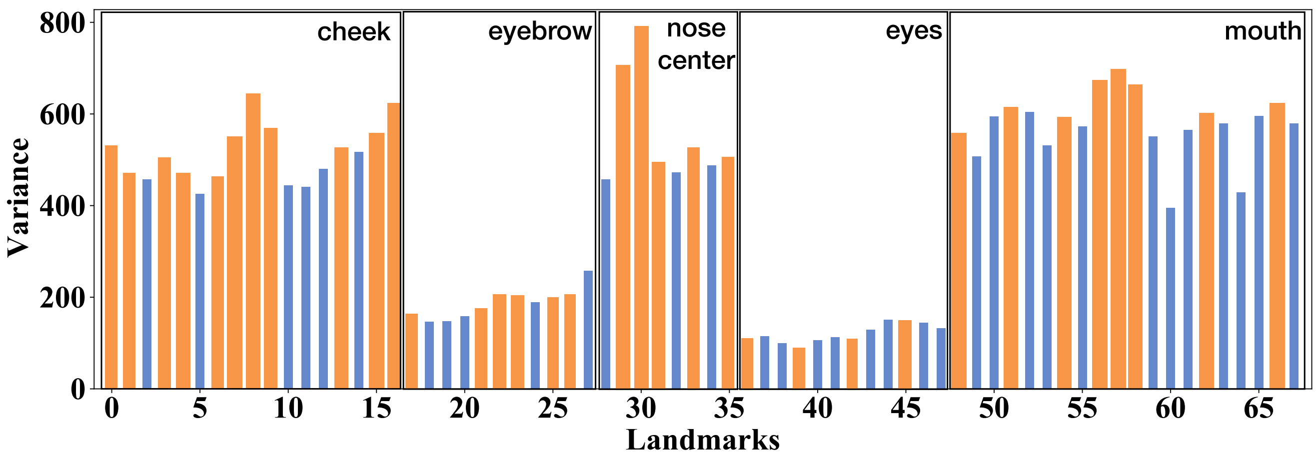

Figure 2. Comparison results *w.r.t.* the landmarks' saliency.

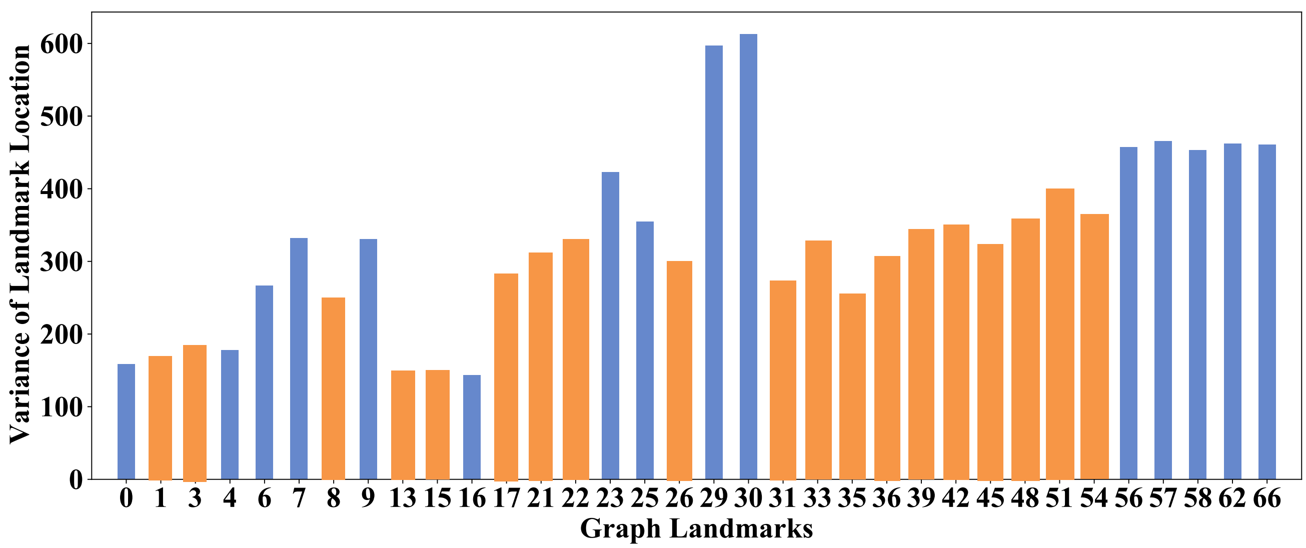

Figure 3. Comparison results *w.r.t.* the landmarks' stableness.

References

[1] Adrian Bulat and Georgios Tzimiropoulos. How far are we from solving the 2d and 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In *ICCV*, 2017.

[2] Hang Du, Hailin Shi, Dan Zeng, and Tao Mei. The elements of end-to-end deep face recognition: A survey of recent advances. *arXiv preprint arXiv:2009.13290*, 2020.

[3] Xiao Sun, Bin Xiao, Fangyin Wei, Shuang Liang, and Yichen Wei. Integral human pose regression. In *ECCV*, pages 529–545, 2018.

[4] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *ICML*. PMLR, 2019.

[5] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li. Face alignment across large poses: A 3d solution. In *CVPR*, 2016.

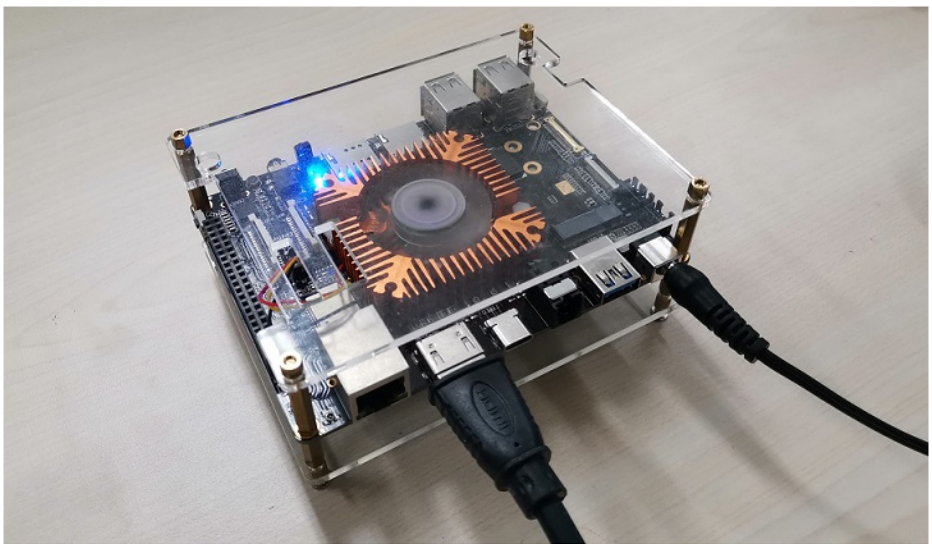

Figure 4. The edge computing device: ARM CPU (2×A72 + 4×A53 cores, 1.8GHz).