

SO-HandNet: Self-Organizing Network for 3D Hand Pose Estimation with Semi-supervised Learning (Supplementary Material)

Yujin Chen¹, Zhigang Tu^{1*}, Liuhaog Ge², Dejun Zhang⁴, Ruizhi Chen¹, Junsong Yuan³

¹Wuhan University, Wuhan, China

²Nanyang Technological University, Singapore

³State University of New York at Buffalo, Buffalo, NY, USA

⁴China University of Geosciences, Wuhan, China

{yujin.chen, tuzhigang, ruizhi.chen}@whu.edu.cn ge0001ao@e.ntu.edu.sg

zhangdejun@cug.edu.cn jsyuan@buffalo.edu

1. More Details about Implementation

The proposed network consists of three modules, including the hand feature encoder (HFE), the hand feature decoder (HFD) and the hand pose estimator (HPE). We adopt a stage-wise training strategy to train the network.

We use Adam [1] with a weight decay of 0.0005 to optimize the network. Instead of directly optimizing the whole network, we pre-train each part of the architecture and then holistically optimize the whole network. Firstly, we train the HFE and the HFD with initial learning rate 0.001 and batch size 18. The learning rate is divided by 2 every 10 epochs. This stage is stopped after 30 epochs. Secondly, we optimize the pre-trained HFE and the HPE with initial learning rate of 0.00001 and batch size 40. We reduce the learning rate by half every 5 epochs. This stage is stopped after 30 epochs. Finally, we optimize the whole network together with initial learning rate of 0.00001 and batch size 18. The learning rate is divided by 10 after 30 epochs and the training process is stopped after 50 epochs.

2. Result Visualization

We study the impact of semi-supervised learning on ICVL dataset. With the same network architecture, we conduct a semi-supervised training experiment which uses 25% of the data with annotation and the rest 75% data without annotation from the training set. We also conduct a contrast experiment which only uses 25% of the annotated data. We show some results from the top view and the front view of the normalized hand point cloud in Figure 1. As shown in this figure, our method using semi-supervised learning obtains overall more accurate 3D hand pose than that without using semi-supervised learning.

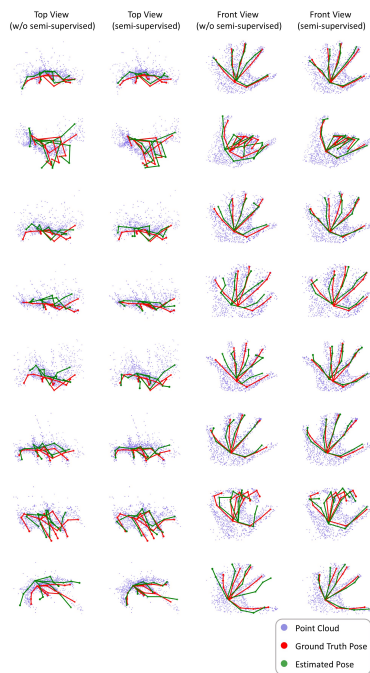


Figure 1. Hand pose estimation results. First column: the top view of estimation results that without semi-supervised learning. Second column: the top view of estimation results that with semi-supervised learning. Third column: the front view of estimation results that without semi-supervised learning. Last column: the front view of estimation results that with semi-supervised learning.

References

- [1] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

*Corresponding author: tuzhigang@whu.edu.cn