

Point-to-Pose Voting based Hand Pose Estimation using Residual Permutation Equivariant Layer

Supplementary Material

Shile Li
Technical University of Munich
li.shile@mytum.de

Dongheui Lee
Technical University of Munich, German Aerospace Center
dhlee@tum.de

1. Comparison of different network structures

Besides the experiment listed in the main paper, we also performed self-comparison at the initial exploration stage to verify the advantage of ResPEL network. From Table 1, the incremental improvements of residual version and voting scheme can be shown. 4PELs is a shallow network with four PELs with (64,256,512,1024) hidden neurons and Maxpooling means the voting scheme is replaced by a max-pooling layer and 4 consequent fully-connected layers to regress the hand pose. By replacing the shallow network with residual version (ResPEL+Maxpooling), the pose error is reduced significantly. Further improvement can be achieved by using our proposed voting scheme (ResPEL+Voting).

	4PELs+ Maxpooling	ResPEL+ Maxpooling	ResPEL+ Voting
mean joint error [mm]	16.60	11.02	8.35

Table 1. Self-comparison of different network structures on NYU dataset (three views)