Learning Semantic Representation via Joint 3D Face Reconstruction and Facial Attribute Estimation
Zichun Weng, Youjun Xiang, Xianfeng Li, Wanliang Huo, Juntao Liang and Yuli Fu

Overview
- Joint learning with two tasks: 3D face reconstruction (3DFR) and Facial Attribute Estimation (FAE)
- Semantic facial representations for both tasks

Related Works
- Fully-supervision / self-supervised: lack of feature explanation
- RingNet (CVPR-19): feature consistency constrained by Triple loss
- Liu et al. (CVPR-18): feature consistency constrained by Face Recognition loss
- Ours: FAE has more explicit correlation with 3DFR than Face Recognition

Joint Framework
- Joint learning with two tasks: 3D face reconstruction (3DFR) and Facial Attribute Estimation (FAE)
- Semantic facial representations for both tasks

Qualitative Results
- Semantic facial representation in the feature space:

Quantitative Results
- Comparison on Florence dataset:

Contribution
- For the first time, we train two highly relevant facial tasks, 3DFR and FAE, in a joint manner. Quantitative evaluation and qualitative visualization indicate the effectiveness and robustness of our method.
- We develop an in-batch hybrid-task training scheme that enables our model to learn from hybrid facial datasets with heterogeneous labels.
- The proposed MTL framework allows CNN to extract semantic facial representations from in-the-wild images, which are significant for unconstrained 3D face reconstruction.