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

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

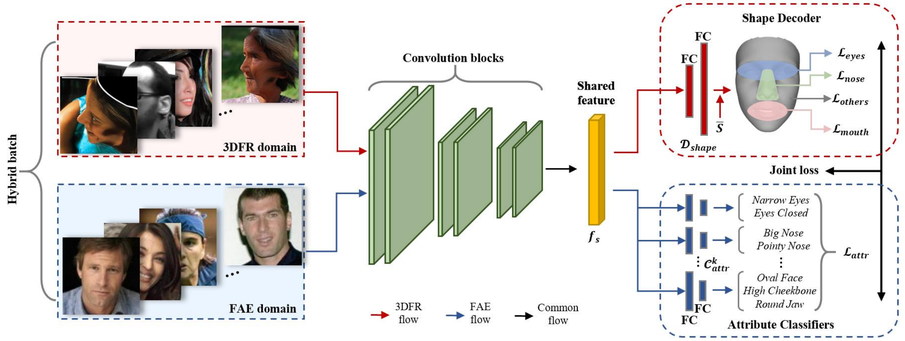
Overview

- **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



- **Losses:**

$$\mathcal{L}_{attr} = -\sum_{k=0}^{K-1}\sum_{m=0}^{M_k-1}[\alpha_m^k A_m^k \log(\hat{A}_m^k)+(1-A_m^k)\log(1-\hat{A}_m^k)] \quad \text{(3D Reconstruction Loss)}$$

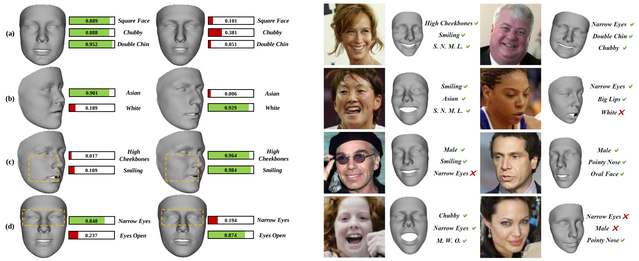
$$\mathcal{L}_{shape} = \frac{1}{N}\sum_l \mathcal{L}_l \cdot W_l = \frac{1}{N}\sum_l \left\| S_l - \hat{S}_l \right\|_2^2 \cdot W_l \quad \text{(Facial Attribute Loss)}$$

Related Works

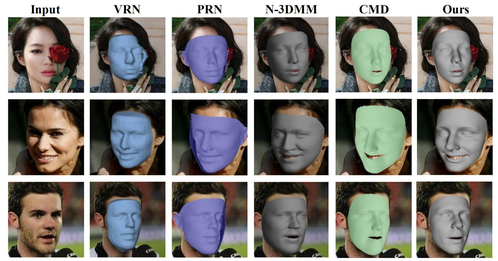
- **Fully-supervision / self-supervised:** lack of feature explanation
- **RingNet (CVPR-19'):** feature consistency constrained by Triple loss
- **Liu *et al.* (CVPR-18'):** feature consistency constrained by Face Recognition loss
- **Ours:** FAE has more explicit correlation with 3DFR than Face Recognition

Qualitative Results

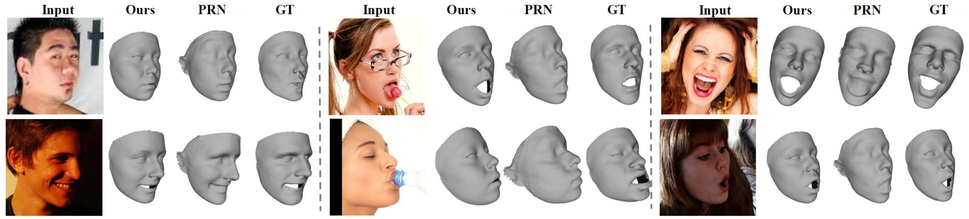
- **Semantic facial representation in the feature space:**



- **Comparison on CelebA dataset:**



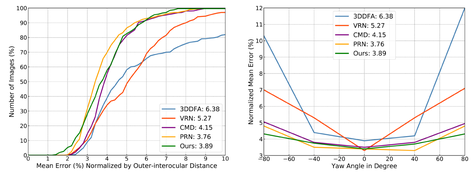
- **Comparison with PRNet and ground truth on AFLW2000 dataset:**



(a) Race (b) Occlusion (c) Extreme expression

Quantitative Results

- **Comparison on Florence dataset:**



(a) CED curves (b) Pose-specific NME

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.