Learning Semantic Representation via Joint 3D Face Reconstruction and Facial Attribute Estimation

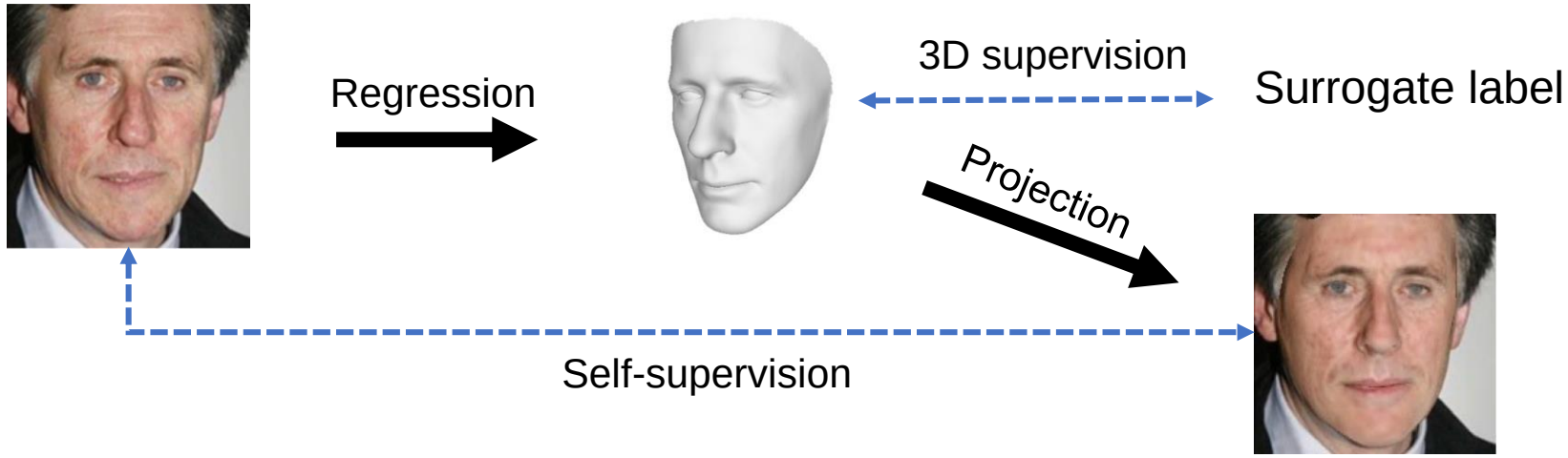
Zichun Weng, Youjun Xiang, Xianfeng Li, *et al.*

South China University of Technology

→ 3D Face Reconstruction

- Supervised by surrogate label
- Self-supervision

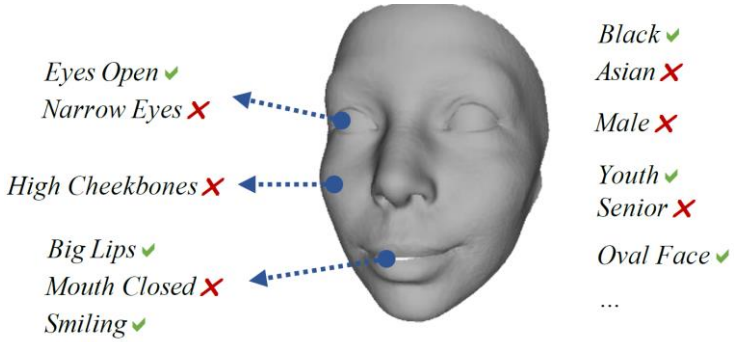
→ Lack of semantic meaning
In the latent space

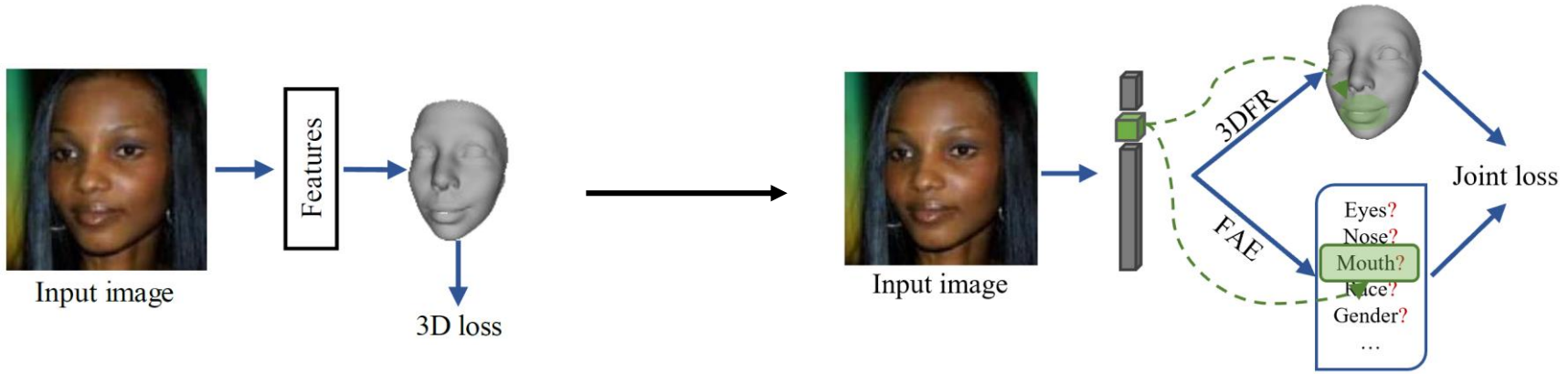


Lin et al. CVPR 2020

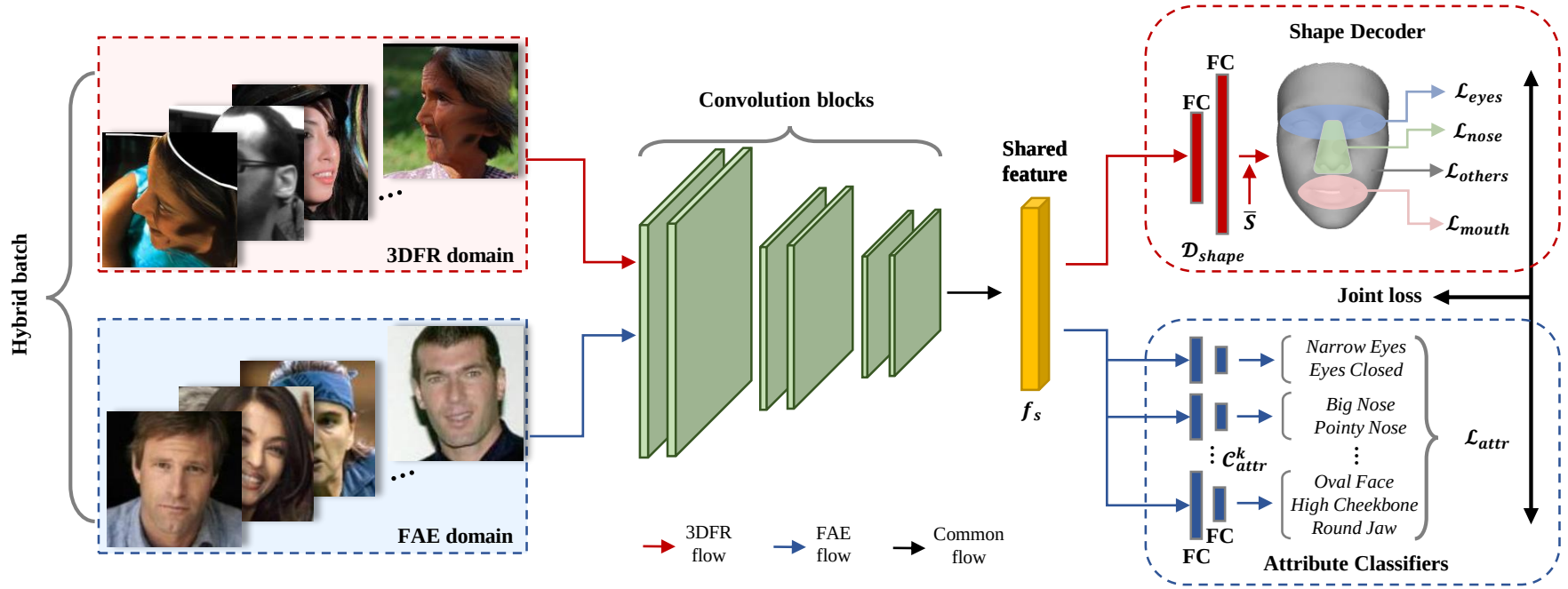
→ Motivation

- 3D face reconstruction (3DFR)
- Facial attribute Estimation (FAE)
- Joint training
- Shared facial representation





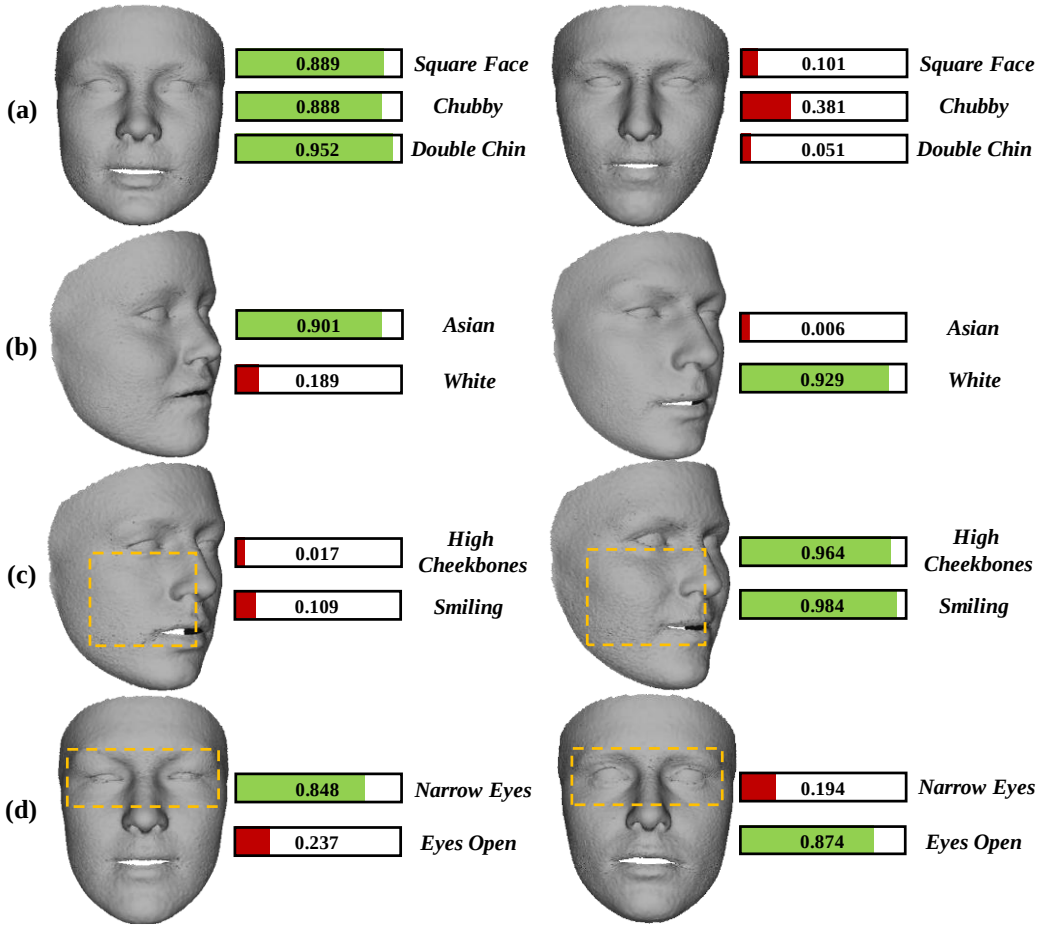
→ Our Joint Framework



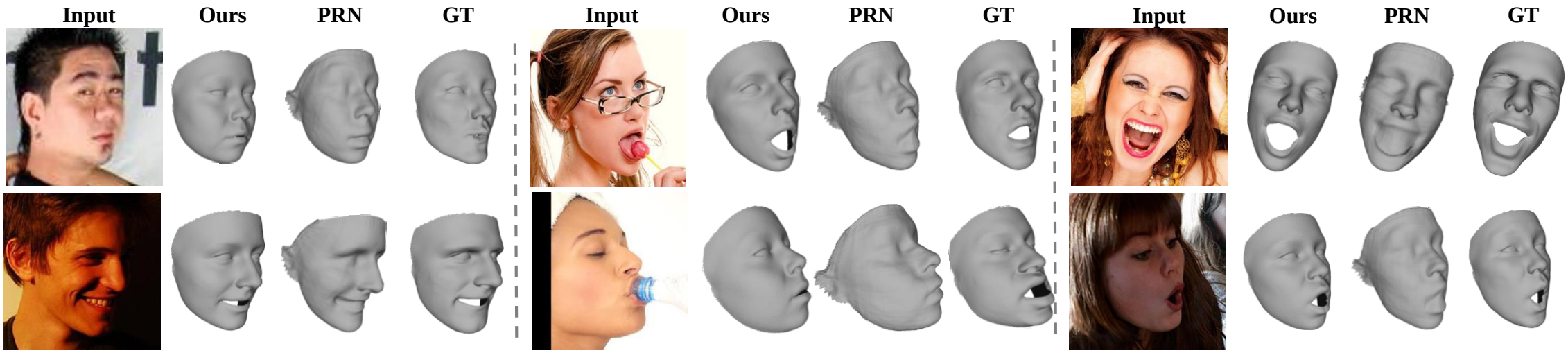
- Advantages:
  - Extracting semantic facial representations for high-fidelity reconstruction
  - Multi-task learning avoids overfitting in either task

→ Semantic Representation for Input Face

- Shape and attributes deviation: changing one dimension in the feature space
- “disentanglement”
- Fine-grained shape manipulation: cheekbones / eyes



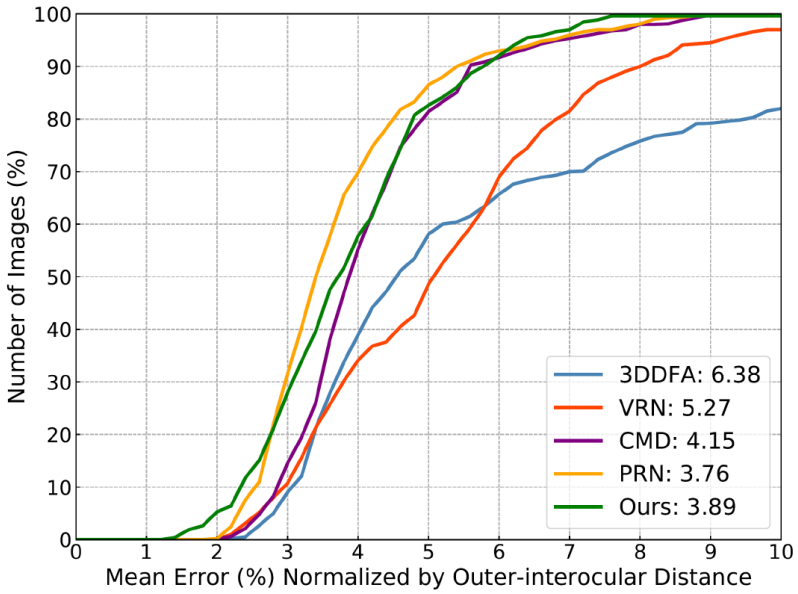
→ Comparison on AFLW2000 Dataset



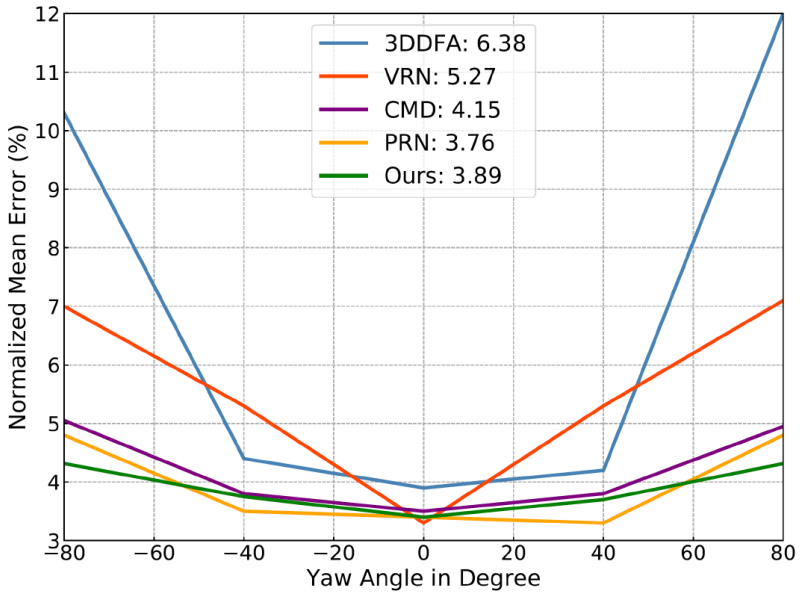
(a) Race (b) Occlusion (c) Extreme expression

- High-fidelity reconstruction with accurate feature extraction
- Robustness across extreme conditions
- Sharper but more reasonable shapes

→ Comparison on Florence Database



(a) CED curves



(b) Pose-specific NME

3DDFA: Zhu et al. CVPR 2016
VRN: Jackson et al. ICCV 2017
CMD: Zhou et al. CVPR 2019
PRN: Feng et al. ECCV 2018

→ Conclusion

- **Joint framework:**
  - Semantic representation
  - Accuracy
  - Robustness

- **Future works:**
  - Disentanglement in feature space
  - Training with scanned 3D data

Thank you!