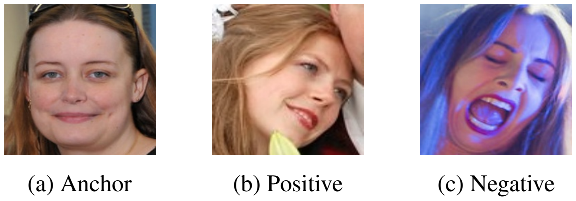
Learning a Facial Expression Embedding Disentangled from Identity Supplementary Material

Wei Zhang[1]*, Xianpeng Ji[1]*, Keyu Chen[2]*, Yu Ding[1]*, Changjie Fan[1]
[1]Virtual Human Group, Netease Fuxi AI Lab
[2]University of Science and Technology of China
{zhangwei05,jixianpeng,dingyu01,fanchangjie}@corp.neatease.com
cky95@mail.ustc.edu.cn

Example of FEC dataset

We demonstrate a triplet example from the FEC [4] dataset to illustrate the similarity between Anchor, Positive, and Negative images. As observed, Anchor is more similar to Positive and more different from Negative.



(a) Anchor (b) Positive (c) Negative

Figure 1: An example of FEC dataset.Anchor and Positive are the more similar images compared with Negative.

Implementation Details

We have described the implementation details for FEC [4] triplet prediction in the main text. Network parameters of our model are 55.93M, compared to FECNet with 223.48M. Our flop is 6.32GMac and FECNet is 8.18GMac. Here, we will supplementally describe the implementation details for applications of emotion recognition and face manipulation.

For emotion recognition, we use the Deviation Learning Network (DLN) without the crowd layer as backbone, and a Softmax layer, in replacing of the layer that produces the expression embedding, to calculate the occurrence probabilities of 7 emotions. The DLN is trained on the AffectNet [3] training set and RAF-DB [2] training set for 10 epochs with a learning rate of 0.001 using SGD optimizer with a momentum of 0.9. The batch size is set to 30. The DLN is supervised by the weighted cross-entropy loss.

For face manipulation, we make use of the pix2pixHD Network [5] and train it by Adam with a learning rate of 2×10^{-4} for Generator and Discriminator. The pipeline can be seen in Fig. 2. We use the supervised data of RaFD [1] for generating manipulated faces during training while we generate faces manipulated by different identities' expression when testing.

References

[1] Oliver Langner, Ron Dotsch, Gijsbert Bijlstra, Daniel HJ Wigboldus, Skyler T Hawk, and AD Van Knippenberg. Presentation and validation of the radboud faces database. *Cognition and emotion*, 24(8):1377–1388, 2010.

[2] Shan Li and Weihong Deng. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. *IEEE Transactions on Image Processing*, 28(1):356–370, 2019.

[3] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10(1):18–31, 2017.

[4] Raviteja Vemulapalli and Aseem Agarwala. A compact embedding for facial expression similarity. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5683–5692, 2019.

[5] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018.

*Equal contribution. Yu Ding is the corresponding author.

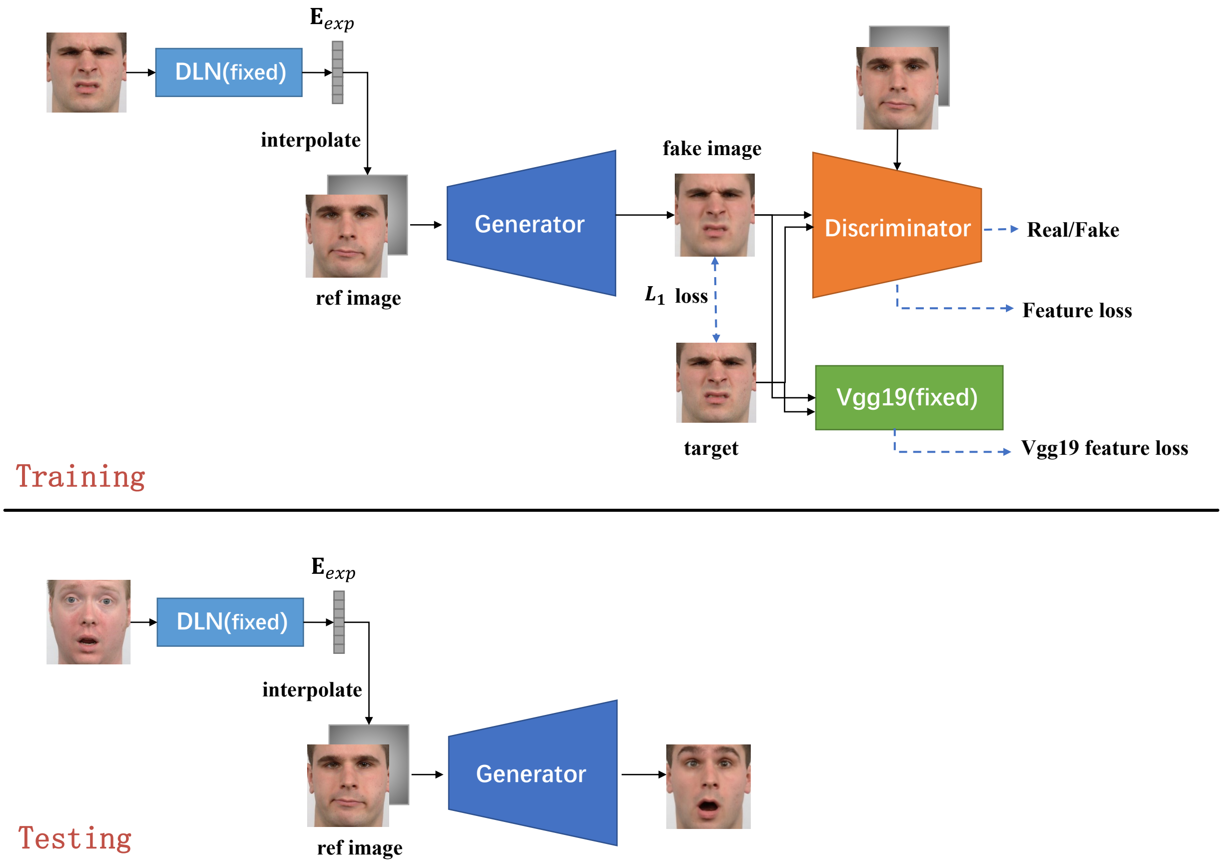


Figure 2: Conditional GAN for face manipulation. Please refer to [5] for more details about the Generator and Discriminator.