FaceInpainter: High Fidelity Face Adaptation to Heterogeneous Domains

Paper ID 5722

1. Broader Impact

As for face editing and face swapping, there are potential threats to the public, *e.g.*, political threats, disinformation attacks, identity theft, and celebrity pornography. For the purpose of preventing these harms, it is urgent to keep up with the latest deepfake technologies, and develop state-of-the-art detection algorithms base on more "faithful" deepfake data. Our contribution to the community mainly is developing an IGFI framework to implement face swapping under heterogeneous domains, which can be used to collect more swapped face data for better deepfake detection. Moreover, the IGFI task under 3D cartoon and art drawing domains has positive meaning for cultural communication.

2. Additional Quantitative Comparison

The test set consists of 10K face images by evenly sampling 10 frames from each video clip in FaceForensics++. We use Cosface to extract the identity vectors of the generated and original frames. For each generated frame, we search for the nearest face in the original frames and check whether that face is from the correct source video. In our work, the state-of-the-art face recognition model namely Arcface is supposed to extract identity feature that disentangles with the expression and pose features.

We implement Sum of Modulus of gray Difference (SMD2) based on https://github.com/Leezhen2014/python–/blob/master/BlurDetection.py.

The Frechet Inception Distance (FID) is used to measure both quality and diversity of swapped faces. As a learning-based generation model, FaceInpainter has the best FID score compared with DeepFakes and FaceShifter. The score of the graphics-based FaceSwap is the lowest, because it applies the original source texture to the deformed 3D face model, resulting in a closer distance to the original data distribution. However, it fails on ID retrieval evaluation metric compared with other learning-based methods.

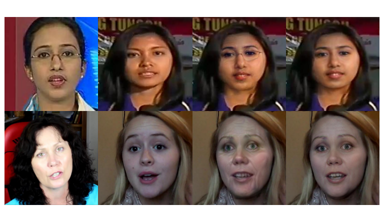

Figure 1. SFI results with the contextual loss.

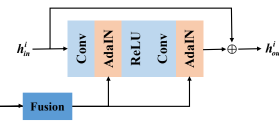

Figure 2. The network architecture of *AdaRes* block.

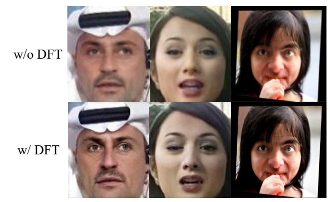

Figure 3. High-quality IGFI results with the DFT module.

Methods	DeepFakes	FaceSwap	FaceShifter	SFI-Net	FaceInpainter
Fid	14.27	**11.13**	13.27	14.22	11.99

Table 1. Fid evaluation on FaceForensics++ dataset.

3. Additional Implementation Details

In JR-Net, we warp $\hat{Y}_{t,t}$ according to 98 facial feature points of X_t based on https://github.com/thesouthfrog/stylealign, making $\triangle X_t$ focus more on the occlusion. Warping is based on Delaunay Triangulation. In the second stage, we utilize the discriminator of StyleGAN to perform the adversarial training. The architecture of *AdaRes* block is shown in Figure 2.

4. Additional Qualitative Results

Face Swapping Results. More qualitative results are shown in Figure 4. It is challenging and meaningful to generalize the model to heterogeneous domains with distinguishing texture features. Our framework can preserve attributes (expression, pose, texture of the target face) and modify corresponding identities under various heterogeneous domains.

Efficient Contextual Loss. The results without contextual loss have obvious texture distortions and additional identity feature, *e.g.* eyeglass (Figure 1 row 1). Whereas the results with contextual loss preserve better attributes. In our experiment, whether wearing glasses is supposed to be same as the target, rather than the source face.

Efficient DFT Module. As shown in Figure 3, our model with the dictionary feature transfer (DFT) module can generate higher-quality swapped faces.

Challenging Cases. As shown in Figure 4 (rows 3&5&7), there are still some artifacts in the facial area when zooming into the swapped face. It is challenging to generalize the model to some complex heterogeneous domains.

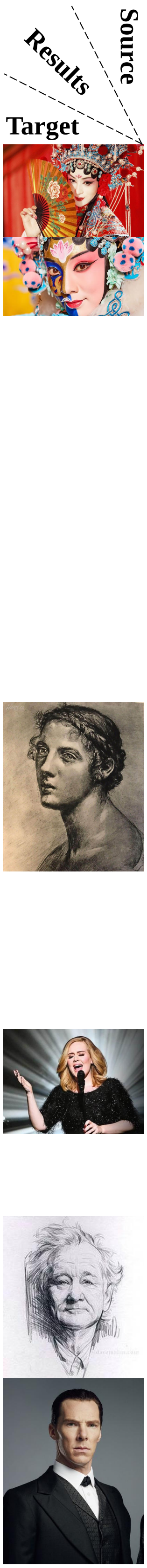

Figure 4. IGFI results based on celebrity identities under heterogeneous domains.