# Appendix

## A. Pretrain of LQ Face Parsing Network

### A.1. Network Architecture

Given a LQ face image of any size, we first upsample it to $512 \times 512$ and treat it as the input $I_L$. FPN is trained to produce a parsing map $\hat{I}_P$ and a HR face $\hat{I_H}$ that approximate the ground-truth parsing map $I_P$ and ground-truth HR face $I_H$ respectively, *i.e.*,

$$\theta_p = \arg\min_{\theta_p} \mathcal{L}_{parse}(\hat{I}_P, I_P) + \mathcal{L}_{pix}(\hat{I_H}, I_H), \quad (11)$$

where $\theta_p$ denotes the parameters of FPN, $\mathcal{L}_{parse}$ is the parsing loss, and $\mathcal{L}_{pix}$ is the pixel space L2 loss. As shown in Fig. 9, FPN adopts an encoder-resnet-decoder architecture. It begins with 4 downsample blocks, followed by 10 resnet blocks and 4 upsample blocks. Finally, two output convolution layers are used to generate $\hat{I}_P$ and $\hat{I_H}$. We adopt BatchNorm-LeakyRelu after every convolution layer.

We use multi-task learning for FPN because we found that $\mathcal{L}_{pix}$ is quite helpful for the prediction of $\hat{I_P}$. Since $I_L$ is degraded, both the pixel values and pattern of the face components are not clear and stable. The network is not able to understand the meaning of each label without the extra supervision of $I_H$. Fig. 10 shows that the parsing results with multi-task learning are much better than that without it, especially in the eyes and eyebrows.

### A.2. Datasets and Implementation

We use CelebA-Mask-HQ [19] to train FPN. The CelebA-Mask-HQ contains $30,000$ HR faces with a size of $1024 \times 1024$ selected from the CelebA [23] dataset. Each image has a segmentation mask of facial attributes corresponding to CelebA. The masks of CelebA-Mask-HQ are manually-annotated with a size of $512 \times 512$ and 19 classes including background, skin, nose, eyes (left and right), eyebrows (left and right), ears (left and right), mouth, lips (up and bottom), hair, hat, eyeglass, earring, necklace, neck, and cloth. The whole dataset is split into a training set ($24,183$ images), a validation set ($2,993$ images), and a test set ($2,824$ images). We use the training set as ground-truth HQ faces and parsing maps, and the LQ faces are generated online with Eq. 12.

We use Adam optimizer [17] to train the FPN. We set $\beta_1 = 0.9, \beta_2 = 0.999$ and learning rate to $0.0002$. The training batch size is set to 8.

## B. Degradation Model

As described in the paper, our degradation model used the following equation:

$$I_L^r = ((I_H \otimes \mathbf{k}_\varrho) \downarrow_r + \mathbf{n}_\delta)_{JPEG_q}, \quad (12)$$

where

- $\mathbf{k}_\varrho$ is the blur kernel. We randomly choose one of the following four kernels: Gaussian Blur ($3 <= \varrho <= 15$), Average Blur ($3 <= \varrho <= 15$), Median Blur ($3 <= \varrho <= 15$), Motion Blur ($5 <= \varrho <= 25$);
- $\downarrow_s$ is the downsample operation. The scale factor $r$ is randomly selected in $[\frac{32}{512}, \frac{256}{512}]$;
- $\mathbf{n}_\delta$ is the addictive white gaussian noise (AWGN) with $0 <= \delta <= 0.1 \times 255$;
- $JPEG_q$ is the JPEG operation. The compression level is randomly chosen from $[10, 65]$, in which higher means stronger compression and lower image quality.

We implement the degradation model using `imgaug`[1] library with code snippets in Fig. 11.

## C. More Results

In this section, we show more results on PSFR-RealTest and Solvay conference test. We mainly compare our model with DFDNet because they provide public codes and test models, and their results are current state-of-the-art. We also provide carefully finetuned results of PULSE on Solvay test.

### C.1. Results of PSFR-RealTest

Fig. 12, Fig. 13 and Fig. 14 show more examples from PSFR-RealTest dataset.

### C.2. Results of Solvay Conference Test

We give the overall results of the 5-th Solvay conference test images in Fig. 15. All faces are cropped out and aligned first, then enhanced by our model and finally paste back to the original photo. Complete results and detailed comparison with other methods are presented in Fig. 16, Fig. 17, Fig. 18, Fig. 19 and Fig. 20.

[1] https://github.com/aleju/imgaug

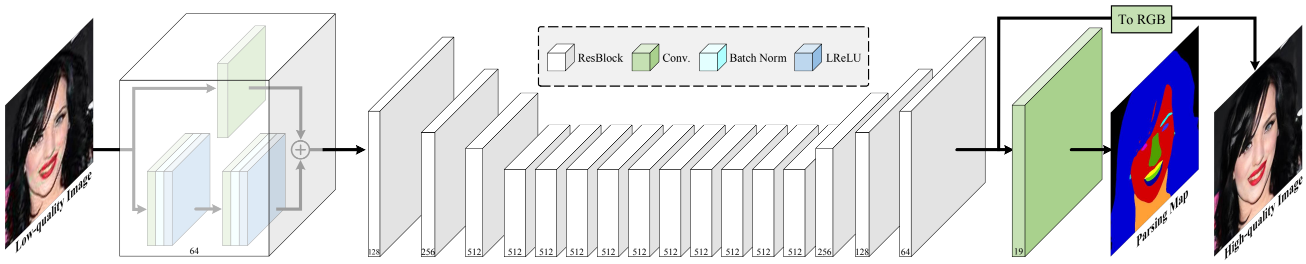



Figure 9: Architecture details of face parsing network (FPN).


Figure 10: Comparison of parsing results of natural LR faces with and without supervision of $I_H$ .

```
import imgaug as ia
import imgaug.augmenters as iaa
scale_size = random(32, 256)
org_size = 512
aug_seq = iaa.Sequential([
        iaa.Sometimes(0.5, iaa.OneOf([
            iaa.GaussianBlur((3, 15)), iaa.AverageBlur(k=(3, 15)),
            iaa.MedianBlur(k=(3, 15)), iaa.MotionBlur((5, 25))
        ])),
        iaa.Resize(scale_size, interpolation=ia.ALL),
        iaa.Sometimes(0.2, iaa.AdditiveGaussianNoise(loc=0, scale=(0.0, 0.1*255), per_channel=0.5))
                                                    ,
        iaa.Sometimes(0.7, iaa.JpegCompression(compression=(10, 65))),
        iaa.Resize(org_size),
    ])
```

Figure 11: Code snippets for degradation model.

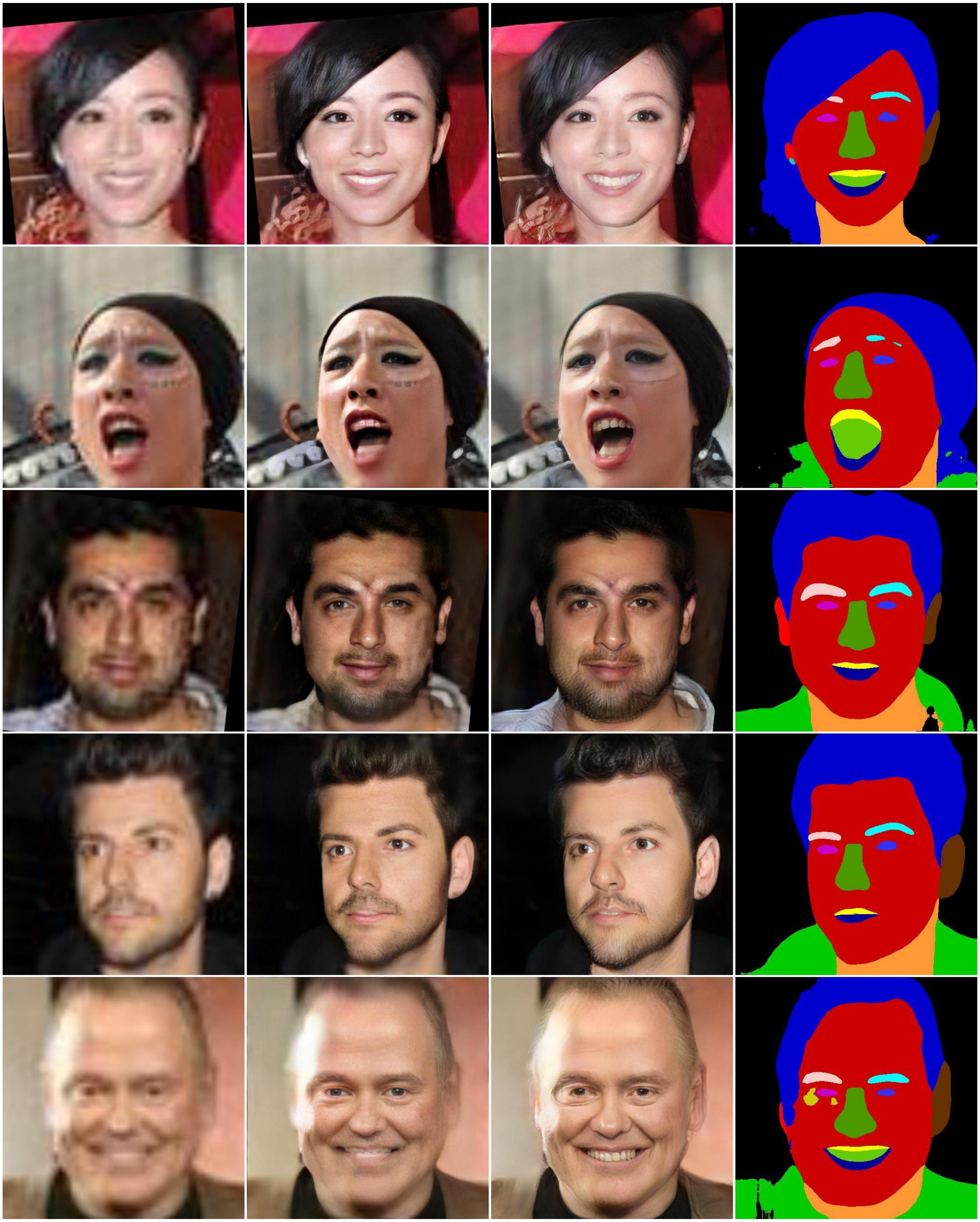
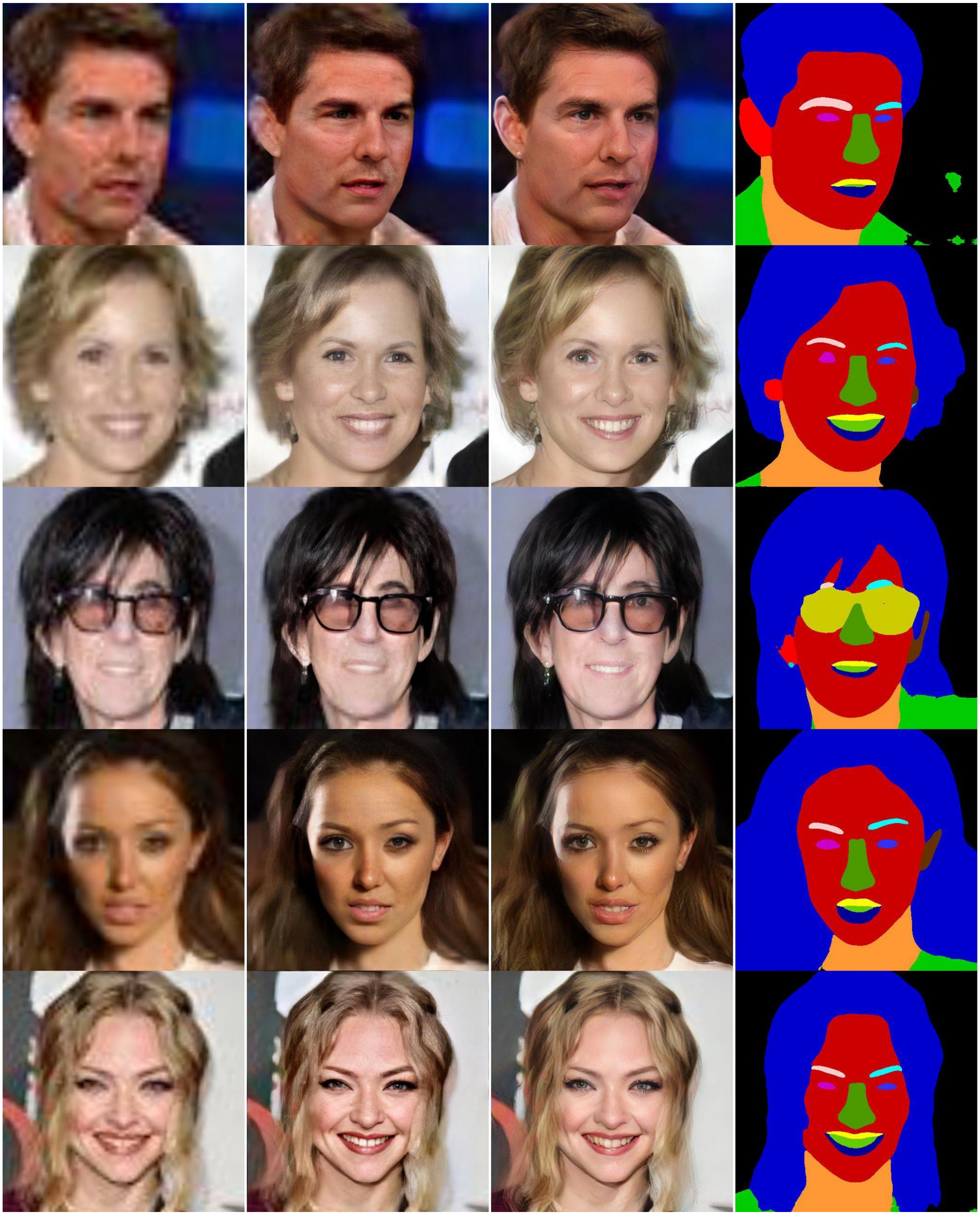



Figure 12: More results from PSFR-RealTest Dataset.



LQ | DFDNet | PSFR-GAN (ours) | Parsing Map (ours)

Figure 13: More results from PSFR-RealTest Dataset.

LQ DFDNet PSFR-GAN (ours) Parsing Map (ours)

Figure 14: More results from PSFR-RealTest Dataset.


Figure 15: Overall result of the 5-th Solvay conference taken in 1927. Please zoom in to see the details.


Figure 16: Results of 5-th Solvay conference test.


Figure 17: Results of 5-th Solvay conference test.


Figure 18: Results of 5-th Solvay conference test.

LQ | PULSE | DFDNet | PSFR-GAN (ours) | Parsing Map (ours)

Figure 19: Results of 5-th Solvay conference test.


Figure 20: Results of 5-th Solvay conference test.