Blemish-aware and Progressive Face Retouching with Limited Paired Data

Lianxin Xie[1], Wen Xue[1], Zhen Xu[1], Si Wu[1,2,3]*, Zhiwen Yu[1] and Hau San Wong[4]

[1]School of Computer Science and Engineering, South China University of Technology

[2]Peng Cheng Laboratory

[3]PAZHOU LAB

[4]Department of Computer Science, City University of Hong Kong

*{cslianxin.xie@mail.scut.edu.cn, cswusi@scut.edu.cn}

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

TUE-PM-140

Proposed Approach

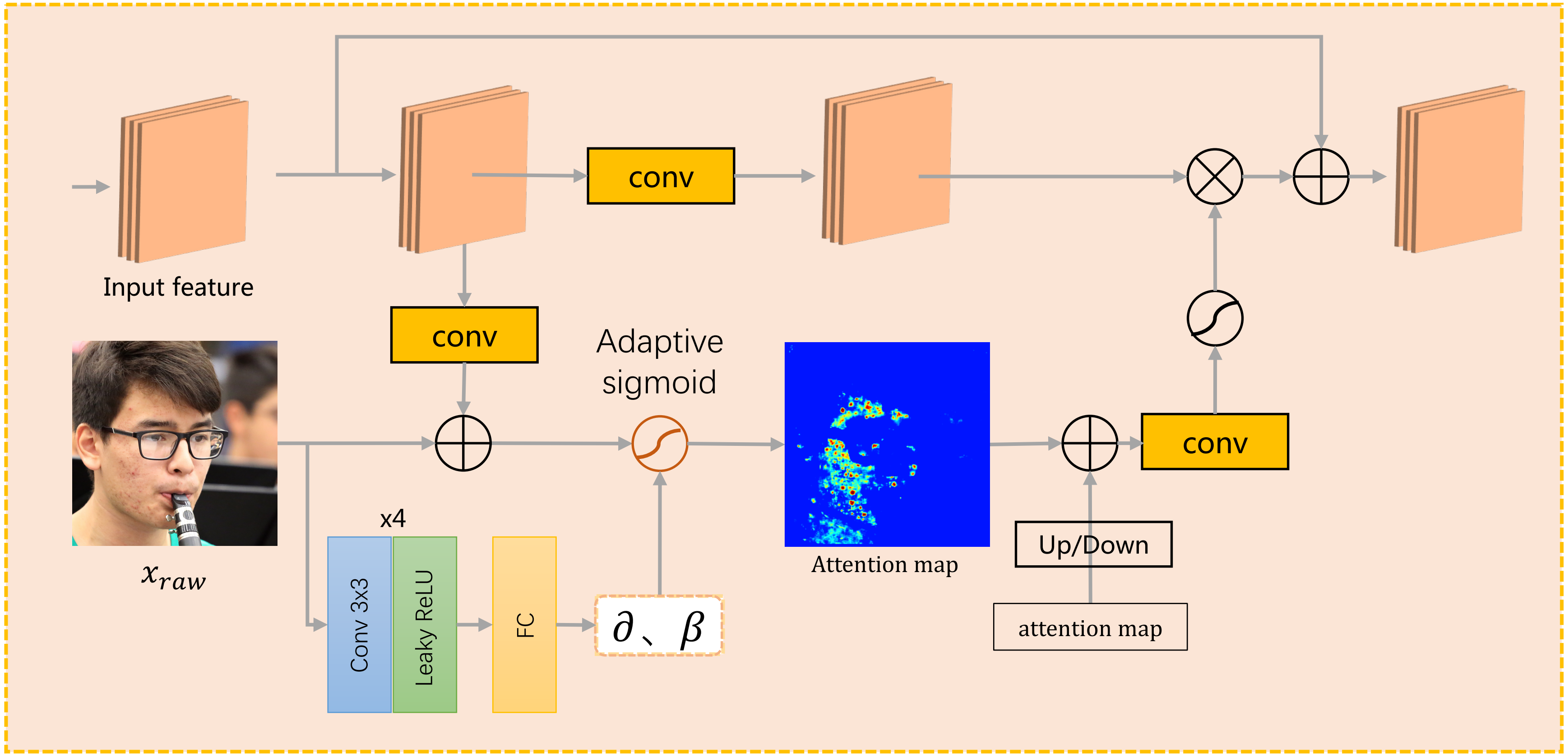
- To deal with a wide range of **facial blemishes**, we exploit the merits of both encoder-decoder and generator architectures by seamlessly integrating them into a unified framework to **progressively remove blemishes**.
- A **blemish-aware attention module** is incorporated to enhance the collaboration between the components by **refining the intermediate features** that are transferred among the components.
- We leverage **unpaired training data** to regularize the proposed framework, which effectively reduces the dependence on paired training data.

Structure of BPFRe

- An encoder-decoder architecture is applied at the first stage to perform **coarse retouching**. At the second stage, we modify the generator architecture of StyleGAN to operate on the multi-scale intermediate features of the decoder and render an image with **finer details**.
- Two **blemish-aware attention modules** are incorporated between the encoder and decoder, and between the decoder and generator, respectively.

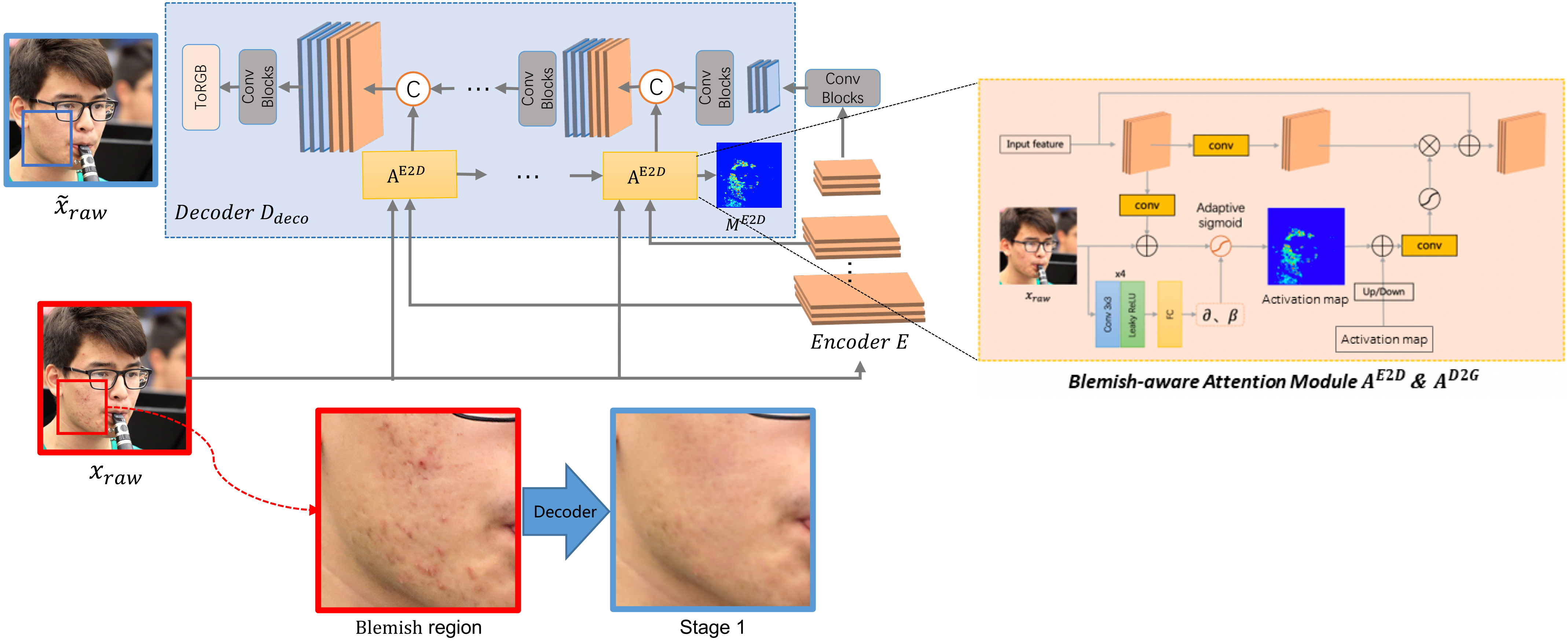
Structure of BPFRe

- Blemish-aware attention module



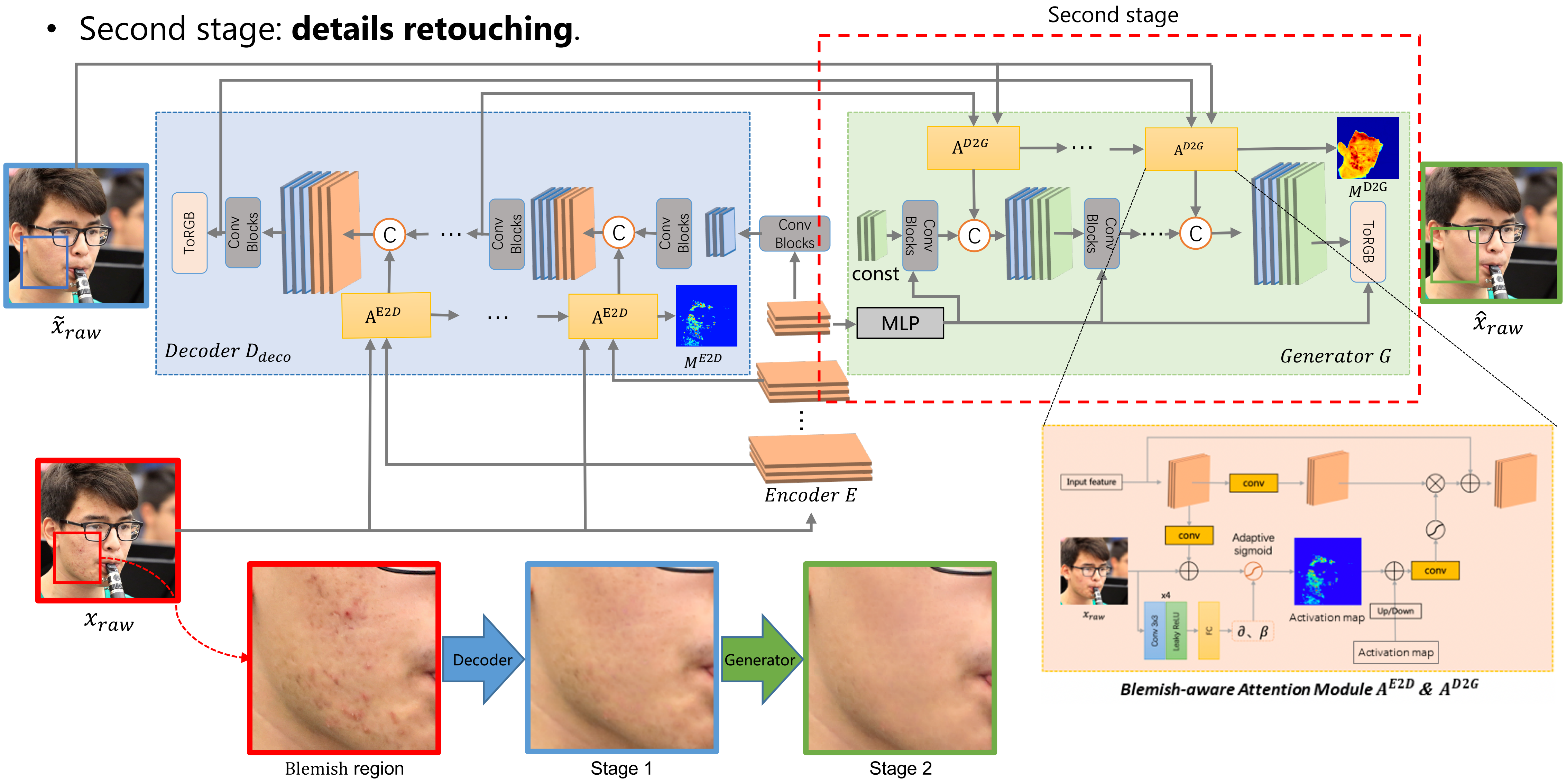
Structure of BPFRe

- First stage: **coarse retouching**.



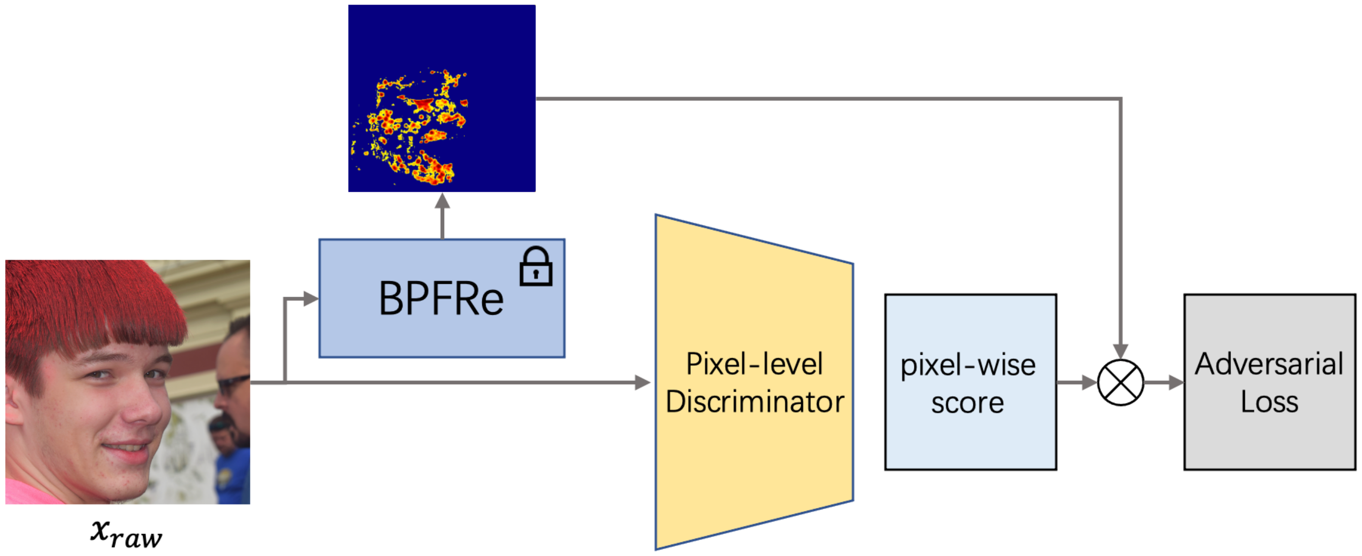
Structure of BPFRe

- Second stage: **details retouching**.



Leveraging unpaired training data

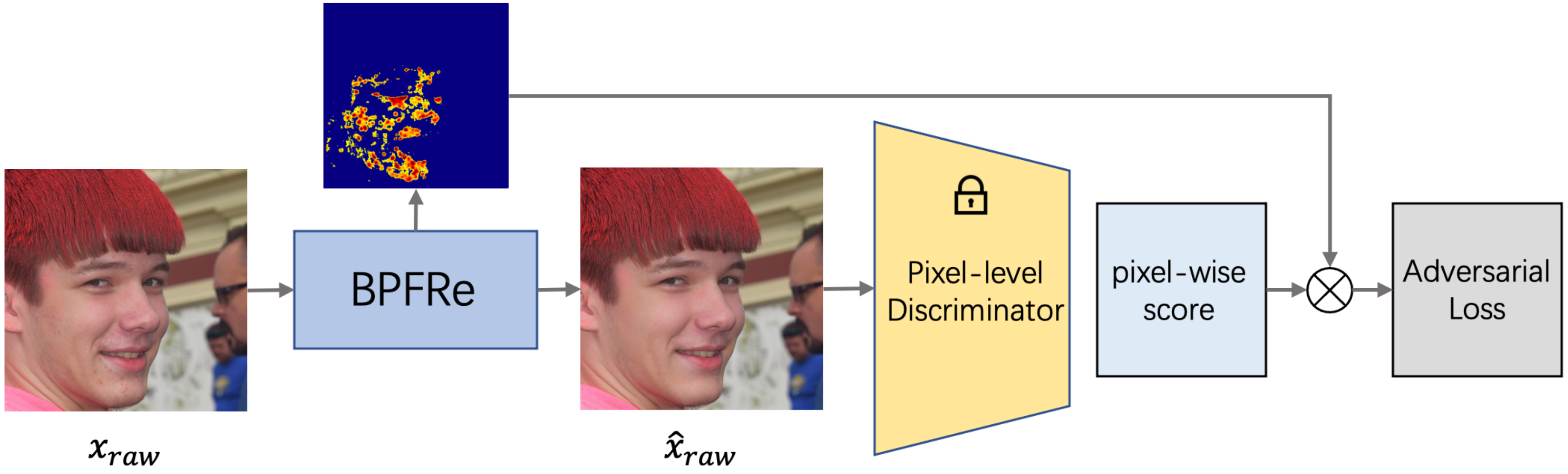
- For blemish images



Making the discriminator focus on identifying blemish regions

Leveraging unpaired training data

- For blemish images



Guiding the generator to remove blemish and deceive the discriminator

Training Loss

- Blemish-aware attention modules

$$\mathcal{L}_{\text{blem}}^{coarse} = \mathbb{E}_{x_{raw}^p}[|\mathcal{M}^{E2D} - \Delta^{coarse}|_1], \quad \mathcal{L}_{\text{blem}}^{refine} = \mathbb{E}_{x_{raw}^p}[|\mathcal{M}^{D2G} - \Delta^{refine}|_1]$$

where $\Delta^{coarse} = |x_{raw}^p - x_{ret}^p|_1$, and $\Delta^{refine} = |\tilde{x}_{raw}^p - x_{ret}^p|_1$

- Coarse retouching

$$\mathcal{L}_{\text{cons}}^{coarse} = \mathbb{E}_{x_{raw}^p}\left[|\tilde{x}_{raw}^p - x_{ret}^p|_1 + |\phi\left(\tilde{x}_{raw}^p\right) - \phi(x_{ret}^p)|_1\right] + \mathbb{E}_{x_{ret}}[|\tilde{x}_{ret} - x_{ret}|_1 + |\phi(\tilde{x}_{ret}) - \phi(x_{ret})|_1]$$

where $\tilde{x}$ represents the output of decoder D_{deco}, $\phi(\cdot)$ denotes the pretrained VGG network.

- Details retouching

$$\mathcal{L}_{\text{cons}}^{refine} = \mathbb{E}_{x_{raw}^p}\left[|\hat{x}_{raw}^p - x_{ret}^p|_1 + |\phi\left(\hat{x}_{raw}^p\right) - \phi(x_{ret}^p)|_1\right]$$

$$\mathcal{L}_{adv}^{disc} = \mathbb{E}_{x_{raw}}\left[\mathcal{M}^{D2G} \otimes log\left(1 - D_{disc}(x_{raw})\right) + \mathcal{M}^{D2G} \otimes log\left(1 - D_{disc}(\hat{x}_{raw})\right)\right] + \mathbb{E}_{x_{ret}}\left[\mathcal{M}^{D2G} \otimes \log\left(1 - D_{disc}(x_{ret})\right)\right]$$

$$\mathcal{L}_{adv}^{synt} = \mathbb{E}_{x_{raw}}\left[\mathcal{M}^{D2G} \otimes log\left(1 - D_{disc}(\hat{x}_{raw})\right)\right]$$

where $\hat{x}$ represents the output of generator G, D_{disc} denotes the pixel-level discriminator, $\mathcal{M}$ is blemish mask.

Visualization of retouching process

- Visualization of attention maps and corresponding retouching results at the two stages of BPFRe

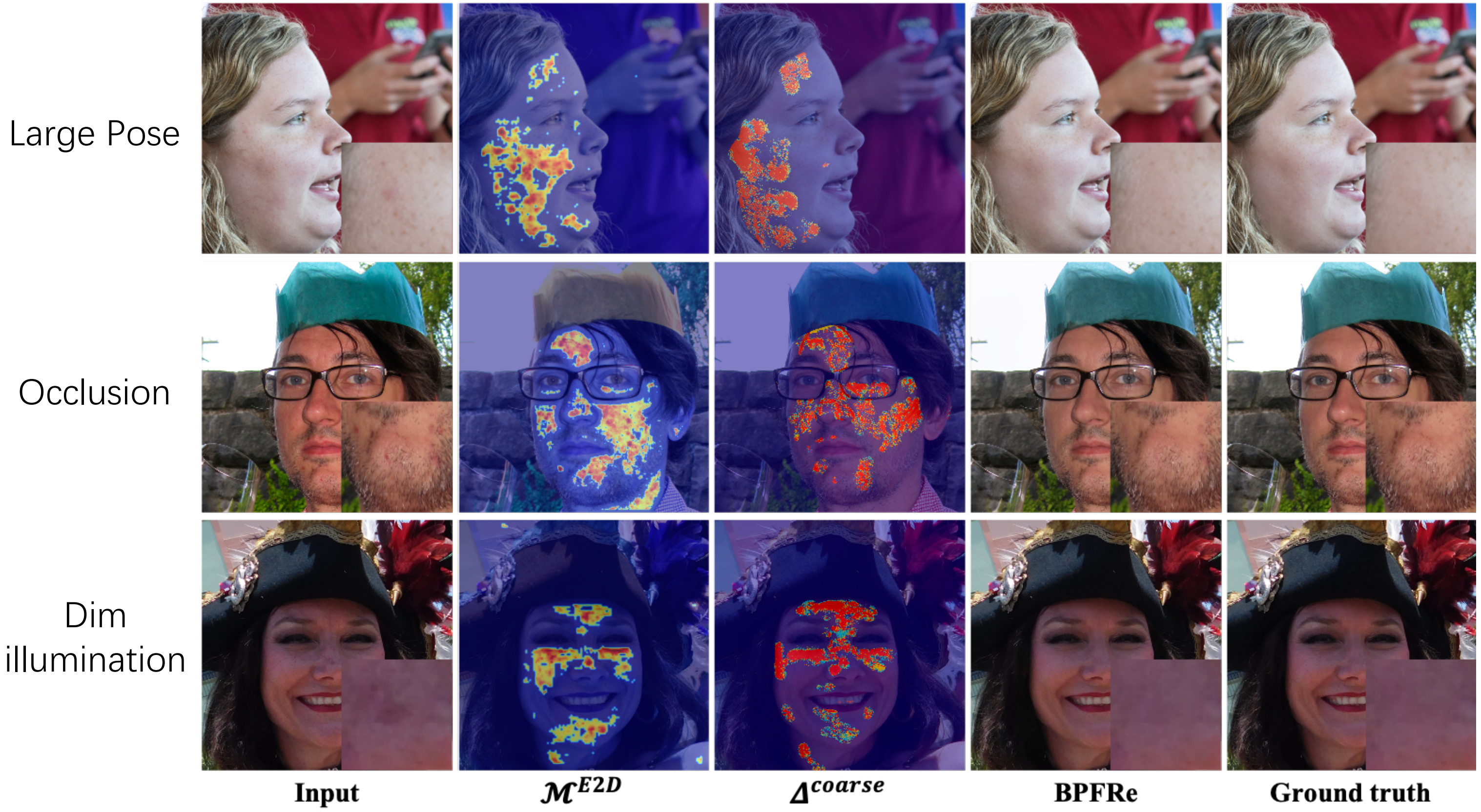
Input | $\mathcal{M}^{E2D}$ | $\mathcal{M}^{D2G}$ | Δ | $\tilde{x}_{raw}$ | $\hat{x}_{raw}$ | **Ground truth**

first stage

second stage

Hard Sample

- We evaluated BPFRe and the competing methods on the selected FFHQR images with large pose, occlusion and dim illumination.



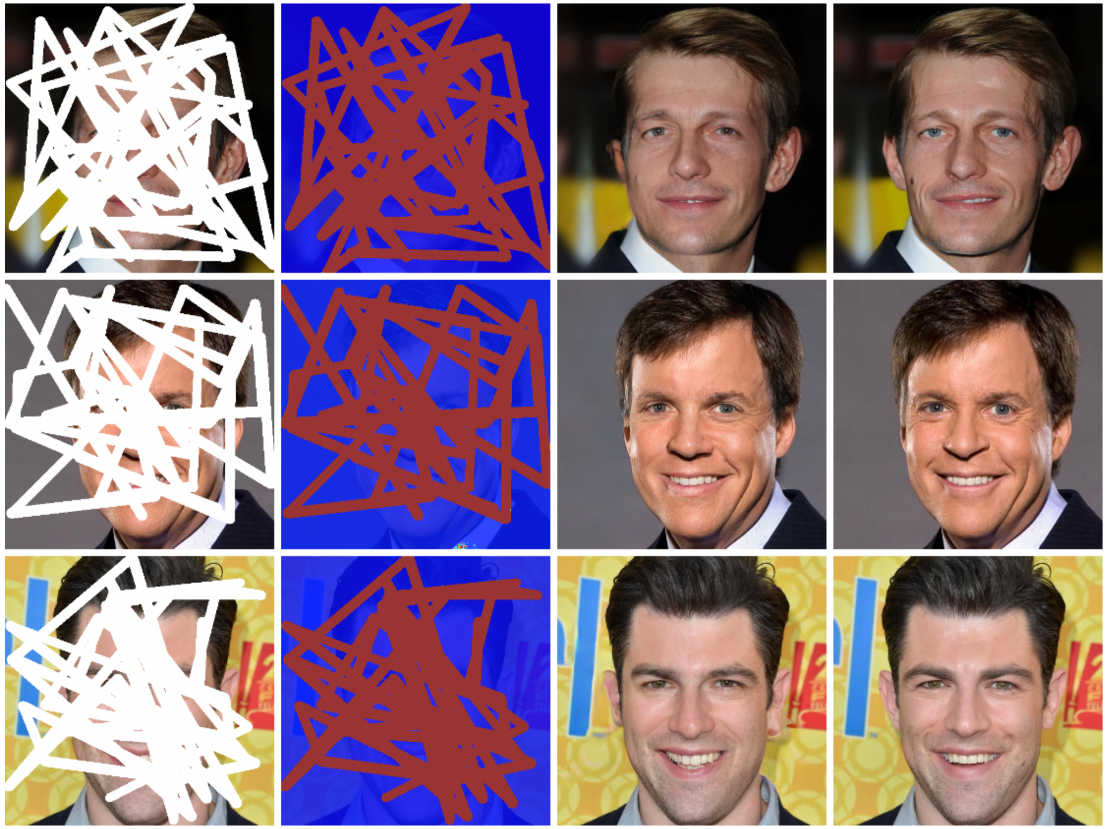
Generalization ability

- Additional synthesis results of BPFRe and the competing methods on in-the-wild face images

Input Pix2pixHD AutoRetouch GPEN MPRNet BPFRe

Applied to Image Inpainting

- The attention-guided two-stage architecture is applied to image inpainting



Input $\mathcal{M}^{E2D}$ $\hat{x}_{raw}$ **Ground truth**

Conclusion

- We propose an attention-guided progressive face retouching framework to remove blemishes naturally and synthesize high-fidelity content.
- We design a two-stage structure to exploit the merit of the U-Net architecture in restoring the image details and that of the GAN generator architecture in generating realistic images.
- The core idea is to explicitly suppress blemishes when transferring the intermediate features from the encoder to the decoder, and from the decoder to the generator
- We adopt a blemish-aware attention module to learn the weighting maps

Thank you