Supplementary File:
Escaping Data Scarcity for High-Resolution Heterogeneous Face Hallucination

Yiqun Mei*, Pengfei Guo*, Vishal M. Patel
Johns Hopkins University

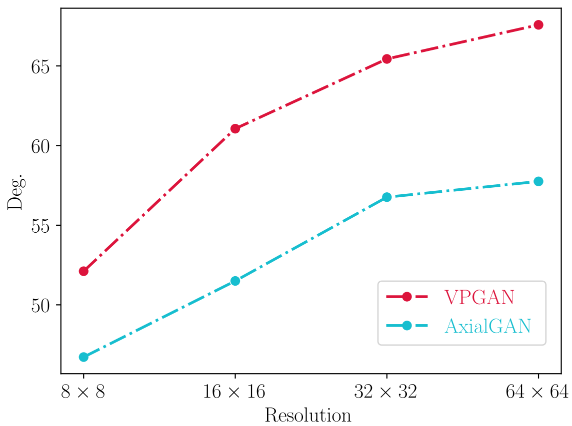


Figure 1. Effect of input resolution

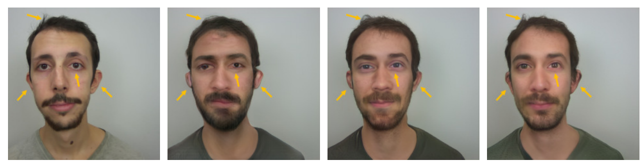
1. Study of Input Resolution

We study the effect of input resolution and compare our method with the previous best AxialGAN [2]. Input resolution is set to 8×8, 16×16, 32×32 and 64×64. For simplicity, we report Deg. which reflects identity preserving ability and is highly correlated to the verification accuracy. Results are shown in Figure 1. There is no surprise that the performance of both methods is improved as resolution increases. Our approach significantly outperforms AxialGAN at all resolutions, validating the benefits of the VPGAN.

2. Visual Comparison for Ablation Study

To better demonstrate the effectiveness of our designs, we conduct visual comparison for the ablation study. Results are shown in Figure 2. Columns 2 and 3 correspond to the variants without visual priors and MSCA, respectively. When comparing w/o VP and VPGAN, one can observe that leveraging visual priors can substantially improve hallucination, by synthesizing more accurate face outlines and detailed structures. While adopting the standard UNet [4] encoder (i.e. w/o MSCA) can still generate plausible images, face details are inconsistent with the ground-truth. This limits its verification accuracy. In contrast, adding MSCA (i.e. VPGAN) can bring apparent improvements on basic facial components and thus better preserve identity information. This proves that multi-scale information encoding is essential for more accurate hallucination.

*equal contribution



GT | w/o VP | w/o MSCA | **VPGAN**

Figure 2. Visual comparison for ablation study

GT | VPGAN | GT | VPGAN

Figure 3. Limitations of our method. First and second row show biased results towards skin and hair color, respectively

Table 1. The data statistics of clients.

Client	Train (images)	Dataset	Test (images)	Total
1	280	VIS-TH	210	1010
2	140			
3	220			
4	160			
5	1000	ARL-VTF	985	4185
6	860			
7	760			
8	580			
Total	4000		1195	5195

3. Limitations

One key design of VPFL is to utilize pre-trained GANs as a facial decoder. While leveraging visual priors greatly improves hallucination results, it limits the potential applications to images that can be produced by GANs. Therefore, it

GT/TH | pixel2pixel [3] | HiFaceGAN [5] | GANVFS [6] | SAGAN [1] | AxialGAN [2] | **VPGAN**

Figure 4. Visual comparison on the VIS-TH dataset.

GT/TH | pixel2pixel [3] | HiFaceGAN [5] | GANVFS [6] | SAGAN [1] | AxialGAN [2] | **VPGAN**

Figure 5. Visual comparison on the ARL-VTF dataset.

will be challenging to extend our method to general heterogeneous object recognition. In addition, although our approach can synthesize more clear and accurate images, the resulting faces may still have biased skin/hair color, an effect similar to existing approaches (Figure 3). This is because TH images cannot encode color information, and thus the produced faces tend to appear with an averaged color of the training dataset. We found using visual priors cannot alleviate this issue. In addition, VPFL relies on multi-institutional collaboration to scale up training, where each client is required to maintain stable connections. Thus, network failures might cause undesirable behaviors. In the future, we will extend this work by improving robustness towards potential network disconnections.

4. The Data Statistics of Clients in FL

Table 1 presents the detailed data statistics of 8 clients.

5. More Visual Results

We present more qualitative results in Figure 4 and Figure 5 to demonstrate the superiority of our approach.

References

[1] Xing Di, Benjamin S Riggan, Shuowen Hu, Nathaniel J Short, and Vishal M Patel. Polarimetric thermal to visible face veri-

fication via self-attention guided synthesis. In *IEEE International Conference on Biometrics*, pages 1–8, 2019. 2

[2] Rakhil Immidisetti, Shuowen Hu, and Vishal M. Patel. Simultaneous face hallucination and translation for thermal to visible face verification using axial-gan. In *IEEE International Joint Conference on Biometrics*, pages 1–8, 2021. 1, 2

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134, 2017. 2

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015. 1

[5] Lingbo Yang, Shanshe Wang, Siwei Ma, Wen Gao, Chang Liu, Pan Wang, and Peiran Ren. Hifacegan: Face renovation via collaborative suppression and replenishment. In *ACM International Conference on Multimedia*, pages 1551–1560, 2020. 2

[6] He Zhang, Benjamin S Riggan, Shuowen Hu, Nathaniel J Short, and Vishal M Patel. Synthesis of high-quality visible faces from polarimetric thermal faces using generative adversarial networks. *International Journal of Computer Vision*, 127(6):845–862, 2019. 2