Learning Flow-based Feature Warping for Face Frontalization with Illumination Inconsistent Supervision Supplementary Material

Yuxiang Wei[1], Ming Liu[1], Haolin Wang[1], Ruifeng Zhu[2,4], Guosheng Hu[3], and Wangmeng Zuo[1,5(✉)]

[1] School of Computer Science and Technology, Harbin Institute of Technology, China
[2] University of Burgundy Franche-Comté, France [3] Anyvision, UK
[4] University of the Basque Country, Spanish [5] Peng Cheng Lab, China
yuxiang.wei.cs@gmail.com, {csmliu, Why_cs}@outlook.com
{reefing.z, huguosheng100}@gmail.com, wmzuo@hit.edu.com

1 Training details for Flow Estimation Networks

It is difficult and expensive to manually obtain two ground-truth flow fields between the profile I and the ground-truth image I^{gt}. Instead, we introduce the landmark loss [4], sampling correctness loss [6] and the regularization term [6] to pretrain the bi-directional flow estimation networks (the forward flow estimation network $\mathcal{F}$ and the reverse flow estimation network $\mathcal{F}'$). For landmark loss, we use the dense landmark detection method[1] to detect 1000 facial landmarks for I and I^{gt}. We then move face contour landmarks in the vertical and horizontal directions and mark new landmarks and correspondences in the above areas between I and I^{gt}. In this way, we can deform the additional face areas (*e.g.*, hair, neck and ears). In our experiments, we pretrain the $\mathcal{F}$ and the $\mathcal{F}'$ for 4 epochs and then all networks are trained in an end-to-end manner.

2 Additional Qualitative Results

Fig. 1 shows the face synthesized results on Multi-PIE within $\pm 90°$ at 12 different poses (except $0°$). It is obvious that our model can synthesize photo-realistic images with delicate details across all pose variations.

More synthesized results on the LFW dataset under large poses are given in Fig. 2. It can be seen that our method exhibits satisfying generalization to in-the-wild face images, and the frontalization results are consistent with the profile face images.

To better understand the Warp Attention Module (WAM), we visualize the learned flow fields, warped images, and attention maps in the Fig. 4. For optical flow visualization, we use the color coding of Butler *et al.* [1]. The color coding

[1] https://www.faceplusplus.com/dense-facial-landmarks/

Figure 1: Synthesis results by our model under different poses of the Multi-PIE dataset. From top to down, the poses are $\pm 15°, \pm 30°, \pm 45°, \pm 60°, \pm 75°, \pm 90°$. The ground-truth frontal images are provided at the last row.

scheme is illustrated in Fig. 3. Hue represents the direction of the displacement vector, while the intensity of the color represents its magnitude. White color corresponds to no motion. As shown in Fig. 4, face frontalization can be viewed as the horizontal rotation of the face (the learned flow fields are mainly blue and red which represent horizontal rotation). Using learned forward flows field can warp the profile to the frontal view. And the learned attention maps can help focus on the critical parts of the warped features.

3 Additional Quantitative Results

Tab. 1 shows the Rank-1 recognition rates of different methods under Multi-PIE Setting 1. The recognition rates of all methods drop as pose degree increases. Missing more facial appearance information leads to the difficulty of synthesis task with pose rotation angle increases. As shown in Tab. 1, our model achieves the best performance across all poses, which demonstrates that our model can synthesize frontal images while preserving the identity information.

Figure 2: Additional synthesis results of the LFW dataset by our model. Each pair presents the profile (left) and the synthesized frontal face (right).

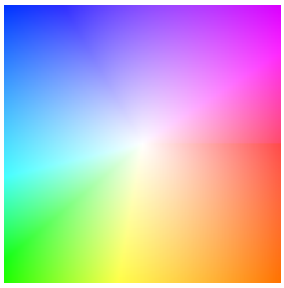

Figure 3: Flow field color coding used in this paper. The displacement of every pixel in this illustration is the vector from the center of the square to this pixel. The central pixel does not move.

Table 1: Rank-1 recognition rates (%) across poses under Setting 1 of the Multi-PIE. The best two results are highlighted by **bold** and underline respectively.

Method	±15°	±30°	±45°	±60°	±75°	±90°	Avg
Light CNN [7]	99.78	99.80	97.45	73.30	32.35	9.00	68.61
TP-GAN [3]	99.78	99.85	98.58	92.93	84.10	64.03	89.88
CAPG-GAN [2]	99.95	99.37	98.28	93.74	87.40	77.10	92.64
PIM [8]	99.80	99.40	98.30	97.70	91.20	75.00	93.57
3D-PIM [9]	99.83	99.47	99.34	98.84	94.34	76.12	94.66
FNM [5]	99.90	99.50	98.20	93.70	81.30	55.80	88.07
Ours	**100.00**	**100.00**	**100.00**	**98.86**	**96.54**	**88.55**	**97.33**

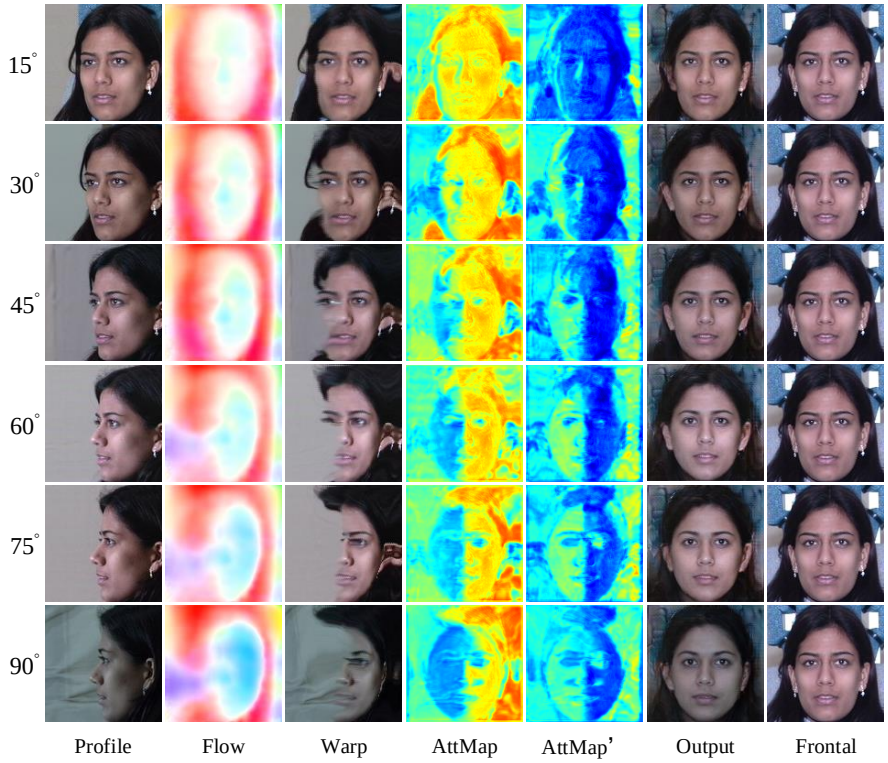

Figure 4: The visualization of learned flow fields, warped images, and attention maps. AttMap and AttMap' represent the attention map of warped feature and its flip, respectively

References

1. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: European conference on computer vision. pp. 611–625. Springer (2012)
2. Hu, Y., Wu, X., Yu, B., He, R., Sun, Z.: Pose-guided photorealistic face rotation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8398–8406 (2018)
3. Huang, R., Zhang, S., Li, T., He, R.: Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2439–2448 (2017)
4. Li, X., Liu, M., Ye, Y., Zuo, W., Lin, L., Yang, R.: Learning warped guidance for blind face restoration. In: Proceedings of the European Conference on Computer Vision. pp. 272–289 (2018)
5. Qian, Y., Deng, W., Hu, J.: Unsupervised face normalization with extreme pose and expression in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9851–9858 (2019)
6. Ren, Y., Yu, X., Chen, J., Li, T.H., Li, G.: Deep image spatial transformation for person image generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7690–7699 (2020)
7. Wu, X., He, R., Sun, Z., Tan, T.: A light cnn for deep face representation with noisy labels. IEEE Transactions on Information Forensics and Security **13**(11), 2884–2896 (2018)
8. Zhao, J., Cheng, Y., Xu, Y., Xiong, L., Li, J., Zhao, F., Jayashree, K., Pranata, S., Shen, S., Xing, J., et al.: Towards pose invariant face recognition in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2207–2216 (2018)
9. Zhao, J., Xiong, L., Cheng, Y., Cheng, Y., Li, J., Zhou, L., Xu, Y., Karlekar, J., Pranata, S., Shen, S., et al.: 3d-aided deep pose-invariant face recognition. In: IJCAI. vol. 2, p. 11 (2018)