Semi-Supervised Wide-Angle Portraits Correction by Multi-Scale Transformer

Fushun Zhu [1*] Shan Zhao [2*] Peng Wang [2*] Hao Wang [2] Hua Yan [1†] Shuaicheng Liu [3,2†]
[1] Sichuan University [2] Megvii Technology
[3] University of Electronic Science and Technology of China

1. Additional Implementation Details

1.1. Image Scaling

In our experiments, the original distortion images from smartphones are not straight-forward for training directly due to the large size. Therefore, we scale all the images to a uniform size with 384×512, and send the resized images to our MS-Unet for training. Consequently, the the output correction flow maps share the same size (i.e., 384×512) with the resized images. Afterwards, the correction flow maps are resized as the same size as the original images, and then they are utilized to correct the original distortion images into normal images. This scaling policy enables our MS-Unet to be executed under a lower complexity, which makes it feasible to apply into smartphones.

1.2. Correction Flow Map & Segmentation Mask

In Tan's method [3], two sub-networks were designed to implement the wide-angle portraits correction. In their method, the LineNet produces the perspective projection flow maps to project the distortion image as flattened, while the ShapeNet predicts the face correction flow maps to correct the flattened images into normal images. To reduce the structural redundancy, we design the one-stage network called MS-Unet to generate the correction flow maps, which can project the distortion image into normal image directly.

Fig. 1 shows some corrected images by our MS-Unet, as well as the corresponding correction flow maps (including horizontal and vertical correction flow maps). Usually the darker the region in the flow map, the stronger correction the region requires. Consequently, we can observe that the flow maps pay more attention to the distortion faces in addition to the corners of the distortion image. Meanwhile, Fig. 1 also illustrates the corresponding segmentation mask of each flow map, which is developed to assist our semi-supervised scheme.

*Equal contribution. †Corresponding authors.

2. Introduction about the Unlabeled Data

We construct the unlabeled dataset containing $5,000$ distortion images with various scenes and smartphones. This dataset make it possible to train the MS-Unet with our proposed semi-supervised scheme.

The samples in the unlabeled dataset are captured with 4 types of smartphones (including Samsung Note 10, Xiaomi 11, vivo X23 and vivo iQOO) with wide-angle lens of different distortion modules. Each smartphone contains various scenes with both horizontal and vertical orientation, and the number of people in each image is also range from 1 to 3.

Besides, our unlabeled dataset contains a variety of complex scenes, and some samples are show in Fig. 2. To be specific, these images are captured with different shooting ways. In particular, the shooting ways include different number of people, shooting angles and shooting distances, as shown in Fig. 2. Different shooting ways are combined with different orientation, covering various types of distortions.

3. Comparison Results

3.1. Compare with general semi-supervised methods.

To further evaluate the effectiveness of our proposed semi-supervised strategy. We implement another representative semi-supervised algorithm FixMatch [2] which is general in semi-supervised field, then the strategy is employed in our MS-Unet. As shown in Table 1, compared with the fully supervised MS-Unet, the FixMatch combined with MS-Unet can improve the performance, but our semi-supervised method can improve the performance more.

3.2. Comparison with Other Methods

We show more visual comparisons with perspective undistortion, Shih' results [1], and Tan's results [3]. We mark the obvious differences in the corrected images with red boxes. In general, our proposed method strikes a better balance between correcting the straight lines and distortion

Table 1. Performance comparison for Fully-Supervised MS-Unet, FixMatch and our method.

Method	LineAcc	ShapeAcc
Fully-Supervised MS-Unet	66.825	97.491
FixMatch [2] + MS-Unet	66.932	97.494
Ours	**67.209**	**97.500**

faces. Moreover, it keeps a more natural transition between the faces and the background.

3.3. Comparison with Other Phones

In addition, Fig. 4 shows more visual results compared with the corrected wide-angle portraits images by iPhone 12 and Xiaomi 11. We can clearly observe that our proposed method is superior to the two commercial solutions. Especially the region marked with red box, the corrected faces is more natural than others.

3.4. Our Future Research Direction

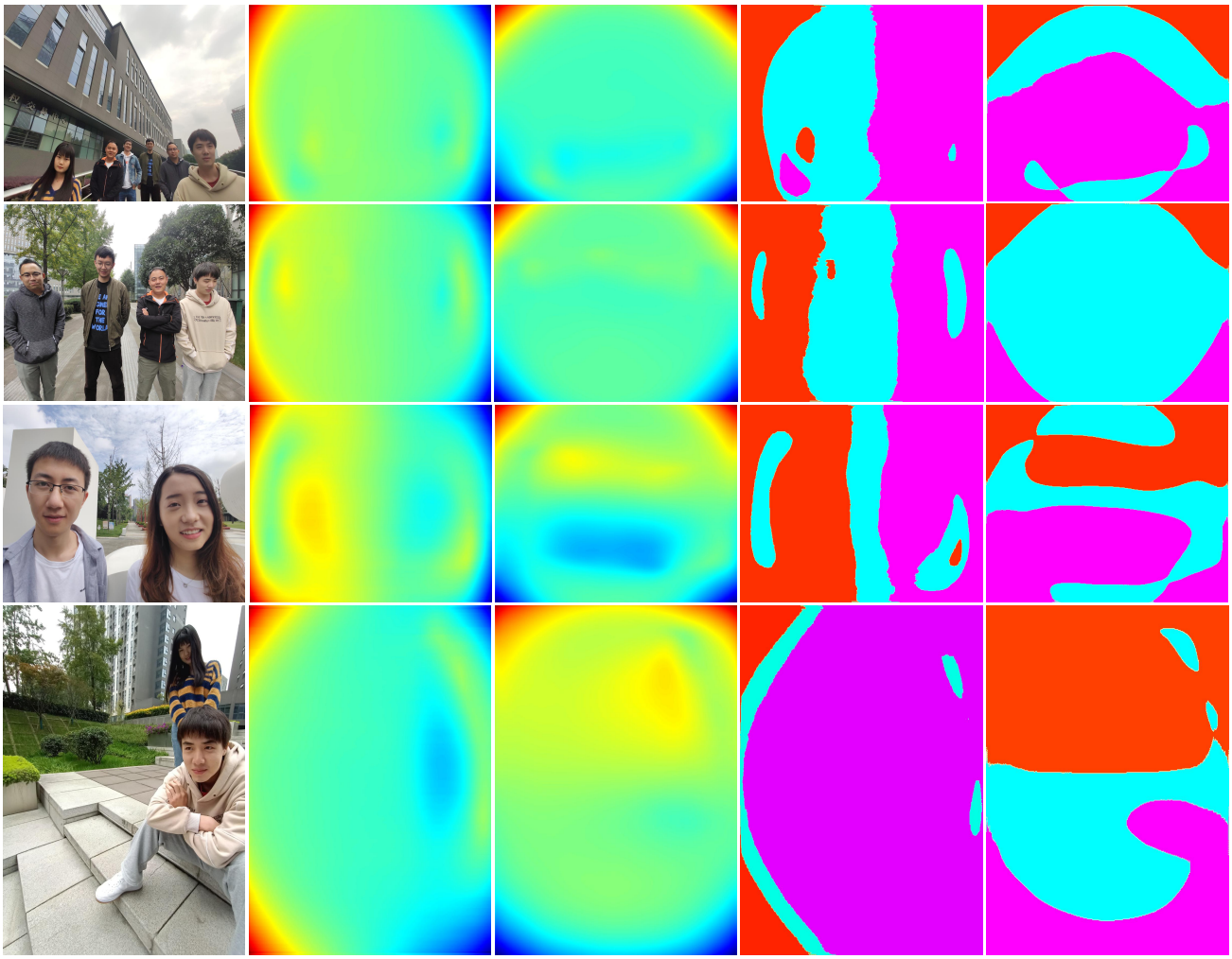
Although our proposed method can correct distortion images well in many scenes, the performance needs to be further improved in some scenes. As shown in Fig. 5, we further evaluate the effectiveness of our proposed method, and explore the space to further enhance. Fig. 5 (a) shows two unsatisfactory examples where the feet are close to the corners of the images. It is inevitable to force the feet to overstretch when correcting the corners. Similarly, the body in Fig. 5 (b) is also overstretch due to the strong correction of corners. The main reason is that the current ground truth flow maps only focus on the correction of distorted faces but ignore the other part of the body, such as feet and arms. In the future, we will continue to expand our work to correct the distortion of the whole body, which can solve the problem mentioned in Fig. 5. Also, it will make the corrected images more natural.

References

[1] YiChang Shih, Wei-Sheng Lai, Chia-Kai Liang, and Chia-Kai Liang. Distortion-free wide-angle portraits on camera phones. *ACM Trans. Graphics*, 38(4):1–12, 2019. 1

[2] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems*, 33:596–608, 2020. 1, 2

[3] Jing Tan, Shan Zhao, Pengfei Xiong, Jiangyu Liu, Haoqiang Fan, and Shuaicheng Liu. Practical wide-angle portraits correction with deep structured models. In *Proc. CVPR*, pages 3498–3506, 2021. 1



(a) Correction Image | (b) Horizontal Correction Flow Map | (c) Vertical Correction Flow Map | (d) Horizontal Segmentation Mask | (e) Vertical Segmentation Mask

Figure 1. Visualization results of our proposed methods. From left to right: (a) the correction images, (b) the horizontal correction flow map, (c) the vertical correction flow map, (d) the horizontal segmentation mask, (e) the vertical segmentation mask.

(a) Scene 1

(b) Scene 2

(c) Scene 3

(d) Different Distances

(d)Different Angles

Figure 2. Some samples of different scenes in our unlabeled dataset. 5 different shooting scenes with various people and backgrounds are shown in this figure.

(a) Projection Image (b) Shih's Result (c) Tan's Result (d) Our Result

Figure 3. Qualitative results of different wide-angle portraits correction methods. Note that the obvious differences in the corrected images are marked with red boxes.

(a) Distortion Image (b) Projection Image (c) iPhone12's Result (d) Xiaomi11 ultra's Result (e) Our Result

Figure 4. Visual comparison between our method and two other smartphones with wide-angle portraits correction methods. We mark the obvious differences in the correction images with red boxes.

(a) The Serious Distortion of Foot (b) The Serious Distortion of Body

Figure 5. Some corrected images that require to be further improved.