Practical Wide-Angle Portraits Correction with Deep Structured Models (Supplementary materials)

Jing Tan[1*] Shan Zhao[1*] Pengfei Xiong[2*] Jiangyu Liu[1] Haoqiang Fan[1] Shuaicheng Liu[3,1 †]

[1]Megvii Research [2]Tencent
[3]University of Electronic Science and Technology of China
{tanjing, zhaoshan, liujiangyu, fhq}@megvii.com
xiongpengfei2019@gmail.com, liushuaicheng@uestc.edu.cn
https://github.com/TanJing94/Deep_Portraits_Correction

1. A Detailed Introduction to Our Training Set

We construct a training dataset containing distorted and undistorted images with various subjects, scenes and camera modules, which provides a foundation for the wide-angle portrait correction.

Table. 1 shows the specific distribution of our training set on different camera modules and scenes divided by the number of people in each image. The training set covers 5 types of smartphones (Samsung A9S, Samsung note10, Vivo x23, Xiaomi 9 and Samsung A21) with wide-angle lens of different distortion modules, and each camera module contains scenes with different number of people from 1 to 6, corresponding to (Scene 1 to Scene 6 in Table 1). In the table, we show the number of images of each scene, as well as its percentage over all dataset. In addition, people are evenly distributed based on both modules and scenes.

Furthermore, the scenes of our training set are diverse, as shown in Fig. 1. Specifically, it includes a variety of common indoor and outdoor scenes. Meanwhile, different scenes include different shooting angles and shooting distances. In particular, the shooting angles include high-angle, flat-angle, and low-angle while distances include long-range and short-range as shown in Fig. 1. Different angles can combine with different ranges, covering various types of distortions.

2. LineNet Detailed Topology

Table. 2 shows the detailed topology of the proposed LineNet without the Line Attention Module(LAM), which is similar to the ShapeNet. In the table, Pooling-F-S denotes the filter size and stride of each pooling operator. Conv-C-F-S-P means the output channel, filter size, stride and padding of each convolution operator, and Deconv-C-F-S-P has the same meaning in terms of C-F-S-P. BN, ReLU, Res are abbreviation of Batch Normalization, the activation operation and the operation of residual adding, respectively.

3. More Comparison Results

3.1. Comparison with Other Methods

We show more qualitative comparisons with perspective undistortion (Fig. 2(a)), stereographic projection (Fig. 2(b)) and Shih' results [1] (Fig. 2(c)) on our testset and Shih's dataset [1]. Generally, the proposed method not only corrects the straight lines, but also corrects the face deformation as well, while keeping the natural transition between the background and the face regions.

3.2. Comparison with Other Phones

In addition, we show more visual results compared with iPhone 12 and Xiaomi 10 which contain the function of wide-angle portrait correction. As can be seen from Fig. 3, the proposed method is clearly superior to the two commercial solutions.

3.3. Visualization of corner cases

As shown in Fig. 4, we further verified the effectiveness of our algorithm on more corner cases, including some failure cases, internet faces and landscape photos. Fig 4 (a) shows a failure example where a straight line is in front of the face. It is inevitable to bend the line when correcting the face. Fig 4 (b) shows an example of Internet image with non-Asian faces, and Fig 4 (c) shows an example of landscape photo that contains no man-made structures, our method can handle both well.

*Equal contribution
†Corresponding author

Table 1. Training set distribution on different camera modules and scenarios with different number of people.

	#Scene 1	#Scene 2	#Scene 3	#Scene 4	#Scene 5	#Scene 6	Total
Samsung A9S	143	162	226	98	254	159	1042(17.8%)
Samsung note10	348	278	170	88	52	190	1126(19.2%)
Vivo x23	282	100	110	207	210	129	1038(17.8%)
Xiaomi 9	265	169	181	391	320	155	1481(25.3%)
Samsung A21	179	154	81	296	302	155	1167(19.9%)
Total	1217(20.8%)	863(14.7%)	768(13.1%)	1080(18.5%)	1138(19.4%)	788(13.5%)	5854

(a) High-angle Shot

(b) Flat-angle Shot

(c) Low-angle Shot

(d) Long-range Shot

(e) Close-range Shot

Figure 1. Examples of different scenes in the training set. The figure shows 5 different shooting scenes, with various people and backgrounds in the scenes. Each column from left to right represents the original distortion image, the projection image and the final correction image.

References

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

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

Figure 2. Qualitative evaluation of undistortion methods. Notice the coordination of face area and line curvatures in the transition area.

Table 2. Network topology of the proposed LineNet without the Line Attention Module(LAM).

Inputs	LineNet	Outputs
Img(3, 256, 384)	$Conv-16-3-2-0, BN$ $Pooling-2-2$	Res-1(16, 128, 192)
Res-1(16, 128, 192)	$Concat, Pooling-2-2, Conv-32-1-1-0$ $\left[\begin{array}{c} Conv-8-1-1-0, BN, ReLU \\ Conv-8-3-1-1, BN, ReLU \\ Conv-32-1-1-0, BN, Res, ReLU \end{array}\right] \times 3$	Res-2(32, 64, 96)
Res-2(32, 64, 96)	$Concat, Pooling-2-2, Conv-32-1-1-0$ $\left[\begin{array}{c} Conv-8-1-1-0, BN, ReLU \\ Conv-8-3-1-1, BN, ReLU \\ Conv-32-1-1-0, BN, Res, ReLU \end{array}\right] \times 4$	Res-3(32, 32, 48)
Res-3(32, 32, 48)	$Concat, Pooling-2-2, Conv-64-1-1-0$ $\left[\begin{array}{c} Conv-16-1-1-0, BN, ReLU \\ Conv-16-3-1-1, BN, ReLU \\ Conv-64-1-1-0, BN, Res, ReLU \end{array}\right] \times 5$	Res-4(64, 16, 24)
Res-4(64, 16, 24)	$Concat, Pooling-2-2, Conv-128-1-1-0$ $\left[\begin{array}{c} Conv-32-1-1-0, BN, ReLU \\ Conv-32-3-1-1, BN, ReLU \\ Conv-128-1-1-0, BN, Res, ReLU \end{array}\right] \times 4$	Dec-4(128, 8, 12)
Dec-4(128, 8, 12)	$Conv-32-3-1-1, BN, ReLU$ $\left[\begin{array}{c} Conv-32-3-1-1, BN, ReLU \\ Conv-32-3-1-1, BN \\ Res \end{array}\right]$ $Deconv-32-2-2-0, BN, ReLU$	Dec-3(32, 16, 24)
Dec-3(32, 16, 24)	$\{Conv-24-3-1-1, BN, ReLU\}(Res-3), Add$ $\left[\begin{array}{c} Conv-24-3-1-1, BN, ReLU \\ Conv-24-3-1-1, BN \\ Res \end{array}\right]$ $Deconv-24-2-2-0, BN, ReLU$	Dec-2(24, 32, 48)
Dec-2(24, 32, 48)	$\{Conv-16-3-1-1, BN, ReLU\}(Res-2), Add$ $\left[\begin{array}{c} Conv-16-3-1-1, BN, ReLU \\ Conv-16-3-1-1, BN \\ Res \end{array}\right]$ $Deconv-16-2-2-0, BN, ReLU$	Dec-1(16, 64, 96)
Dec-1(16, 64, 96)	$\{Conv-8-3-1-1, BN, ReLU\}(Res-1), Add$ $\left[\begin{array}{c} Conv-8-3-1-1, BN, ReLU \\ Conv-8-3-1-1, BN \\ Res \end{array}\right]$ $Deconv-8-2-2-0, BN, ReLU$	Dec-0(8, 128, 192)
Dec-0(8, 128, 192)	$\{Conv-4-3-1-1, BN, ReLU\}(Res-0), Add$ $\left[\begin{array}{c} Conv-4-3-1-1, BN, ReLU \\ Conv-4-3-1-1, BN \\ Res \end{array}\right]$ $Deconv-4-2-2-0, BN, ReLU$ $Conv-2-3-1-1$	Flow(2, 256, 384))

(a) Input Image (b) Projection Image (c) iPhone12's Result (d) Xiaomi10's Result (e) Our Result

Figure 3. Qualitative comparison between our method and some phones with wide-ange portrait correction.

(a)

(b)

(c)

Input Our Result

Figure 4. (a) an failure example. (b) an internet image with non-Asain faces. (c) a landscape photo without man-made structures.