

Combining Generative and Geometry Priors for Wide-Angle Portrait Correction

Lan Yao¹, Chaofeng Chen², Xiaoming Li¹✉, Zifei Yan¹, and Wangmeng Zuo^{1,3}

¹ Harbin Institute of Technology

² Nanyang Technological University

³ Pazhou Lab, Huangpu

{mrha011010, chaofenghust, csxmli}@gmail.com {yanzifei, wmzuo}@hit.edu.cn

Abstract. Wide-angle lens distortion in portrait photography presents a significant challenge for capturing photo-realistic and aesthetically pleasing images. Such distortions are especially noticeable in facial regions. In this work, we propose encapsulating the *generative face prior* as a guided natural manifold to facilitate the correction of facial regions. Moreover, a notable central symmetry relationship exists in the non-face background, yet it has not been explored in the correction process. This *geometry prior* motivates us to introduce a novel constraint to explicitly enforce symmetry throughout the correction process, thereby contributing to a more visually appealing and natural correction in the non-face region. Experiments demonstrate that our approach outperforms previous methods by a large margin, excelling not only in quantitative measures such as line straightness and shape consistency metrics but also in terms of perceptual visual quality. All the code and models are available at <https://github.com/Dev-Mrha/DualPriorsCorrection>.

Keywords: Portrait correction · Generative prior · Geometry prior

1 Introduction

Due to the convenience of taking pictures with mobile phones and the advantages of wider viewing angles, people frequently choose the wide-angle mode for photographing individuals. However, the distortion caused by wide-angle lenses often results in images with distorted lines and stretched faces, greatly affecting the aesthetic quality of portrait photography.

Traditional methods focus on optimizing a warping map to correct distortions brought by different projections, such as perspective, Mercator and stereographic projections [5, 6, 16]. Since the human faces are usually quite close to the camera and severely affected by these distortions, it cannot be simply performed simultaneously with background correction. Therefore, previous works usually handle face and background regions with different methods. Shih *et al.* [38] proposed a solution by combining perspective projection for the background and stereographic projection for the face region, using a warp mesh. They approached it as an optimization problem, seeking an optimal balance between redressing

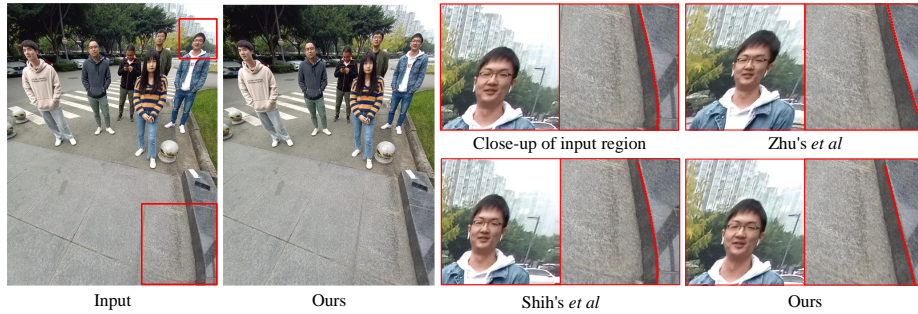


Fig. 1: Example of distorted and corrected images compared with other methods. Thanks to the proposed generative and geometry priors, our results have straighter lines in the background and more natural-looking faces compared to other methods.

the background and faces. Recent approaches [40, 48] introduced deep learning methods to tackle this issue, eliminating the need for additional parameters. Although great progress has been achieved, the real-world faces may suffer from various types of distortions that are challenging to correct without additional information, especially when training data is limited (see Fig. 1).

In this paper, we introduce a framework that leverages generative and geometry priors to rectify wide-angle distortions in faces and backgrounds, respectively. We follow previous practice [40] and separate our framework into two modules: FaceCNet, which targets facial distortion, and LineCNet, which addresses the straightness of buildings and lines in the background. The FaceCNet employs the generative face structure prior from pre-trained StyleGAN to enhance facial correction. Our key motivation is that the pre-trained StyleGAN is able to map the latent vectors \mathcal{W} to a normally shaped face through GAN inversion [1, 41]. Given that \mathcal{W} is a compact latent space, it facilitates the mapping of distorted faces into \mathcal{W} thereby restoring facial structures. This structural guidance enables FaceCNet to effectively correct wide-angle facial distortions with complex projection. For the background regions, we empirically observe that they mainly suffer from barrel distortion and exhibit central symmetry. We therefore introduce symmetric regularization in the training of LineCNet. This simple geometric prior proves effective in straightening lines in the background. Finally, we design a post-processing process to harmoniously fuse the corrected face region into the background. The experimental results show our method produces undistorted images with well-proportioned faces, surpassing existing techniques in both visual quality and quantitative evaluation metrics. The main contributions are:

- We integrated the generative face prior from StyleGAN to aid in correcting wide-angle distortions of facial regions. This face structure prior enables our method to handle faces distorted by complex projections.
- To the best of our knowledge, we are the first to employ a symmetry prior in the background of wide-angle photos, which is highly effective in correcting curved lines.

2 Related Works

2.1 Wide-angle Portrait Correction

Traditional approaches to correcting wide-angle lens distortion often rely on mathematical formulations and optimization problems, identifying specific formulas to mitigate the distortion effect [7, 42, 46, 49]. However, due to the complexity of real-world problems that often resist ideal mathematical modeling, recent methods have increasingly turned to deep learning-based approaches. Zhao *et al.* [47] utilized camera calibration to obtain parameters for predicting flow with U-Net, facilitating eye-fish undistortion. Shih *et al.* [38] employed the field of view (FOV) to compute a mesh, integrating various distortions in the face region and background. Cao *et al.* [4] introduced an adaptive triangle mesh for wide-angle portrait correction. However, these methods require extra camera parameters which are often not available. Therefore, recent methods tried to directly learn to correct the distortion. Tan *et al.* [40] pioneered the use of deep learning methods to remove distortions without relying on camera parameters, employing a two-stage network. Zhang *et al.* [45] They proposed a method for wide-angle correction through content-aware conformal mapping, applying different distortion recovery models to different parts. Zhu *et al.* [48] presented a semi-supervised transform-based network to exploit semantic information for portrait correction. Unlike these approaches, our method employs geometry and generative priors for background and facial region respectively, achieving better results for real-world portraits.

2.2 Face Prior as Reference

Structure prior has been proven effective in many low-level vision tasks, *e.g.*, depth image enhancement [15, 28], image inpainting [12, 23, 32, 34], and image restoration [8, 9, 11, 17, 22, 24–26, 43, 44]. With the rapid advancement of generative adversarial networks (GANs), particularly exemplified by StyleGAN [20], researchers have explored the potential of pre-trained GAN models to provide comprehensive priors, including both structure and textural attributes for many low-level tasks [8, 27, 43, 44]. Numerous GAN inversion methods [2, 14, 35, 41] have emerged, enabling the manipulation of real images by optimizing their latent codes within the GAN space [29]. This capability has also led to the adoption of StyleGAN in various face restoration tasks [8, 31, 33, 44]. For example, GFP-GAN [43] integrates a degradation module with a pre-trained StyleGAN on facial data, achieving remarkable performance in various face restoration tasks, such as super-resolution, inpainting, colorization *etc.* The primary advantage of employing generative priors lies in their capability to effectively generate missing regions within an image. This generative power holds particular promise for correcting wide-angle distortions in facial images, as the missing regions pose a challenge to straightforward correction via direct warping techniques.

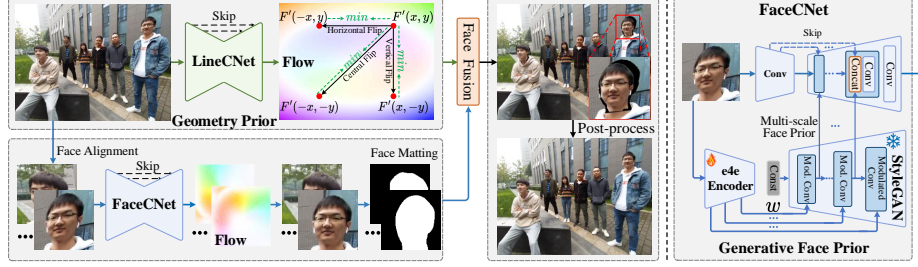


Fig. 2: Overview of our framework. It contains three parts: LineCNet for the geometric distortions on the background, FaceCNet for correcting the face region using generative face prior, and post-process for fusing the corrected face region into the background.

3 Method

3.1 Architecture Overview

Our approach is capable of correcting the wide-angle portrait that preserves the natural appearance of both the face and background regions. As illustrated in Fig. 2, the whole framework is divided into three parts: the LineCNet, FaceCNet, and face fusion block. Given a wide-angle image, the LineCNet is applied to address geometric distortions that affect straight lines in the background while the FaceCNet is introduced to correct the face regions with the guidance of generative face prior from StyleGAN. A face fusion block finally merges the undistorted faces generated by FaceCNet with the background corrected by LineCNet.

3.2 Generative Prior for Face Correction

In comparison with other objects, face region owns specific structures that are easily perceptible when exhibiting unnatural distortions. Therefore, in the correction process, we should concentrate more on the facial region. However, existing methods mainly consider facial landmarks as a constraint, overlooking overall facial consistency and naturalness. Notably, the generative models (*e.g.*, StyleGAN) have encapsulated the ability to generate photo-realistic face images, enabling us to facilitate face correction towards a more natural performance. In the following, we first review StyleGAN inversion for real-world images and then introduce the method of encapsulating the generative prior for face correction.

The original StyleGANs generate face images from random Gaussian noises. To adopt StyleGANs on real-world images, previous works map the face image into the \mathcal{W} space. This inversion process is to predict the latent representation of an image generated by a Generative Adversarial Network (GAN), aiming to find the latent code that, when input into the GAN generator, reproduces the target image. In our approach, we follow these methods and employ the encoder for editing (e4e) framework [41] for StyleGAN inversion. The e4e method [41] revolves around training an encoder network to map real images to their

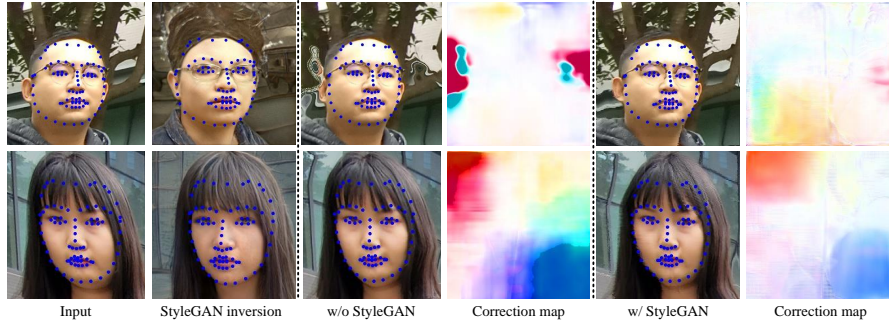


Fig. 3: Examples of our face correction w/ and w/o using generative prior. Through the results of correcting flow, it can be observed that the correction is concentrated at the position where the StyleGAN face structure deviates from the input.

corresponding latent codes within the $\mathcal{W}+$ latent space of StyleGAN2, which is a high-dimensional approximation of StyleGAN’s \mathcal{W} space. Once trained, this encoder network can deduce the latent code of an input image by locating the closest code in the learned latent space that would produce a similar image when processed by the StyleGAN2 generator. The inferred latent code enables various manipulations such as facial expression alterations, age progression/regression, and other image modifications. The e4e encoder provides a precise means for image editing by directly manipulating the latent space of StyleGAN2.

During the inversion process, it is inevitable to lose facial structures, resulting in lower fidelity. This implies that directly applying StyleGAN to correct the face region fails to preserve identity textures (refer to the StyleGAN inversion results in Fig. 3). Since the intermediate features in StyleGAN can generate photo-realistic structures aligning well with the natural manifold, we propose using these priors to guide the correction process. Specifically, for face images, we first obtain their \mathcal{W} vector using the pre-trained e4e encoder. We then utilize the intermediate features of StyleGAN to aid the prediction of deformation flow, preserving higher fidelity. Additionally, the e4e encoder is pre-trained on high-quality face images. To generalize to unnatural faces, we fine-tune the encoder with our synthetic distorted faces exhibiting characteristics similar to wide-angle distortion stretching. Here, we adopt a synthetic optical flow similar to the distortion produced by a wide-angle lens to generate distorted faces for training our FaceCNet. Leveraging the fine-tuned e4e model [41], distorted faces are inverted into the StyleGAN domain, facilitating the correction of distorted face images.

Fig. 2 (right) demonstrates the details of our FaceCNet. It predicts the face correction flow with the input of faces and their corresponding StyleGAN features. To achieve this, we employ the U-Net architecture [36] as a baseline, chosen for its ability to calculate cost volumes for flow prediction. The design of U-Net enables the fusion of features from various receptive fields, facilitating comprehensive information processing.

Moreover, we introduce multi-scale feature fusion into the U-Net framework to integrate StyleGAN features in different feature spaces. We analyze that the incorporation of multi-scale feature fusion serves a dual purpose. Firstly, it enables the seamless integration of StyleGAN features, leveraging the rich priors associated with facial structures. This integration is crucial for an effective correction process, as it allows FaceCNet to better understand and preserve intricate details inherent to facial images. Secondly, the multi-scale feature fusion contributes to the overall robustness of the network by accommodating information from different scales, ensuring that FaceCNet can capture both fine-grained and global features in the correction task. By incorporating StyleGAN features, FaceCNet gains access to additional information that goes beyond the standard input faces. This additional information proves instrumental in preserving natural facial details and enhancing the overall quality of corrected images. Consequently, the described architecture and feature integration approach contribute to FaceCNet’s ability to produce natural-looking results in the facial regions.

3.3 Symmetry Prior for Background Correction

Wide-angle lenses typically produce images with barrel distortion, which is a specific type of distortion related to the lens itself. Incorporating the radial distortion figure highlights the lens-induced nature of radial distortion, further supporting the argument for symmetry within barrel distortion. Barrel distortion arises as a geometric distortion due to the optical design of the lens. This occurs because the outer regions of the lens magnify the image more than the central portion, leading to a noticeable bulging effect towards the edges of the frame. Barrel distortion follows a specific format, and it is related to camera parameters [13, 46], as described by Eq. (1):

$$r_u = r_d \times (1 - k_1 \times r_d^2 - k_2 \times r_d^4), \quad (1)$$

where r_d represents the distance from the center of distortion in the distorted image, r_u represents the undistorted distance, and k_1 and k_2 are coefficients related to the lens’s distortion characteristics. These coefficients are determined during camera calibration and are used to correct barrel distortion in image processing workflows. From this formula and comparison chart, it can be found that barrel distortion is only related to the camera lens, and increases as the distance from the center grows.

Indeed, perspective distortion is another aspect of visual distortion commonly encountered in wide-angle photography, but it differs from barrel distortion in its origin and characteristics [3, 37]. Unlike barrel distortion, which is a geometric distortion caused by the optical properties of the lens, perspective distortion is an undesired effect in real images that is not caused by lens flaws. Perspective distortion is related to the relative sizes and distances of objects within the scene, as well as the positioning and angle of view of the camera. Objects close to the lens appear abnormally large relative to more distant objects, and distant objects appear abnormally small and hence farther away. When perspective distortion

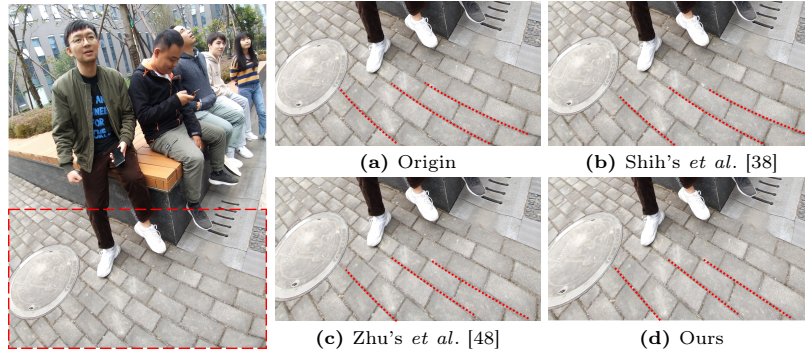


Fig. 4: Comparison results of background correction with others. We can observe that the proposed LineCNet can better correct the distorted lines.

interacts with stereo objects, it can disrupt the accurate alignment of features between images, complicating depth perception and 3D reconstruction in stereo imaging applications. That’s why we cannot use common undistortion methods to correct stereo objects like human faces.

Therefore, without considering spherical objects such as human faces, the distortion introduced by wide-angle lenses can be considered symmetrical. From this, we introduce symmetric loss to help correct the optical flow training of the background part. According to the optical flow to recover distortion should have horizontal, vertical, and central symmetry, a symmetric regularization loss is designed and applied in our LineCNet to effectively correct the distortion and reduce the complexity of the network without predicting the camera parameters.

In particular, our LineCNet, similar to FaceCNet, utilizes the U-Net network as its baseline architecture. As mentioned above, the distortion caused by the wide-angle lens to the image is consistent with symmetry, so no other modules except symmetry loss are added to the LineCNet. This is driven by the aim of maintaining simplicity and efficiency in addressing background distortions. By leveraging the U-Net and symmetry constraint, LineCNet can effectively correct the background distortions in wide-angle images as shown in Fig. 4.

3.4 Face Fusion and Postprocess

In contrast to previous approaches, we adopt a novel strategy of isolating the face from the background and subsequently reassembling the faces during the merging process. This strategy ensures that the straight lines surrounding the face in the background remain unaffected. Therefore, our FaceCNet can correct the face region towards a more photo-realistic performance while struggling to have pleasing results on the other region (see background in Fig. 3). To address this issue, we employ face parsing on cropped facial regions, allowing only the precise facial components to be inserted. ParseNet [9] is utilized for this purpose which separated the face and background in the crop face image and obtained the

mask of the face part. When reattaching the face to the image, we calculate the translation position of the face before and after LineCNet correction to ensure a more accurate fit between the face and the corrected background. In this process, face correction and background correction will cause a certain gap in the region, especially the region near the border. Image in-painting algorithm Lama [39] or other diffusion-based methods [30] is finally used to fill this missing region.

This comprehensive approach ensures that both facial and background components are corrected separately and then integrated seamlessly, leading to high-quality, and distortion-free images.

3.5 Learning Objectives

Since our FaceCNet and LineCNet are designed for different regions, one for the face and the other for the background, they can be trained in parallel.

As for training FaceCNet, it is natural to require the predicted flow vector Φ close to the ground-truth Φ^{gt} , and we thus define the face deformation loss as:

$$\mathcal{L}_{face}^f = \|\Phi_{face} - \Phi_{face}^{gt}\|_2^2. \quad (2)$$

Meanwhile, we also need to ensure that the warped face image I_{face}^w using Φ approaches to the ground-truth image I_{face}^{gt} in the pixel domain:

$$\mathcal{L}_{face}^p = \|I_{face}^w - I_{face}^{gt}\|_2^2, \quad (3)$$

where I_{face}^w is obtained using the flow interpolation:

$$I_{face}^w(i, j) = \sum_{(h, w) \in \mathcal{N}} I_{face}(h, w) \max(0, 1 - |\Phi_{i,j}^y - h|) \max(0, 1 - |\Phi_{i,j}^x - w|). \quad (4)$$

\mathcal{N} denotes the 4-pixel neighbors and I_{face} is the face region cropped from the wide-angle portraits.

Finally, we add total variation (TV) loss as smoothness regularization to avoid too large changes of adjacent areas:

$$\mathcal{L}_{face}^{tv} = \sum_{i,j} \sqrt{(\Phi_{i,j-1} - \Phi_{i,j})^2 + (\Phi_{i+1,j} - \Phi_{i,j})^2}. \quad (5)$$

To sum up, the final learning objective for FaceCNet is defined as:

$$\mathcal{L}_{FaceCNet} = \mathcal{L}_{face}^f + \lambda_1 \mathcal{L}_{face}^p + \lambda_2 \mathcal{L}_{face}^{tv}, \quad (6)$$

where λ_1 and λ_2 are set to 2 and 0.5, respectively.

In terms of optimizing LineCNet, the loss function is defined as:

$$\mathcal{L}_{LineCNet} = \|\Phi_{bg} - \Phi_{bg}^{gt}\|_2^2 + \lambda_3 \|I_{bg}^w - I_{bg}^{gt}\|_2^2 + \lambda_4 \mathcal{L}_{Sym}, \quad (7)$$



Fig. 5: Example of our synthetic face pairs for training FaceCNet.

where Φ_{bg} and I_{bg}^w are the predicted background flow field and warped background image. Φ_{bg}^{gt} and I_{bg}^{gt} denote their corresponding ground-truth. The symmetry loss \mathcal{L}_{Sym} is defined as:

$$\mathcal{L}_{Sym} = \|\Phi_{bg}^v - \Phi_{bg}\|_2^2 + \|\Phi_{bg}^h - \Phi_{bg}\|_2^2 + \|\Phi_{bg}^c - \Phi_{bg}\|_2^2, \quad (8)$$

where Φ_{bg}^v , Φ_{bg}^h , and Φ_{bg}^c present the vertical, horizontal, and central flip of predicted flow field Φ_{bg} , respectively (see flow map illustration in Fig. 2). This loss function reinforces symmetrical alignment, enhancing the accuracy of line correction. In our experiment, λ_3 and λ_4 are set to 1 and 2, respectively.

4 Experiments

4.1 Data Preparation

Face dataset The GAN inversion network is trained and evaluated with the public dataset, *i.e.*, FFHQ [19] and CelebA-HQ [18]. To generate image pairs for training, we designed a flow that stretches the four corners outward to simulate the distortion caused by a wide-angle lens (see the examples in Fig. 5).

Wide angle portrait dataset We use the public wide-angle portrait dataset from Tan et al. [40]. This dataset consists of over 5000 images for training and 129 images for testing. Each item in the dataset contains the ground truth image, the line-corrected image, and the correction flow maps for training. It also provides landmark points of line correction and face correction for quantitative evaluation.

4.2 Implementation Details

As shown in Fig. 2, each network of our framework is trained separately and the results are combined with a fusion postprocess block. The FaceCNet is trained with the synthetic faces first and then finetuned with real images cropped from [40]. The face images are resized to 256×256 as inputs to FaceCNet. LineCNet is trained with 512×384 inputs which is consistent with previous works [40, 48]. Both of the networks are trained with Adam optimizer [21] and the learning rate is set to $1e - 4$. The batch size is 4 and it takes 5 days to train FaceCNet and 3 days for LineCNet on a 3060 NVIDIA GPU.

4.3 Evaluation Metrics

We follow previous works [40] and employ several evaluation metrics including LineAcc and ShapeAcc to assess the performance of our method.

LineAcc is designed to evaluate the straightness of lines within the image by comparing the slope variation between the output and ground truth images:

$$LS = 1 - \frac{1}{n} \sum_{i=0}^{n-1} \left(\frac{y_{d_i} - y_{d_{i-1}}}{x_{d_i} - x_{d_{i-1}}} - \frac{y_{g_o} - y_{g_n}}{x_{g_o} - x_{g_n}} \right) \quad (9)$$

where LS is the similarity between the slopes of lines in the output and ground truth images. n is the number of points in each line. (x_{g_i}, y_{g_i}) and (x_{d_i}, y_{d_i}) are the corresponding positions in the reference and distortion images, respectively.

ShapeAcc focuses on evaluating the similarity of facial landmarks between the output and reference images. This metric calculates the cosine similarity between corresponding landmark vectors, which is defined as follows:

$$FC = \frac{1}{n} \sum_{i=0}^{n-1} \cos(L_{g_i}, L_{d_i}) \quad (10)$$

where L_{g_i} and L_{d_i} denote the corresponding face landmark points in the reference and distortion images, respectively.

Landmark Distance To provide a more comprehensive measurement of facial correction, we introduce the Landmark Distance metric, which calculates the similarity of faces based on the Euclidean distance between corresponding landmark points as below:

$$LD = \frac{1}{n} \sum_{i=0}^{n-1} \|L_{g_i} - L_{d_i}\|_2^2 \quad (11)$$

For a fair comparison, we align the position of the nose position in reference and distorted images before calculating the distance.

4.4 Ablation Study

To evaluate the performance of our proposed framework, we first conduct ablation experiments to analyze the effectiveness of symmetry prior and face prior. Then, we conduct experiments to show the importance of a multi-scale feature fusion strategy for face prior.

Table 1: Ablations on the proposed generative and geometry priors.

Method	Symmetry Prior	Generative Prior	LineAcc	ShapeAcc	Landmark Distance
Baseline			66.192	97.027	5.991
Baseline + Sym.	✓		67.304	97.266	5.5464
Baseline + Gen.		✓	66.192	99.012	5.3397
Ours (<i>Full</i>)	✓	✓	67.304	99.012	5.013

**Fig. 6:** Visual comparison of our method with and without geometric symmetry prior.

Effectiveness of Priors We explore how the priors affect the network performance. We utilize the network without any priors as the baseline, and evaluate the performance of four networks with different prior configurations as follows:

- 1) Baseline: LineCNet and FaceCNet with simple U-Net structure and no priors;
- 2) Baseline + Sym. (Symmetry Prior): add symmetry loss for LineCNet;
- 3) Baseline + Gen. (Generative Prior): add generative face prior features from StyleGAN into FaceCNet;
- 4) Ours (*Full*): add both symmetric prior and generative prior.

The results of the quantitative comparison are provided in Tab. 1. We can observe that the symmetric prior improves the LineAcc while the generative prior benefits the ShapeAcc and Landmark Distance. The effectiveness of them is more obvious from the visual examples in Fig. 6 and Fig. 7.

In Fig. 6, the comparison shows that the predicted flow with L_{Sym} in Fig. 6 (b) is smoother and more symmetric than that without L_{Sym} in Fig. 6 (c). From the left bottom corner, we can see that with a more symmetric flow, the contours of bricks in Fig. 6 (c) are much straighter compared to those in Fig. 6 (b). The results clearly demonstrate the superiority of using symmetric prior.

In Fig. 7, the benefit of employing the StyleGAN prior is more evident. Without the StyleGAN prior, the network predicts an unusual flow near the cheek in Fig. 7 (b), leading to an odd face shape due to excessive shrinking. Meanwhile, the result in Fig. 7 (c) appears more natural. This demonstrates that



Fig. 7: Visual comparison of our method with and without generative face prior.

Table 2: Ablations on the StyleGAN feature concatenation

Method	PSNR	SSIM	Landmark Distance
1) Baseline	25.586	0.848	3.091
2) Baseline + Single Scale	24.335	0.797	2.908
3) Baseline + Multi-Scale	26.732	0.871	2.357

with the guidance of the StyleGAN prior, the network is less likely to produce odd results and is more effective at restoring natural-looking facial structures.

Multiscale facial prior features To evaluate the impact of incorporating multiscale StyleGAN features on face correction, we use a U-Net architecture as our baseline and explore three variants: 1) Baseline U-Net without StyleGAN; 2) Baseline + single scale StyleGAN feature of the smallest size; 3) Baseline + multi-scale StyleGAN features. We also introduce PSNR and SSIM from [10] as the image quality metrics for evaluating face correction.

The quantitative results are summarized in Tab. 2. Both variants 2) and 3) lead to notable improvements in facial landmarks, with variant 3) showing a more significant enhancement, achieving the best Landmark Distance of 2.35. This highlights the effectiveness of multi-scale features in enhancing face correction. It is noteworthy that variant 3) yields better landmark results but slightly lower PSNR/SSIM values. Given that 2) utilizes high-level scale features, this suggests that the crucial information for face structure correction primarily originates from high-level features, whereas low-level features predominantly influence color and details, thereby improving PSNR/SSIM. Fig. 8 shows the qualitative comparisons between different variants as well as the StyleGAN inversion results. Despite the StyleGAN inversion results presenting variations in facial identity, they offer valuable structural guidance for face correction. The results achieved with multiscale features demonstrate improved landmark similarity to StyleGAN inversion compared to those from a single-scale model, highlighting the advantages of multiscale feature fusion in enhancing the accuracy and effectiveness of face correction.

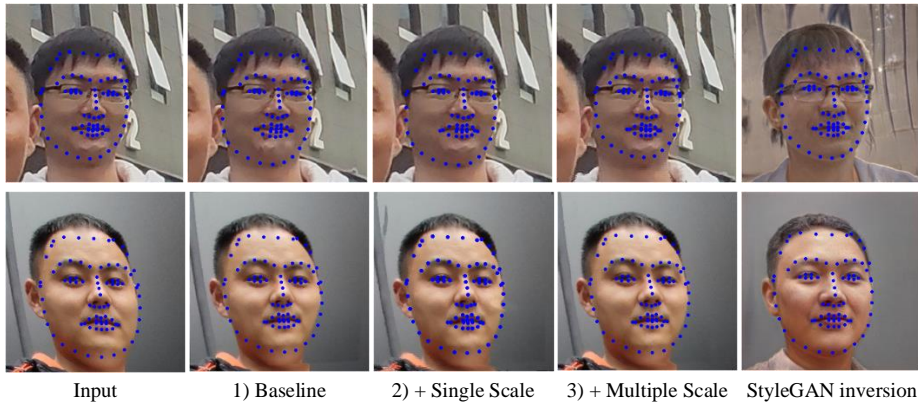


Fig. 8: Analyses of multiple scale face prior for face correction.

Table 3: Quantitative comparison of our proposed method and previous works. Results of Shih *et al.* [38] and Tan *et al.* [40] are directly taken from their papers.

Method	Note		Vivo		All dataset		
	LineAcc	ShapeAcc	LineAcc	ShapeAcc	LineAcc	ShapeAcc	Landmark Dis.
Shih [38]	-	-	-	-	66.143	97.253	6.035
Tan [40]	68.683	97.115	65.148	98.363	66.784	97.490	-
Zhu [48]	66.381	97.746	66.572	96.076	67.209	97.500	5.840
Ours	69.891	98.697	65.459	99.091	67.304	99.012	5.013

4.5 Comparison with Others Methods

We conduct comprehensive comparisons with competing SOTA methods, *i.e.*, Shih *et al.* [38] and Zhu *et al.* [48], and evaluate its effectiveness in both straight line and face restoration tasks. To ensure a fair comparison, all results are obtained without additional calibration preprocessing. Tab. 3 provides a detailed comparison of the results achieved by our method against those of existing approaches. The quantitative comparison highlights our method’s enhanced ability to preserve straight lines and accurately restore facial features. Notably, our approach significantly reduces the ShapeAcc error by 60%, from 2.5 to 1.0. To complement the quantitative evaluation, we also conducted visual comparisons, as depicted in Fig. 9. We can observe that our line correction results consistently surpass those of other methods, highlighting the advantage of using a symmetric prior. Furthermore, the faces corrected by our approach appear more natural and are free from artifacts, in contrast to those by Shih *et al.* in the second row. These comparisons further affirm the efficacy of our method, demonstrating its capability to maintain the original face size while effectively correcting distortions. Through these analyses, we have demonstrated the superior performance of our approach in handling both background and facial distortions.

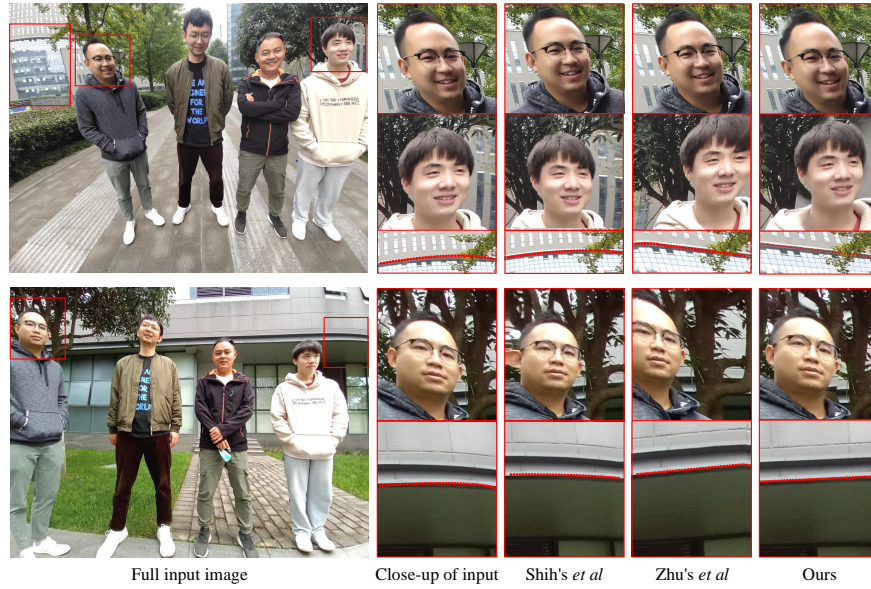


Fig. 9: Qualitative results of different correction methods. Our results are significantly better in line correction and yield more natural-looking faces without artifacts.

4.6 Limitation

The generative face prior may struggle to handle faces with large poses. This can be alleviated by expanding the face diversity for the generative models. Another limitation is the unnatural body region (*e.g.*, obviously longer feet in Fig. 9). We will solve it by introducing the generative body prior in the future.

5 Conclusion

This paper introduces an innovative method to correct wide-angle distortion by leveraging the synergistic capabilities of the generative face prior and geometric symmetry prior. Our approach not only successfully mitigates background distortions through LineCNet, which is trained with symmetric regularization, but also adeptly addresses intricate facial distortions with the incorporation of a StyleGAN prior. The presented visualizations of the correction flow map and facial landmarks highlight the effectiveness of these two priors. Through comprehensive experimentation, our method sets a new benchmark, outperforming existing techniques and showcasing significant visual enhancements.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (NSFC) under grant U22B2035.

References

1. Abdal, R., Qin, Y., Wonka, P.: Image2stylegan: How to embed images into the stylegan latent space? In: Proceedings of the IEEE/CVF International Conference on Computer Vision (October 2019)
2. Alaluf, Y., Patashnik, O., Cohen-Or, D.: Restyle: A residual-based stylegan encoder via iterative refinement. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6711–6720 (2021)
3. Bousaid, A., Theodoridis, T., Nefti-Meziani, S., Davis, S.: Perspective distortion modeling for image measurements. *IEEE Access* **8**, 15322–15331 (2020)
4. Cao, J., Lin, B., Chen, Z.: Adaptive content-aware correction for wide-angle portrait photos. In: Computational Visual Media Conference (2022)
5. Carroll, R., Agarwala, A., Agrawala, M.: Image warps for artistic perspective manipulation. *Acm Transactions on Graphics* **29**(4CD), p.127.1–127.9 (2010)
6. Carroll, R., Agrawala, M., Agarwala, A.: Optimizing content-preserving projections for wide-angle images. *ACM Trans. Graph* (2009)
7. Carroll, R., Agrawala, M., Agarwala, A.: Optimizing content-preserving projections for wide-angle images. *ACM Trans. Graph.* **28**(3), 43 (2009)
8. Chan, K.C., Wang, X., Xu, X., Gu, J., Loy, C.C.: Glean: Generative latent bank for large-factor image super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14245–14254 (2021)
9. Chen, C., Li, X., Yang, L., Lin, X., Zhang, L., Wong, K.Y.K.: Progressive semantic-aware style transformation for blind face restoration. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11896–11905 (2021)
10. Chen, C., Mo, J.: IQA-PyTorch: Pytorch toolbox for image quality assessment. [Online]. Available: <https://github.com/chaofengc/IQA-PyTorch> (2022)
11. Chen, C., Shi, X., Qin, Y., Li, X., Han, X., Yang, T., Guo, S.: Real-world blind super-resolution via feature matching with implicit high-resolution priors. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 1329–1338 (2022)
12. Dolhansky, B., Ferrer, C.C.: Eye in-painting with exemplar generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018)
13. Fitzgibbon, A.W.: Simultaneous linear estimation of multiple view geometry and lens distortion. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. vol. 1, pp. I–I (2001)
14. Gu, J., Shen, Y., Zhou, B.: Image processing using multi-code gan prior. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3012–3021 (2020)
15. Gu, S., Zuo, W., Guo, S., Chen, Y., Chen, C., Zhang, L.: Learning dynamic guidance for depth image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2017)
16. Hold-Geoffroy, Y., Sunkavalli, K., Eisenmann, J., Fisher, M., Lalonde, J.F.: A perceptual measure for deep single image camera calibration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)
17. Jiang, Y., Chan, K.C., Wang, X., Loy, C.C., Liu, Z.: Robust reference-based super-resolution via c2-matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2103–2112 (2021)

18. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. In: International Conference on Learning Representations (2018)
19. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4401–4410 (2019)
20. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (June 2020)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
22. Li, X., Chen, C., Zhou, S., Lin, X., Zuo, W., Zhang, L.: Blind face restoration via deep multi-scale component dictionaries. In: European Conference on Computer Vision (2020)
23. Li, X., Hu, G., Zhu, J., Zuo, W., Wang, M., Zhang, L.: Learning symmetry consistent deep cnns for face completion. IEEE Transactions on Image Processing pp. 7641–7655 (2020)
24. Li, X., Li, W., Ren, D., Zhang, H., Wang, M., Zuo, W.: Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature fusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020)
25. Li, X., Liu, M., Ye, Y., Zuo, W., Lin, L., Yang, R.: Learning warped guidance for blind face restoration. In: European Conference on Computer Vision (2018)
26. Li, X., Zhang, S., Zhou, S., Zhang, L., Zuo, W.: Learning dual memory dictionaries for blind face restoration. IEEE Transactions on Pattern Analysis and Machine Intelligence (2022)
27. Li, X., Zuo, W., Loy, C.C.: Learning generative structure prior for blind text image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023)
28. Li, Y., Huang, J.B., Ahuja, N., Yang, M.H.: Deep joint image filtering. In: European Conference on Computer Vision (2016)
29. Liu, M., Wei, Y., Wu, X., Zuo, W., Zhang, L.: Survey on leveraging pre-trained generative adversarial networks for image editing and restoration. Science China Information Sciences **66**(5), 151101 (2023)
30. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L.: Re-paint: Inpainting using denoising diffusion probabilistic models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11461–11471 (2022)
31. Menon, S., Damian, A., Hu, S., Ravi, N., Rudin, C.: Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2437–2445 (2020)
32. Nazeri, K., Ng, E., Joseph, T., Qureshi, F., Ebrahimi, M.: Edgeconnect: Structure guided image inpainting using edge prediction (2019)
33. Pan, X., Zhan, X., Dai, B., Lin, D., Loy, C.C., Luo, P.: Exploiting deep generative prior for versatile image restoration and manipulation. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(11), 7474–7489 (2021)
34. Ren, Y., Yu, X., Zhang, R., Li, T.H., Liu, S., Li, G.: Structureflow: Image inpainting via structure-aware appearance flow. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2019)

35. Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., Cohen-Or, D.: Encoding in style: a stylegan encoder for image-to-image translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2287–2296 (2021)
36. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer (2015)
37. Santana-Cedr s, D., Gomez, L., Alem n-Flores, M., Salgado, A., Esclar n, J., Mazorra, L., Alvarez, L.: Automatic correction of perspective and optical distortions. *Computer Vision and Image Understanding* **161**, 1–10 (2017)
38. Shih, Y., Lai, W.S., Liang, C.K.: Distortion-free wide-angle portraits on camera phones. *ACM Transactions on Graphics (TOG)* **38**(4), 1–12 (2019)
39. Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V.: Resolution-robust large mask inpainting with fourier convolutions. arXiv preprint arXiv:2109.07161 (2021)
40. Tan, J., Zhao, S., Xiong, P., Liu, J., Fan, H., Liu, S.: Practical wide-angle portraits correction with deep structured models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3498–3506 (2021)
41. Tov, O., Alaluf, Y., Nitzan, Y., Patashnik, O., Cohen-Or, D.: Designing an encoder for stylegan image manipulation. arXiv preprint arXiv:2102.02766 (2021)
42. Valente, J., Soatto, S.: Perspective distortion modeling, learning and compensation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (June 2015)
43. Wang, X., Li, Y., Zhang, H., Shan, Y.: Towards real-world blind face restoration with generative facial prior. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9168–9178 (2021)
44. Yang, T., Ren, P., Xie, X., Zhang, L.: Gan prior embedded network for blind face restoration in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 672–681 (2021)
45. Zhang, Q., Li, H., Wang, Q.: Wide-angle rectification via content-aware conformal mapping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17357–17365 (June 2023)
46. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. vol. 1, pp. 666–673 (1999)
47. Zhao, Y., Huang, Z., Li, T., Chen, W., LeGendre, C., Ren, X., Shapiro, A., Li, H.: Learning perspective undistortion of portraits. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7849–7859 (2019)
48. Zhu, F., Zhao, S., Wang, P., Wang, H., Yan, H., Liu, S.: Semi-supervised wide-angle portraits correction by multi-scale transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19689–19698 (2022)
49. Zorin, D., Barr, A.H.: Correction of geometric perceptual distortions in pictures. In: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques. pp. 257–264 (1995)