Generative Adversarial Networks Based Facial Image Restoration for Enhanced Image Quality

Submitted November 7, in partial fulfillment of
the conditions for the award of the degree **Bsc in Internet of Things and Robotics Engineering.**

Pallab Sarkar,Saurabh Joardar, Sieadatun Nahar Kanon, Azmary Sara
1801016,1801028,1901018, 190106

Supervised by Mahir Mahbub

Bangabandhu Sheikh Mujibur Rahman Digital University,Bangladesh

I hereby declare that this dissertation is all my own work, except as indicated in the text:

Signature ___________________

Date ______ / ______ / ______

I hereby declare that I have all necessary rights and consents to publicly distribute this dissertation via the Bangabandhu Sheikh Mujibur Rahman Digital University,Bangladesh

Public access to this dissertation is restricted until: 07/11/2023

Abstract

Blind face restoration usually relies on facial priors, such as facial geometry or reference prior, to restore realistic and faithful details. However, very low-quality inputs cannot offer accurate geometric prior while high-quality references are inaccessible, limiting the applicability in real-world scenarios. In this work, we propose a GFP-GAN that leverages rich and diverse priors encapsulated in a pre-trained face GAN for blind face restoration. This Generative Facial Prior (GFP) is incorporated into the face restoration process via spatial feature transform layers, which allow our method to achieve a good balance of realness and fidelity. Extensive experiments show that our method achieves superior performance to the prior art on both synthetic and real-world datasets.

Acknowledgements

We would like to express our deepest appreciation to our supervisor, **Mahir Mahbub**, for his invaluable guidance, support, and mentorship throughout the project. His insights and expertise have been instrumental in shaping our work and pushing us to achieve our best. We also wish to acknowledge our fellow team members for their dedication, collaboration, and hard work, which were essential in realizing our project goals. Together, we navigated the challenges and celebrated the successes, and this project's completion stands as a testament to our collective effort and commitment. Lastly, we acknowledge the support of our families, friends, and peers who provided encouragement and understanding during this endeavor.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Blind face restoration aims at recovering high-quality faces from low-quality counterparts suffering from unknown degradation, such as low-resolution, noise, blur, compression artefacts, etc. When applied to real-world scenarios, it becomes more challenging, due to more complicated degradation, diverse poses and expressions. Previous works typically exploit face-specific priors in face restoration, such as facial landmarks [9], parsing maps, and facial component heatmaps [69], and show that those geometry facial priors are pivotal to recovering accurate face shapes and details. However, those priors are usually estimated from input images and inevitably degrades with very low-quality inputs in the real world. In addition, despite their semantic guidance, the above priors contain limited texture information for restoring facial details. Another category of approaches investigates reference priors, i.e., high-quality guided faces or facial component dictionaries, to generate realistic results and alleviate the dependency on degraded inputs. However, the inaccessibility of high-resolution references limits its practical applicability, while the limited capacity of dictionaries restricts its diversity and richness of facial details. In this study, we leverage Generative Facial Prior (GFP) for real-world blind face restoration, i.e., the prior implicitly encapsulated in pre-trained face Generative Adversarial Network (GAN) models such as StyleGAN. These face GANs are capable of generating faithful faces with a high degree of variability, thereby providing rich and diverse priors such as geometry, facial textures and colours, making it possible to jointly restore facial details and enhance colours. However, it is challenging to incorporate such generative priors into the restoration process. Previous attempts typically use GAN inversion. They first 'invert' the degraded image back to a latent code of the pre-trained GAN, and then conduct expensive image-specific optimization to reconstruct images. Despite visually realistic outputs, they usually produce images with low fidelity, as the low-dimension latent codes are insufficient to guide accurate restoration. To address these challenges, we propose the GFP-GAN with delicate designs to achieve a good balance of realness and fidelity in a single forward pass. Specifically, GFP- GAN consists of a degradation removal module and a pre-trained face GAN as facial prior. They are connected by a direct latent code mapping, and several Channel-Split Spatial Feature Transform (CS-SFT) layers in a coarse-to-fine manner. The proposed CS-SFT layers perform spatial modulation on a split of features and leave the left features to directly pass through for better information preservation, allowing our method to effectively incorporate generative prior while retraining high fidelity. Besides, we introduce facial component loss with local discriminators to further enhance perceptual facial details, while employing identity-preserving loss to further improve fidelity. We summarize the contributions as follows. We leverage rich and diverse generative facial priors for blind face restoration. Those priors contain sufficient facial textures and colour information, allowing us to jointly perform face restoration and colour enhancement. We propose the GFP-GAN framework with delicate designs of architecture and losses to incorporate generative facial prior. Our GFP-GAN with CS-SFT layers achieves a good balance of fidelity and texture faithfulness in a single forward pass. Extensive experiments show that our method achieves superior performance to prior art on both synthetic and real-world datasets.

1.1 Motivation

The motivation for writing the paper lies in the pursuit of excellence, adaptability to real-world challenges, robustness, and continued relevance in a dynamic research landscape. By revisiting and improving the existing work, the authors can contribute to the advancement of the field and ensure that their method remains a valuable tool for various applications. The motivation behind this project report lies in the recognition of the following key factors:

1. Advancing State-of-the-Art: In the rapidly evolving field of computer vision and image processing, it is imperative to continuously advance the state-of-the-art to stay relevant and competitive. By revisiting and rewriting this paper, the authors have an opportunity to incorporate recent advancements in generative models and deep learning techniques, ensuring that their proposed method remains at the forefront of blind face restoration research.

2. Improving Real-World Applicability: The paper acknowledges the limitations of existing methods in real-world scenarios, where very low-quality inputs and inaccessible high-quality references are common. By revising the paper, the authors can focus on further enhancing the applicability of their proposed Generative Facial Prior (GFP)-GAN to address these real-world challenges more effectively.

3. Enhancing Robustness and Generalization: The original paper introduces the GFP-GAN approach, but there may be room for improving the robustness and generalization of the method. A rewrite can focus on refining the model to perform well across a wider range of scenarios, including those with highly degraded input images and diverse facial characteristics.

4. Expanding Experimentation: While the paper mentions superior performance on synthetic and real-world datasets, a rewritten version could include an expanded set of experiments. This may involve testing the method on more diverse and challenging datasets, thereby providing a more comprehensive evaluation of its capabilities and limitations.

1.2 Aims and Objectives

This involves incorporating recent developments in generative models and deep learning to ensure the proposed Generative Facial Prior (GFP)-GAN remains cutting-edge. Another key aim is to make the blind face restoration method more relevant in real-world scenarios. This means addressing common challenges, such as low-quality input images and the unavailability of high-quality references, to enhance its practical applicability. Additionally, the objective is to improve the robustness and generalization of the GFP-GAN model. This will help the method perform effectively across a wider range of scenarios, including those involving highly degraded input images and diverse facial features. Expanding the scope of experimentation is another aim. This includes testing the method on a more diverse set of datasets, both synthetic and real-world, to offer a more comprehensive evaluation of its capabilities and limitations. Furthermore, addressing feedback from the research community and users is important. By incorporating suggestions and improvements, the method can become more user-friendly and effective in addressing real-world needs. Ethical considerations are also essential. The paper aims to discuss potential biases and privacy concerns related to facial restoration techniques, proposing methods and strategies to mitigate these issues and uphold ethical standards.

Finally, the paper seeks to serve as an educational resource. This involves providing a detailed and up-to-date guide for researchers, practitioners, and students interested in blind face restoration and generative models. Clear explanations, improved visual aids, and practical examples are included to enhance understanding.

Chapter 2

Related Work

Image Restoration typically includes super-resolution, denoising, deblurring and compression removal [5]. To achieve visually pleasing results, a generative adversarial network is usually employed as loss supervision to push the solutions closer to the natural manifold, while our work attempts to leverage the pre-trained face GANs as generative facial priors (GFP) Face Restoration. Based on general face hallucination, two typical face-specific priors: geometry priors and reference priors, are incorporated to further improve the performance [6]. The geometry priors include facial landmarks, face parsing maps and facial component heatmaps. However, 1) those priors require estimations from low-quality inputs and inevitably degrade in real-world scenarios. 2) They mainly focus on geometry constraints and may not contain adequate details for restoration. Instead, our employed GFP does not involve an explicit geometry estimation from degraded images and contains adequate textures inside its pre-trained network [2]. Reference priors usually rely on reference images of the same identity. To overcome this issue, DFD- -Net suggests constructing a face dictionary of each component (e.g., eyes, mouth) with CNN features to guide the restoration [4]. However, DFDNet mainly focuses on components in the dictionary and thus degrades in the regions beyond its dictionary scope (e.g., hair, ears and face contour), instead, our GFP-GAN could treat faces as a whole to restore. Moreover, the limited size of the dictionary restricts its diversity and richness, while the GFP could provide rich and diverse priors including geometry, textures and colours [3]. Generative Priors of trained GANs are previously exploited by GAN inversion, whose primary aim is to find the closest latent codes given an input image. PULSE iteratively optimizes the latent code of StyleGAN until the distance between outputs and inputs is below a threshold. mGANprior attempts to optimize multiple codes to improve the reconstruction quality. However, these methods usually produce images with low fidelity, as the low-dimension latent codes are insufficient to guide the restoration. In contrast, our proposed CS-SFT modulation layers enable prior incorporation of multi-resolution spatial features to achieve high fidelity [1]. Besides, expensive iterative optimization is not required in our GFP-GAN during inference. Channel Split Operation is usually explored to design compact models and improve model representation ability. MobileNet proposes depthwise convolutions and GhostNet splits the convolutional layer into two parts and uses fewer filters to generate intrinsic feature maps [7]. Dual path architecture in DPN enables feature re-usage and new feature exploration for each path, thus improving its representation ability. A similar idea is also employed in super-resolution. Our CS-SFT layers share similar spirits but with different operations and purposes. We adopt spatial feature transformation on one split and leave the left split as an identity to achieve a good balance of realness and fidelity. Local Component Discriminators. The local discriminator is proposed to focus on local patch distributions. When applied to faces, those discriminative losses are imposed on separate semantic facial regions. Our introduced facial component loss also adopts such designs but with further style supervision based on the learned criminative features [8].

Chapter 3

Methodology

3.1 Overview of GFP GAN

We describe GFP-GAN framework in this section. Given an input facial image x suffering from unknown degradation, the aim of blind face restoration is to estimate a high-quality image yˆ, which is as similar as possible to the ground-truth image y, in terms of realness and fidelity. The overall framework of GFP-GAN is depicted in Fig. 2. GFP-GAN is comprised of a degradation removal module (U-Net) and a pre-trained face GAN (such as Style- GAN2 [5]) as prior. They are bridged by latent code mapping and several Channel-Split Spatial Feature Transform (CS-SFT) layers. Specifically, the degradation removal module is designed to remove complicated degradation and extract two kinds of features, i.e. 1) latent features Flatent to map the input image to the closest latent code in StyleGAN2, and 2) multi-resolution spatial features Fspatial for modulating the StyleGAN2 features. After that, Flatent is mapped to intermediate latent codes W by several linear layers. Given the close latent code to the input image, StyleGAN2 could generate intermediate convolutional features, denoted by FGAN. These features provide rich facial details captured in the weights of pre-trained GAN. Multi-resolution features Fspatial are used to spatially modulate the face GAN features FGAN with the proposed CS-SFT layers in a coarse-to-fine manner, achieving realistic results while preserving high fidelity. During training, except for the global discriminative loss, we introduce facial component loss with discriminators to enhance the significant face components.

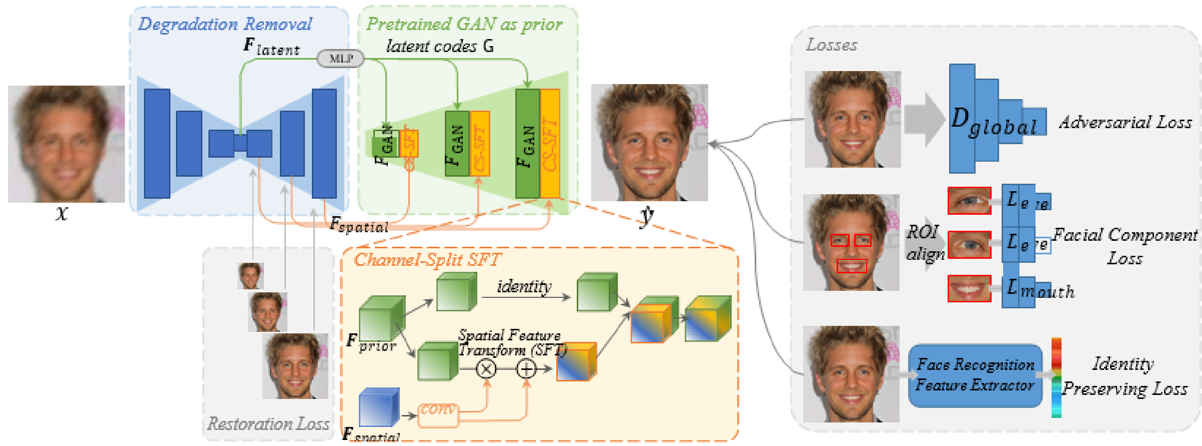

Figure 3.1: Overview of GFP-GAN framework. It consists of a degradation removal module (U-Net) and a pre-trained face GAN as a facial prior. They are bridged by latent code mapping and several Channel-Split Spatial Feature Transform (CS-SFT) layers. During training, we employ 1) intermediate restoration losses to remove complex degradation, 2) Facial component loss with discriminators to enhance facial details, and 3) identity preserving loss to retain face identity.

3.2 Degradation Removal Module

Real-world blind face restoration faces with complicated and severer degradation, which is typically a mixture of low-resolution, blur, noise and JPEG artifacts. The degra- dation removal module is designed to explicitly remove the above degradation and extract 'clean' features Flatent and Fspatial, alleviating the burden of subsequent modules. We adopt the U-Net structure as our degradation remove module, as it could 1) increase receptive field for large blur elimination, and 2) generate multi-resolution features.

The latent features Flatent is used to map the input image to the closest latent code in StyleGAN2. The multi-resolution spatial features Fspatial are used to modulate the StyleGAN2 features. In order to have intermediate supervision for removing degradation, we employ the L1 restoration loss in each resolution scale in the early stage of training. Specifically, we also output images for each resolution scale of the U-Net decoder and then restrict these outputs to be close to the pyramid of the ground-truth image.

3.3 Generative Facial Prior and Latent Code Mapping

The concepts of "Generative Facial Prior" and "Latent Code Mapping" are central to the field of generative modelling and face restoration. Let's briefly explore each of these concepts:

- **Generative Facial Prior:** A "Generative Facial Prior" is a framework or model that encapsulates prior knowledge about human faces in a generative manner. It represents the statistical and structural information that is typically associated with facial images. This prior can encompass various aspects of facial attributes, including facial geometry, texture, expressions, and other relevant features. It is often learned from a large dataset of facial images. In the context of face restoration, a Generative Facial Prior is a valuable resource as it guides the generation or restoration process, ensuring that the output remains faithful to the expected facial characteristics.

- **Latent Code Mapping:** "Latent code mapping" refers to the process of mapping input data or conditions, typically represented as latent variables or codes, to a corresponding output in a generative model. In the case of generative models like Variational Autoencoders (VAEs) or Generative Adversarial Networks (GANs), this mapping involves transforming random or encoded latent codes into data that conforms to a specific distribution, such as an image. In facial image generation or restoration, latent code mapping could involve taking a set of latent variables (e.g., a random vector or learned codes) and mapping them to a realistic facial image. These codes can represent various facial attributes like pose, expression, identity, and lighting conditions. The quality and effectiveness of this mapping process are critical in ensuring that the generated or restored facial images are coherent, realistic, and adhere to the Generative Facial Prior.

3.4 Channel Split Spatial Feature Transform

This term refers to a particular idea or method in the fields of deep learning, computer vision, or image processing. As of my most recent knowledge update in January 2022, this phrase is not, however, commonly acknowledged in the area.

- **Channel Split:** In image processing and neural networks, "channel" often refers to individual colour or feature channels in an image. For example, in an RGB image, you have three channels: red, green, and blue. "Channel splitting" may involve separating or processing these individual channels independently.

- **Spatial Feature Transform:** This term suggests a transformation applied to image features that consider the spatial arrangement of elements. It could relate to techniques like convolutional layers in deep learning, which are designed to capture spatial patterns in data.

Chapter 4

Implementation

4.1 Dataset

We use the Kaggle dataset "humanface8000" for this project. There are 16.3k files in the collection. There are two folders among the landscape photos: "color" and "gray." There are 8164 files in the gray directory and 8164 files in the color directory. Deep learning may be applied to the restoration of ancient images. This dataset is automatically curated by applying a filter to add noise or scratches to the picture using a function and then mapping those scratches onto an overlay image. Here, we cloned the dataset to complete the project.

4.2 Implementation

We adopt the pre-trained StyleGAN2 outputs as our generative facial prior. The channel multiplier of StyleGAN2 is set to one for compact model size. The UNet for degradation removal consists of seven downsamples and seven upsamples, each with a residual block. For each CSSFT layer, we use two convolutional layers to generate the affine parameters and respectively. The training mini-batch size is set to 12. We augment the training data with horizontal flips and color jittering. We consider three components: left eye, right eye, and mouth for face component loss as they are perceptually significant. Each component is cropped by ROI aligned with face landmarks provided in the origin training dataset. We train our model with Adam Optimizer for a total of 800k iterations. We implement our models with the PyTorch framework and train them using four NVIDIA Tesla P40 GPUs.

4.3 Comparisons with State-of-the-art Methods

We compare our GFP-GAN with several state-of-the-art face restoration methods: HiFaceGAN, DFDNet, PSFRGAN, and SuperFAN. GAN inversion methods for face restoration: PULSE and mGANprior is also included for comparison. We also compare our GFP-GAN with image restoration methods: RCAN, ESRGAN, and DeblurGANv2, and we finetune them on our face training set for fair comparisons. We adopt their official codes except for Super-FAN, for which we use a re-implementation. For the evaluation, we employ the widely used non-reference perceptual metrics: FID and NIQE. We also adopt pixel-wise metrics (PSNR and SSIM) and the perceptual metric (LPIPS) for the CelebA-Test with Ground Truth (GT). We measure the identity distance with angels in the ArcFace feature embedding, where smaller values indicate a closer identity to the GT.

- **Synthetic CelebA-Test:** The comparisons are conducted under two settings: 1) blind face restoration whose inputs and outputs have the same resolution. 2) 4× face super resolution. Note that our method could take unsampled images as inputs for face super-resolution. In both settings, GFP-GAN achieves the lowest LPIPS, indicating that our results are perceptually close to the ground truth. GFP-GAN also obtained the lowest FID and NIQE, showing that the outputs have a close distance to the real face distribution and natural image distribution, respectively. Besides the perceptual performance, the method also retains a better identity, as indicated by the smallest degree in the face feature embedding. Note that 1) the lower FID and NIQE of our method than GT do not indicate that our performance is better than GT, as those 'perceptual' metrics are well correlated with the human opinion scores on a coarse scale but not always well correlated on a finer scale [2]; 2) the pixel-wise metrics PSNR and SSIM are no correlation well with the subjective

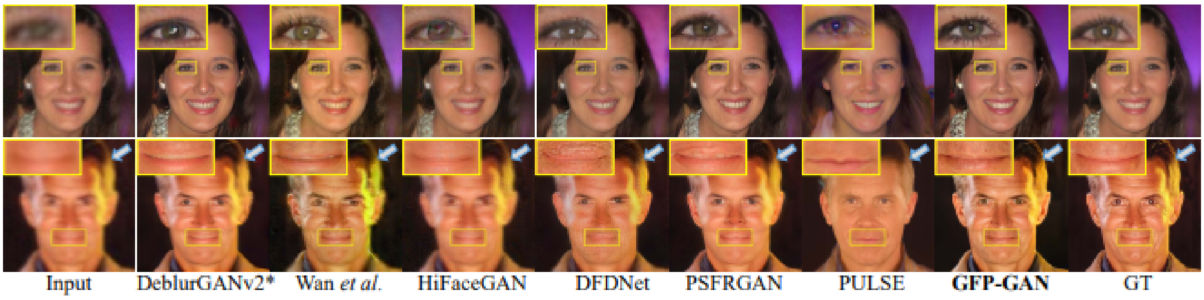

Figure 4.1: Qualitative comparison on the CelebA-Test for blind face restoration. Our GFP-GAN produces faithful details in eyes, mouth and hair.

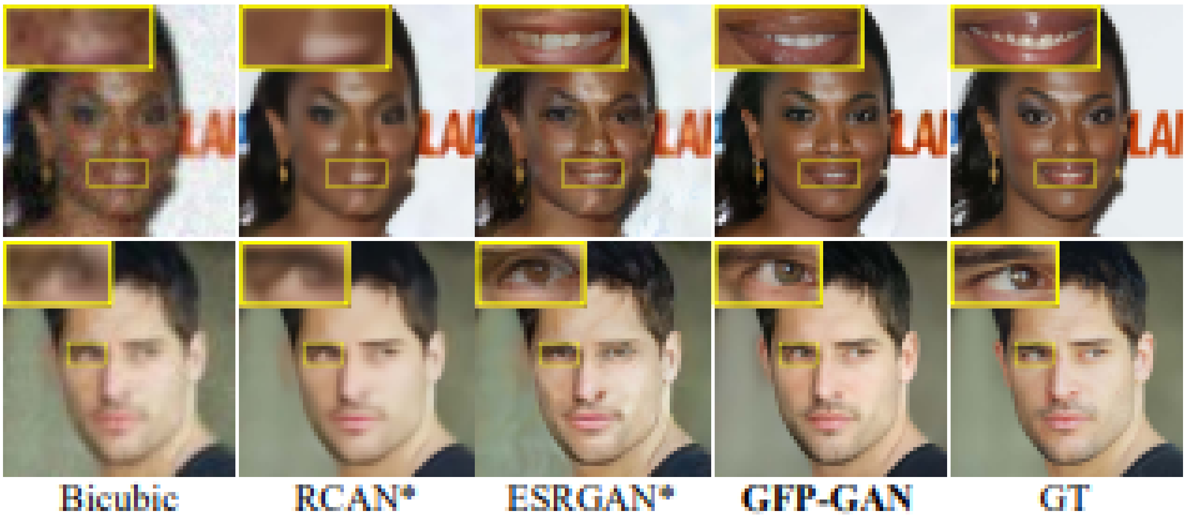

Figure 4.2: Comparison on the CelebA-Test for ×4 face super-resolution. Our GFP-GAN restores realistic teeth and faithful eye gaze direction.

evaluation of human observers [2] and our model is not good at these two metrics. Qualitative results are presented in Fig. 4.1 and Fig. 4.2 1) Thanks to the powerful generative facial prior, our GFP-GAN recovers faithful details in the eyes (pupils and eyelashes), teeth, etc. 2) Our method treats faces as whole in restoration and could also generate realistic hair, while previous methods that rely on component dictionaries (DFDNet) or parsing maps (PSFRGAN) fail to produce faithful hair textures (2nd row, Fig. 4.1). 3) GFP-GAN can retain row, Fig. 4.1). And in Fig. 4.2, GFP-GAN also restores reasonable eye gaze direction.

- **Real-World LFW, CelebChild and WedPhoto-Test:** The qualitative comparisons are shown in Fig. 4.3. GFP-GAN could jointly conduct face restoration and color enhancement for real-life photos with the powerful generative prior. Our

Figure 4.3: Qualitative comparisons on three real-world datasets.

method could produce plausible and realistic faces on complicated real-world degradation, while other methods fail to recover faithful facial details or produce artifacts (especially in the WebPhoto-Test in Fig 4.3). Besides the common facial components like eyes and teeth, GFP-GAN also performs better in hair and ears, as the GFP prior takes the whole face into consideration rather than separate parts. With SC SFT layers, our model can achieve high fidelity. As shown in the last row of Fig. 4.3, most previous methods fail to recover the closed eyes, while ours could successfully restore them with fewer artifacts.

Chapter 5

Ablation Studies

5.1 Pretrained GAN as GFP

Pretrained GAN provides rich and diverse features for restoration. A performance drop is observed if we do not use the generative facial prior, as shown in Fig. 5.1.

5.2 Pyramid Restoration Loss

Pyramid restoration loss is employed in the degradation removal module and strengthens the restoration ability for complicated degradation in the real world. Without this intermediate supervision, the multi-resolution spatial features for subsequent modulations may still have degradation, resulting in inferior performance, as shown in Fig. 5.1.

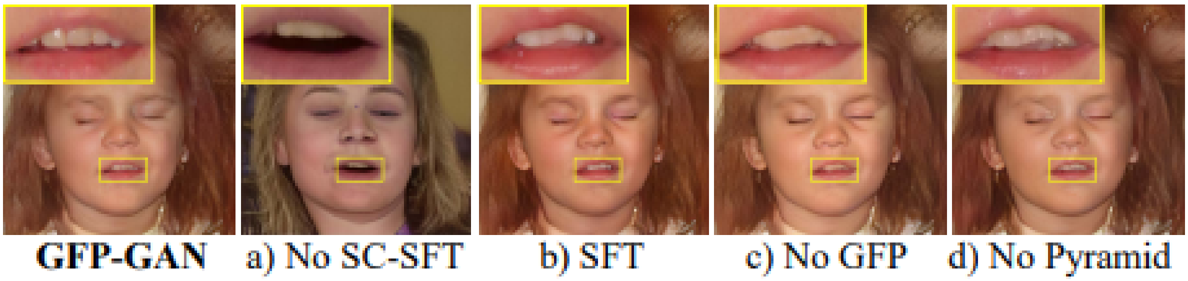

Figure 5.1: Ablation studies on GFP prior and pyramid restoration loss

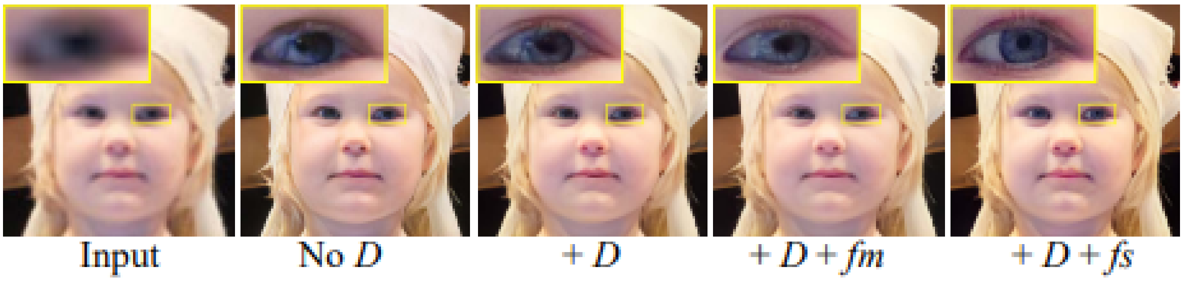

Figure 5.2: Ablation studies on facial component loss.

5.3 Facial Component Loss

We compare the results of 1) removing all the facial component loss, 2) only keeping the component discriminators, 3) adding extra feature matching loss, and 4) adopting extra feature style loss based on Gram statistics. It is shown in Fig 5.2 that the component discriminators with feature style loss could better capture the eye distribution and restore the plausible details.

Chapter 6

Summary and Reflections

6.1 Conclusion

We have proposed the GFP-GAN framework that leverages the rich and diverse generative facial prior to the challenging blind face restoration task. This prior is incorporated into the restoration process with channel-split spatial feature transform layers, allowing us to achieve a good balance of realness and fidelity. Extensive comparisons demonstrate the superior capability of GFP-GAN in joint face restoration and color enhancement for real-world images, outperforming prior art.

Bibliography

[1] Binder, A., and Yuan, J.-S. Optimization of an enhancement-mode algan/gan/algan dhfet towards a high breakdown voltage and low figure of merit. In *2017 IEEE 5th Workshop on Wide Bandgap Power Devices and Applications (WiPDA)* (2017), pp. 122–126.

[2] Chen, B., Liu, X., Zheng, Y., Zhao, G., and Shi, Y.-Q. A robust gan-generated face detection method based on dual-color spaces and an improved xception. *IEEE Transactions on Circuits and Systems for Video Technology 32*, 6 (2022), 3527–3538.

[3] Conmy, A., Mukherjee, S., and Schönlieb, C.-B. Stylegan-induced data-driven regularization for inverse problems. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2022), pp. 3788–3792.

[4] Khan, M. Z., Jabeen, S., Khan, M. U. G., Saba, T., Rehmat, A., Rehman, A., and Tariq, U. A realistic image generation of face from text description using the fully trained generative adversarial networks. *IEEE Access 9* (2021), 1250–1260.

[5] Kumar, A., and Vatsa, A. Influence of gfp gan on melanoma classification. In *2022 IEEE Integrated STEM Education Conference (ISEC)* (2022), pp. 334–339.

[6] Luo, X., Wang, Y., Hao, Y., Li, X.-j., Liu, C.-M., Fei, X.-X., Yu, C.-H., and Cao, F. Research of single-event burnout and hardening of algan/gan-based misfet. *IEEE Transactions on Electron Devices 66*, 2 (2019), 1118–1122.

[7] Wu, X., Zhang, Q., Wu, Y., Wang, H., Li, S., Sun, L., and Li, X. F^3a-gan: Facial flow for face animation with generative adversarial networks. *IEEE Transactions on Image Processing 30* (2021), 8658–8670.

[8] You, J., Huang, G., Han, T., Yang, H., and Shen, L. A unified framework from face image restoration to data augmentation using generative prior. *IEEE Access 11* (2023), 2907–2919.