International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET229344

Video Regeneration and Quality Enhancer using GFP-GAN

Girija V[1], Sunny Nehra[2], Himanshu Kumar[2], Avinash Yadav[2], Karan R [2]
[1]Assistant Professor, CiTeh, Bangalore, Karnataka, India
[2]Student, CiTech, Bangalore, Karnataka, India

Article Info
Volume 9, Issue 3

Page Number : 148-151

Publication Issue :
May-June-2022

Article History
Accepted : 05 May 2022
Published: 15 May 2022

ABSTRACT

Blind face restoration usually relies on facial priors, such as facial geometry prior 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 GFP-GAN that leverages rich and diverse priors encapsulated in a pretrained 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. Thanks to the powerful generative facial prior and delicate designs, our GFP-GAN could jointly restore facial details and enhance colors with just a single forward pass, while GAN inversion methods require image-specific optimization at inference. Extensive experiments show that our method achieves superior performance to prior art on both synthetic and real-world datasets.

Keywords : GFP-GAN, Generative Facial Prior, Video Regeneration Quality Enhancer

I. INTRODUCTION

Blind face restoration aims at recovering high-quality faces from the low-quality counterparts suffering from unknown degradation, such as low-resolution, noise, blur, compression artifacts etc.

Previous works typically exploit face-specific priors in face restoration, such as facial landmarks, parsing maps, facial component heat maps, and show that those geometry facial priors are pivotal to recover accurate face shape 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. GFPGAN consists of a degradation removal module and a pretrained face GAN as facial prior.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.We generate a different coordinate value for the grid vertices. Our framework adopts a differentiable rendering layer for self-supervised training. HPENet is used to conduct head pose estimation on guided face image B and store face position feature information and pose direction vector. We input the low-quality face image A and guide face image B to fit 3D face information through the PTNet and convert the face posture with the mapping function, so as to generate face B′ with the same posture as face A.

II. Literature Survey

Blind Face Restoration [1]

Multimedia, including audio, image, and video, etc., is a ubiquitous part of modern life. Quality evaluation, both objective and subjective, is of fundamental importance for various multimedia applications. In this letter, a novel quality-aware feature is proposed for blind/no-reference (NR) image quality assessment (IQA). The new quality-aware feature is generated from the proposed joint generalized local binary pattern (GLBP) statistics.

The proposed joint generalized local binary pattern (GLBP) statistics generate the new quality-aware feature. The images are initially divided into multi-scale sub-band images using the Laplacian of Gaussian (LOG) filters in this method. The sub-band images are then encoded using the proposed GLBP operator, and quality-aware features are derived from the joint GLBP histograms derived from each sub-band image's encoding maps. Finally, the quality-aware features are mapped to the image's subjective quality score for NR-IQA using support vector regression (SVR). The suggested method is substantially connected to subjective quality ratings and competitive to state-of-the-art NR-IQA methods, according to the experimental results for two typical datasets.

Real-World Blind Face Restoration with Generative Facial Prior [2]

This paper analyzes the motion blur that occurs during the process of image sampling by the camera on Turntable. In order to effectively restore scene blurred image with moving objects which is generated in the process, a restoration algorithm is proposed. Firstly, the moving objects are extracted from the scene blurred images.

Synthetic CelebA-Test [3]

This method uses improvement in white balance to detect the structure in different circumstances. The improved automatic white balance algorithm can reduce the colour distortion and get a clear image with the colour correction of the restored image. According to the contrast experiments of four different underwater images, we can see that the algorithm has some advantages on subjective and objective evaluation indexes, and the sharpness and the colour fidelity of the enhanced image are better.

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 up-sampled images as inputs for face super-resolution.

However, it is challenging to incorporate such generative priors into the restoration process. Previous attempts typically use GAN inversion [19, 54, 52]. They first 'invert' the degraded image back to a latent code of the pretrained 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.

III. SYSTEM ANALYSIS

The architecture here depicts the design and flow of data for the system to predict the subsequent clips of enhanced video. The model is initially trained with CelebA-HQ dataset and as then model receives a new set of images it will collect the new data as a resource for the next set of possibilities. The database will store all the images and which can be later used for training the model again.

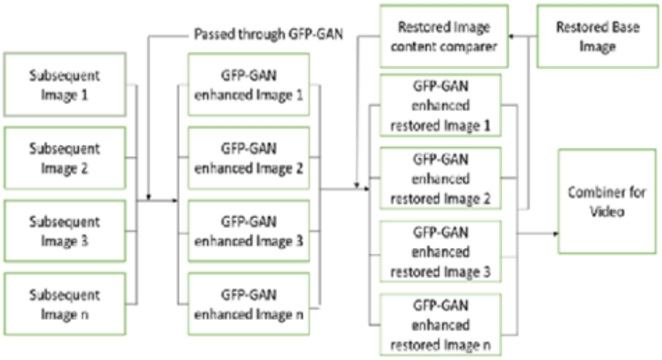

Fig 1 : System Architecture

1) 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 ^y, 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 pertained face GAN (such as Style-GAN2 [36]) as prior. They are bridged by a 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 Flatten to map the input image to the closest latent code in StyleGAN2, and 2) multi-resolution spatial features Spatial for modulating the StyleGAN2 features. After that, flatten is mapped to intermediate latent code 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 pretrained 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 perceptually significant face components, i.e., eyes and mouth. In order to retrain identity, we also employ identity preserving guidance.

IV. RECURRENT NEURAL NETWORK

A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes form a directed or undirected graph along a temporal sequence. This allows it to exhibit temporal dynamic behaviour. Derived from feedforward neural networks, RNNs can use their internal state (memory) to process variable length sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. Recurrent neural networks are theoretically Turing complete and can run arbitrary programs to process arbitrary sequences of inputs. The term "recurrent neural network" is used to refer to the class of networks with an infinite impulse response, whereas "convolutional neural network" refers to the class of finite impulse response. Both classes of networks exhibit temporal dynamic behaviour. A finite impulse recurrent network is a directed acyclic graph that can be unrolled and replaced with a strictly feedforward neural network, while an infinite impulse recurrent network is a directed cyclic graph that cannot be unrolled. Both finite impulse and infinite impulse recurrent networks can have additional stored states, and the storage can be under direct control by the neural network. The storage can also be replaced by another network or graph if that incorporates time delays or has feedback loops. Such controlled states are referred to as gated state or gated memory, and are part of long short-term memory networks (LSTMs) and gated recurrent units. This is also called Feedback Neural Network (FNN). This feedback Neural Network is used to give feedback to the initial layer. This information is known as context of the program and is transferred using RNN. This RNN structure passes this information recurrently. Similar to NLP techniques, this RNN structure modifies the already modified pixels of the video file to ensure that the context of the system is preserved. This preserved information is passed every time to each new frame to make the video's information context clear and relative to the base image.

V. RESULTS

After going through the entire process, we used a video clip of 6 secs and tried to extract it frame by frame. After extraction we got almost 31 frames. These 141 frames are later passed through GFP-GAN Layer and the results are shown accordingly.

The first frame in above results are the frame before GFP-GAN network. There are 31 frames total, and each frame is shown above after clearing. After integrating, a very generalized video could be generated, which is relatively more clear than the earlier one. We later combined all the frames together to form the final clip.

VI. CONCLUSION

Extensive experiments show that our method achieves superior performance to prior art on both synthetic and real-world datasets.

VII. REFERENCES

[1]. Rameen Abdal, Yipeng Qin, and Peter Wonka. Image2stylegan: How to embed images into the stylegan latent space? In ICCV, 2019.

[2]. Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In ECCVW, 2018.

[3]. Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096, 2018.

[4]. Adrian Bulat and Georgios Tzimiropoulos. Super-fan: Integrated facial landmark localization and super-resolution of real-world low-resolution faces in arbitrary poses with gans. In CVPR, 2018.

[5]. Qingxing Cao, Liang Lin, Yukai Shi, Xiaodan Liang, and Guanbin Li. Attention-aware face hallucination via deep reinforcement learning. In CVPR, 2017.

[6]. Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network-based noise modeling. In CVPR, 2018.

[7]. Yunpeng Chen, Jianan Li, Huaxin Xiao, Xiaojie Jin, Shuicheng Yan, and Jiashi Feng. Dual path networks. In NeurIPS, 2017.2

Cite this article as :

Girija V, Sunny Nehra, Himanshu Kumar, Avinash Yadav, Karan R, "Video Regeneration and Quality Enhancer using GFP-GAN", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 3, pp. 148-151, May-June 2022. Available at doi : https://doi.org/10.32628/IJSRSET229344

Journal URL : https://ijsrset.com/IJSRSET229344