Skin Textural Generation via Blue-noise Gabor Filtering based Generative Adversarial Network

Anonymous Author(s)
Submission ID 1365

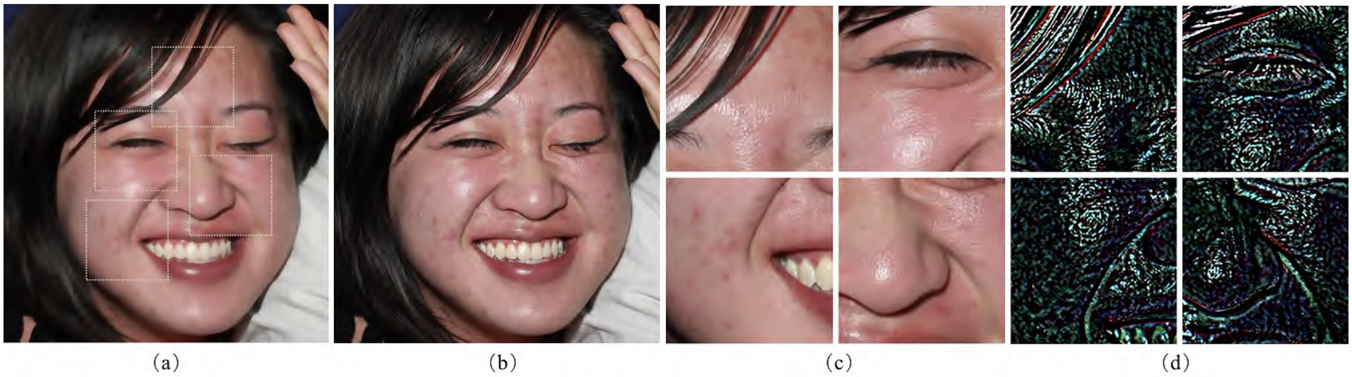

Figure 1: Our skin enhancement result. (a) is the Low-quality face input; (b) is the High-quality generated result, (c) is the texture details of (b) in the white dotted regions of (a); (d) is the learned texture difference of generated high-quality image (c) with the low-quality image.

ABSTRACT

Facial skin texture systhesis is a fundamental problem in high quality facial image generation and enhancement. The key behind is how to effectively synthesize plausible textured noise for the faces. With the development of CNNs and GANs, most works cast the problem as an image to image translation problem. However, these methods lack explicit machenism to simulate the facial noise pattern, so that the generated images are of obvious artifacts. To this end, we propose a new facial noise generation method. Specifically, we utilize the property of blue noise and gabor filter to implicitly guide the asymmetrical sampling for the face region as a guidance map, where non-uniform point sampling is conducted. Thus we propose a novel Blue-Noise Gabor Module so as to produce a spatial-variant noisy image. Our proposed two-branch framework combined facial identity enhancing with textures details generation to jointly produce a high-quality facial image. Experimental results demonstrate the superiority of our method compared with the state-of-the-arts, which enables the generation of high quality facial texture based on a 2D image only, without the involvement of any 3D models.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

skin enhancement, neural networks, Gabor kernel, blue-noise

ACM Reference Format:
Anonymous Author(s). 2018. Skin Textural Generation via Blue-noise Gabor Filtering based Generative Adversarial Network. In *ACM MM '20: ACM Multimedia 2020.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Skin enhancement task aims to regenerate a high-quality face image with rich texture details from a low-quality input face image, which is originated from image restoration and enhancement. Generally speaking, this task is typically ill-posed and very challenging, due to there are always multiple high-quality images corresponding to a single low-quality image.

Most widely studied image restoration and enhancement works are focused on image denoising, demosaicing, and compression artifacts reduction. The degradation in obtaining face images would causes the reduction in face image quality. Image enhancement is basically to improve the interpretability or perception of information in images for human viewers. By modifying the image attributes it could generate the desired result for given specific tasks, which in our case, is to enhance the skin details. The low-level features are more important for image restoration.

Procedural noise functions are widely investigated due to its ability for modeling and creating procedural textures with arbitrarily complex, compared with the difficulties via previously manual

ACM MM '20, June 03–05, 2020, Seattle, WA
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00
https://doi.org/10.1145/1122445.1122456

methods. By enriching the fundamental texture of images, the visual complexity and quality of computer-generated image could be significant improved. Therefore Gabor filters are generated from Gabor function and have been extensively used in computer vision tasks as they show impressive ability to model texture information for images. It is a procedural noise texture that could generate textures with stretched and rotated patterns. In addition, the direct control over the frequency characteristics makes it easier for specifying the feature density and smoothness. Gabor filter responds to the edges and textures of varying frequencies and orientations, and it exhibits a vision similar to human visual perception. However, manually designed Gabor kernels exist the problems of fixed parameters, thus it may not be approaiate for specific texture frequencies, which needs to test many times. Motivated by the learning ability of deep neural networks, we thus Incorporated Gaber kernels into the networks as convolution layer to extract and analysis the texture of image, which could be learned and modified during training.

Inspired by the learning ability of deep generative networks, in this paper, we propose a novel end-to-end trainable generative model for skin enhancement of face images, namely SkinNet. As show in Figure1, our model takes a low-quality face image as input and generates a high-quality face image. It consists of a two-branch sub-network and a discriminator, which target to restore the face identity content and enhance the skin texture details. In the training phise, the blue-noise is adopted to modeling the high frequency noises on face skins. Convolution Gabor kernels are also used extract and analysis the high frequency skin details features. By embedding our Blue-Noise Gabor module into network, our model generate good results and shows a good representation learning ability for textures, which could better restore the datails of Low-quality face images. After training through adviserial learning, our discriminator is able to distingush the generated face images with ground truth high-quality face images.

To summarize, our main contributions are:

- A novel deep generative model, SkinNet is proposed to enhance the detail features of facial skin, which include two-branch, one is Identity content branch, the other one is texture details generation branch.
- A novel Blue-Noise Gabor Module is proposed aiming at the enhancement and synthesis of higher frequencies detail textures, e.g. pores and tiny wrinkle, with different magnitude and orientation.
- A spectral-spatial loss is proposed which enfoces the similarity both of high-frequency details together with low-content features encoded by perceptual network.
- we are the first to apply deep generative methods on the synthesis of details of human face skin, which greatly improve the photo-realistic of the low-quality face image.

2 RELATED WORK

Image Restoration. Image restoration as an important image processing technique for recovering clean images from the corrupted ones has been widely explored in the past decades. Early methods [25, 31, 33] mainly rely on hand-craft features or priors, and may fail on complex cases. Recently, deep learning based network architectures have shown their great success on a variety of image restoration tasks such as image super resolution [10, 19, 22, 32], image deblurring [5, 21], image denoise [26] and etc. A series of network architectures have been designed for extracting features from low quality input and recovering high quality details. Generative Adversarial Network (GAN) [13], as a promising technique for data generation, has been widely studied for image restoration tasks [19, 32]. The merit of GAN is its ability for modeling complex real data distributions and learn to map the randomly sample latent code to the real data via adversarial training. Such property also plays an important role in our architecture for generating fine details.

Face restoration [4, 7, 23] as a special kind of image restoration, has also attracted great attention in the research community. Compared with general image restoration, some prior information such as facial landmark [4] and 3D face model [7] can be utilized for generating face images. However, previous works still have limited ability in generating high frequency facial details. In our work, we use Gabor noise to regress the high frequency noises details of face images and show superior performance in generating facial details.

Image Enhancement. Image Enhancement which aims at adjusting the colors and enhancing the details of the images have been widely explored in the research community recent years. Traditional methods [1, 30] mainly rely on heuristic rules and apply hand-crafted operations for processing the images. Recently, with the rapid development of deep learning technique, various neural network architectures have been designed for image enhancement. [34] is the pioneering work for image adjustment. [6, 11] propose to use neural networks for approximating image filters and achieve considerable improvements, while those methods are limited to learning existing image filters. Like other image generation related tasks, GAN and its variants have also been actively studied in image enhancement tasks. Generally, GAN is used to learn the mapping from the original inputs to the enhanced image domain, and the training data can be both paired [14] or unpaired [8].

In our work, we propose a two-branches skin enhancement network for enhancing local texture details while also keep facial identity content of the input image. Adversarial training strategy is also incorporated to help generate realistic details.

Procedural Noise Functions. Procedural noises have been widely used as a modeling tool for texture synthesis, which has become an essential part in computer graphics applications. Besides the first introduced Perlin noise [27]. Many other procedural noise functions have also been proposed, including the sparse convolution noise [20], [18], Perlin noise [28], wavelet noise [9] and anisotropic noise [12]. Among these noises, sparse convolution noise can be constructed around a specific evaluation functions. One drawback of sparse convolution noise is the generated poor kernels violating the ideal noise requirements. However, Gabor noise [18] solves this problem by using a kernel that is a combination of a Gaussian curve and a cosine curve, which is procedual, could be spectral controled, and support anisotropy. With the development of deep learning techniques, varies networks has been designed for image and texture synthesis. Among which Gabor kernel has also been adopte in the network applications [24], due to the it is sensitive to textures details. In this work, we proposed blue noise Gabor

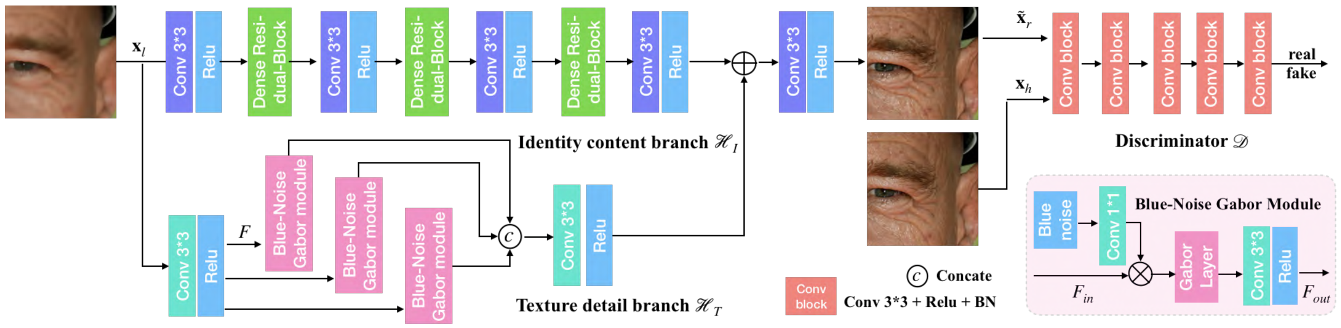

Figure 2: The framework of our two-branch skin enhancement network.

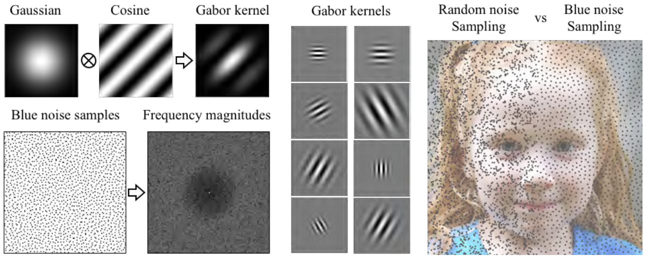

Figure 3: The illustration of Gabor kernel (Up left) and blue noise (Bottom left). The difference between random sampling and blue noise sampling is shown in the right.

module, which takes the advantage of random evenly sampling properties of blue noise and the ability of Gabor kernel to extract textures with different orientation and scales to generate spatially varying textures details.

3 APPROACH

Generic single image restoration and reconstruction aims to learn a mapping function $\tilde{\mathbf{x}} = f(\mathbf{x})$ to estimate a high quality image $\tilde{\mathbf{x}}$ with a given input $\mathbf{x}$. For the skin enhancement task, let's denote $\mathbf{x}_l$ as the low-quality face image with noisy, blurred, or compression artifacts, and denote $\mathbf{x}_h$ as its ground truth high-quality face image. The reconstructed skin image $\tilde{\mathbf{x}}_r$ can be obtained by training through $\tilde{\mathbf{x}}_r = \mathcal{H}_{SkinNet}(\mathbf{x}_l)$, where $\mathcal{H}_{SkinNet}$ is a two-branch skin enhancement network as detailed later. The key to generate high-quality skin textural details behind this network is a Blue-Noise Gabor Module and a novel loss formulated in both spectral and spatial domains, as introduced in Section 3.2 and 3.3 respectively. The former one focuses on regressing the high frequency noisy details, and the latter one is designed to capture the face identity structural information and local skin textural details.

3.1 Two-branch Skin Enhancement Network

Our SkinNet, $\mathcal{H}_{SkinNet}$, follows a deep generative model for enhancing visually pleasing face skin images. Its network structure is illustrated in Figure 2, which is composed of a texture detail generation branch, an identity content enhancement branch, and a discriminator network to jointly restore the fine face skin. We target to disentangle the face image into face identity content and texture details. Specifically, the identity content enhancement branch focuses on shape improvement, e.g. boundary enhancing, deblurring and denoising, while the texture detail generation branch focuses on modeling and generating the skin texture details via Gabor Blue-noises. The feature maps produced by two branches are then fused together to obtain the global fused feature maps,

$$\mathcal{H}_{SkinNet}(\mathbf{x}) = Fuse(\mathcal{H}_{identity}(\mathbf{x}) \oplus \mathcal{H}_{texture}(\mathbf{x})) \qquad (1)$$

where $\mathcal{H}_I(\mathbf{x}) \oplus \mathcal{H}_T(\mathbf{x})$ refers to element-wise addition of the feature maps produced by the identity branch $\mathcal{H}_I(\mathbf{x})$ and the texture branch $\mathcal{H}_T(\mathbf{x})$. $Fuse(\cdot)$ is composed of 1×1 and 3×3 convolution layers, and is designed for fusing features from different levels. We also incorporate a discriminator to distinguish the fake skin detail enhanced images from the real ones, which is demonstrated useful in generating realistic details.

3.1.1 Identity Content Enhancement Branch. We define the Identity structure as the part of image which contains the most energy region of the spectrum in spectral domain. In low quality image, the face shape is better preserved compared with the detailed skin texture. Our Identity content Enhance Branch focus on the improving of these structural identities.

The identity branch takes a low-quality image $\mathbf{x}_l$ as input and extract feature maps F_I via a 3×3 convolution layer followed by a ReLU layer. The feature maps F_I then go through several sequential Dense Residual blocks [35], followed by another 3×3 convolution layer with a ReLU layer, and added with the input $\mathbf{x}_l$ to reconstruct the Identity face skin. The dense residual blocks consist of dense connected layers and local feature fusion with local residual learning, which could effectively extract abundant local features. The connected layers are formed by 3×3 convolution layer with ReLU activation. Each connected layer has direct connection to all subsequent connected layers.

3.1.2 Texture Detail Generation Branch. The texture detail generation branch is designed for modeling the high frequency features on skin surface. It takes as input a low-quality face image and output feature maps containing enhanced high frequency detail information from multiple levels. Specifically, it first extracts the feature maps with a 3×3 convolution layer followed by a ReLU layer, and the feature maps are then passed through several parallel Blue-noise Gabor modules with different Gabor kernel size, scale magnitude, frequency, bandwidth and orientation for enhancing high frequency detail information. Finally, the enhanced feature maps with different Blue-noise Gabor modules are then added together and fused by a 3 × 3 convolution layer with a Relu layer to modeling the textures and generating feature maps with high frequency details. The Gabor modules with different parameters in this branch represent different high frequency patterns, which makes the network being capable of modeling different types of low-level skin texture details for high-quality image restoration tasks, resulting better performance. The details of our proposed Blue-noise Gabor Module is introduced later in Section 3.2.

3.1.3 Global-local Residual Learning. With the usage of global and local residual learning in spatial and spectral space, our network could concentrate on learning the degradation components including blurring, noisy, or compressed artifacts. Our skin enhancement is similar with image-to-image translation task where the input image is highly correlated with the target image. The network learns the residuals between input and output, namely global residual learning. The skin enhancement avoids learning a complicated transformation from one image to another completely different image, instead it only requires learning a residual map to restore the missing high-frequency skin details. The local residual learning among the Blue-noise Gabor Module and dense residual blocks are similar to the residual learning in ResNet. Since the ever-increasing network depths will cause the degradation problem, the local residual learning is used to alleviate this problem and reduce training difficulty and improve the learning ability.

3.2 Blue-noise Gabor Module

Face image with low-quality usually contains more low-frequency content, which lacks the high frequency texture details. These textures could be regarded as a specific type of face noise distribution. In other words, the key to reconstruct a high-quality face image is to effectively model and generate such a distribution. To this end, we propose a novel Blue-noise Gabor Module, which is designed for adding frequency details to the feature maps via blue noise-sampling convolution with Gabor kernels, so as to be vital for generating high quality realistic facial skin details.

3.2.1 Gabor Noise. Gabor noise [18] has a power spectrum which could be used as a basis to approximate arbitrary power spectra of images. It is defined as a sum of weighted and randomly positioned Gabor kernels $\mathcal{G}$,

$$\mathcal{N}_{K,F_0,a,\omega_0}(x,y) = \sum_i \omega_i \mathcal{G}_{K,F_0,a,\omega_0}(x - x_i, y - y_i) \tag{2}$$

where ω_i denotes the weights, K, F, a, ω_0 and (x_i, y_i) denote the magnitude, frequency, bandwidth, orientation and position of the Gabor kernel respectively. The Gabor kernel $\mathcal{G}$ can be regarded as the product of a Gaussian kernel and a 2D sinusoidal function, which is defined as:

$$\mathcal{G}_{K,F_0,a,\omega_0}(x,y) = Ke^{-\pi a^2(x^2+y^2)} \cos\left[2\pi F_0(x_0 \cos\omega_0 + y_0 \sin\omega_0)\right] \tag{3}$$

where K and a represent the magnitude and the width of the Gaussian, and F_0 is the frequency of the cosine and ω_0 controls the orientation. Our proposed module consists of the convolution of blue-noise with the Gabor kernel dipicted as,

$$\mathcal{N}_{K,F_0,a,\omega_0}(x,y) = \left[\sum_i \omega_i \delta_{\{x_i,y_i\}} \otimes \mathcal{G}_{K,F_0,a,\omega_0}\right](x,y) \tag{4}$$

where the weights ω_i are learned during network training. Positions (x_i, y_i) are distributed according to a specific point distribution δ.

3.2.2 Blue Noise Sampling. Blue noise as a kind of point distribution contains higher amounts of high-frequencies and lower amounts of low-frequencies. It has a more even distribution compared with the random noise sampling as seen in Figure 3. For this reason, we choose it as the point distribution to facilitate enhancing the local high frequency skin texture details of the face images. Specifically, blue noise is a set $X = \{\mathbf{x}_i \in D; i = 1, 2, ..., N\}$ of N samples generated from the following random process,

$$\forall \mathbf{x}_i \in X, \forall S \subseteq D : P(\mathbf{x}_i \in S) = \int_S \mathbf{dx}; \forall \mathbf{x}_i, \mathbf{x}_j \in X : \left\|\mathbf{x}_i - \mathbf{x}_j\right\| \geq 2r \tag{5}$$

where r is the distribution radius enforcing the minimum distance constraint between any pair of sample points. A uniformly distributed random sample x_i of X has a probability of falling inside a subset S of D. $\int_D \mathbf{dx} = 1$ denotes for unit space. This is a process that distributes uniform random samples on a domain space based on a minimum distance criterion between samples. Its Fourier spectrum shown in Figure 3 exhibits the blue-noise property, which has low anisotropy and small amount of low frequency energy. Thus, the blue-noise samples are evenly located but still remain at least a minimum distance r apart from one another.

3.2.3 Module Structure. As shown in Figure 2 (bottom right), the proposed Blue-Noise Gabor module takes the feature maps F extracted from low-quality image $\mathbf{x}_l$ as input, and produces a set of feature maps containing enhanced high frequency detail information. Specifically, we randomly generate the blue noise and weight it via a 1 × 1 convolution layer. The weighted Blue-noise is then added to the input feature maps F. The added feature maps are further fed into a Gabor layer followed by a 3 × 3 convolution layer with ReLU as activation for enhancing high frequency detail information. The Gabor layer is implemented as a convolution layer initialized with Gabor kernels with different parameters K, F_0, a, and ω_0, which will be updated continuously during training. The design choice of the Blue-noise and the Gabor layer endows the network with the capability of capturing different visual details and encoding different frequency, spatial localization and orientation information into the feature maps, which is essential to produce the high quality skin texture details on the face images. More importantly, the produced skin texture details are stationary by chosen blue-noise as distribution, which is translation-invariant and has no location bias.

3.3 Spectral-Spatial Skin Enhancement Loss

We design a customized loss to train the aforementioned network in an adversarial manner. The loss is composed of a spectral and a spatial loss, which enable the network to learn the statistics of face images and reconstruct photo realistic face skin with high-quality detail textures.

3.3.1 *Spatial Loss for Visual Identity.*

The spatial loss is designed to force the reconstructed face image to have clearer structure compared with the low-quality input image. We define the spatial similarity loss as the sum of pixel-wise $L1$ loss and the feature-level perceptual loss[15],

$$\mathcal{L}_s = \lambda_1 \sum_{i \in \mathcal{I}} \|\tilde{\mathbf{x}}_r - \mathbf{x}_h\|_1^2 + \sum_{j \in \phi} \|\phi_j(\tilde{\mathbf{x}}_r) - \phi_j(\mathbf{x}_h)\|_2^2, \qquad (6)$$

where $\phi(\tilde{\mathbf{x}}_r)$ and $\phi(\mathbf{x}_h)$ correspond to the feature maps extracted from the reconstructed image $\tilde{\mathbf{x}}_r$ and the ground truth image $\mathbf{x}_h$ respectively with a pretrained VGG-16 network [29], subscripts i is the pixel in $\mathcal{I}$; subscripts j in ϕ_j indicate pixel j in ϕ-th layer of the VGG feature maps, and λ_1 refers to the balancing coefficient.

Such design choice simultaneously takes pixel value, sementic feature and high level structure information into consideration, and encourages the reconstructed images to have the same statistics with the high-quality ground truth images.

3.3.2 *Spectral Loss for Frequency Details.*

Real-valued periodic period image signals can be expressed as the sum of sinusoidal oscillations of various frequencies, magnitudes and phase shifts [2]. Fourier transform, which can be used to transfer the signals into spectral domain have been widely used in image analysis. It could solve the difficulty to recognise noises in spatial domain by transforming it into spectral domain. Our proposed spectral loss $\mathcal{L}_f$ is designed to calculate the Fourier coefficients and enforce the similarity of reconstructed image with ground truth in the spectral domain. The spectral loss $\mathcal{L}_f$ is a weighted version of l_2 loss of Foriour transformed images, defined as,

$$\mathcal{L}_f = \sum_{i \in \mathcal{I}_f} \|(\mathcal{F} \otimes \tilde{\mathbf{x}}_r)_i - (\mathcal{F} \otimes \mathbf{x}_h)_i\|_2^2 \qquad (7)$$

where $\mathbf{x}_h$ is the high-quality ground truth image, $\tilde{\mathbf{x}}_r$ is the reconstructed image. $\mathcal{F}\otimes$ denotes the Foriour Transform. i is the pixel from the transformed image $\mathcal{I}_f$ in spectral domain. The discrete Fourier transform $\mathcal{F}\otimes$ on image $\mathcal{I}$ is defined as,

$$\mathcal{F}(\mathcal{I}_f) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(\mathcal{I}) e^{-j2\pi(\frac{ui}{M} + \frac{vj}{N})}, (u,v) \in \mathcal{I}_f, (i,j) \in \mathcal{I} \qquad (8)$$

where $f(\mathcal{I})$ denotes a $M \times N$ image in spatial domain, $i = 0, 1, ..., M-1$, $j = 0, 1, ..., N-1$. $\mathcal{F}(\mathcal{I}_f)$ denotes the Fourier transform of image $\mathcal{I}$, where $\mathcal{I}_f$ has the same size with $\mathcal{I}$.

The spectral loss used in our pipeline enforces the reconstructed images to maintain similar frequency spectrum with the ground truth images, and lead to high quality skin texture details.

3.3.3 *Total Energy Function .*

Our model is trained through adversarial learning. The two-branch network generates the reconstructed face images, where the discriminator distinguishes the reconstructed images from the ground truth. The adversarial loss is defined as below:

$$\mathcal{L}_{adv} = \min_G \max_{\mathcal{D}} \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x})} \left[log(\mathcal{D}(\mathbf{x}) \right] + \mathbb{E}_{\mathbf{x} \sim p(\mathbf{x})} \left[log(1 - \mathcal{D}(G(\mathbf{x}))) \right]$$

where $\mathbf{x}$ denotes the training low-quality image samples, the generator G and discriminator $\mathcal{D}$ are trained alternately by solving the mini-max optimization problem.

We incorporate three additional loss terms into the adversarial training process $\mathcal{L}_{adv}$, i.e. the spatial identity loss ($\mathcal{L}_s$), the spectral frequency loss ($\mathcal{L}_f$) and the regularized term $\mathcal{L}_{tv}$, The loss function is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{adv} + \lambda_s \mathcal{L}_s + \lambda_f \mathcal{L}_f + \lambda_{tv} \mathcal{L}_{tv}, \qquad (9)$$

where λ_s, λ_f and λ_{tv} are balancing weights. The total variation (TV) regularizer is used to encourage the spatial smoothness of the enhanced face skin image.

4 EXPERIMENTAL RESULTS

4.1 Datasets and Implementation details

We conduct extensive experiments on two datasets: CelebA-HQ [16] and Flickr-Faces-HQ (FFHQ) [17]. The image resolution of both CelebA-HQ and FFHQ is 1024 × 1024. CelebA-HQ consists of 30,000 high-quality images, while FFHQ consists of 70,000. The FFHQ dataset includes vastly more variation than CelebA-HQ in terms of age, ethnicity.

The degradation process could be affected by various factors including sensor and speckle noise, compression artifacts, anisotropic degradations, which process is unknown makes it difficult to model the degradation mapping. We pre-pocess the degradation to original high-quality images to get low-quality images as paired data to bulid our training datasets. In the training, we crop the training images in the face regions with size to 256 × 256 without other pre-alignment operation. Both input and output images are of size 256 × 256. For testing, we send the 1024 × 1024 image as the input. The Gabor kernel in our proposed module are implemented as the basic convolution filter layers. Poisson-disk sampling [3] is used here to implement blue-noise. In the training phase, the blue-noise is adopted to model the high frequency noises on face skins. Convolution Gabor kernels are also used to extract and analysis the high frequency skin details features.

4.2 Quantitative and Qualitative comparisons

In order to evaluate the performance of our proposed network, we compare with the state-of-the-art methods, including SRCNN [] EDSR [22], SRGAN [19], ESRGAN [32] qualitatively and quantitatively. In the skin enhancement task, We compare our method with other state-of-the-art methods by evaluating the performance of all the methods quantitatively on the entire test dataset with the settings as described before. Average PSNR and the structural similarity (SSIM) scores are used as the evaluation metrics. The comparisons between our model and other state-of-the-art methods are presented in Table 1.

Table 1 provides quantitative comparisons with state-of-the-art super-resolution techniques for skin enhancement. It indicates that our method achieves superior performance compared to other methods, i.e., outperforming the second best with a large margin of 5.63 dB in PSNR and 0.075 in SSIM. A qualitative comparison of face skin enhancement between our method and other state-of-the-art methods are shown in Figure 4. Obviously, EDSR fails to generate authentic facial details and the boundary of faces are still blurry. SRGAN has got better results than EDSR, however it still could not generate the tiny pores or wrinkles. ESRGAN can produce better results, however, it has ringing artifacts around facial

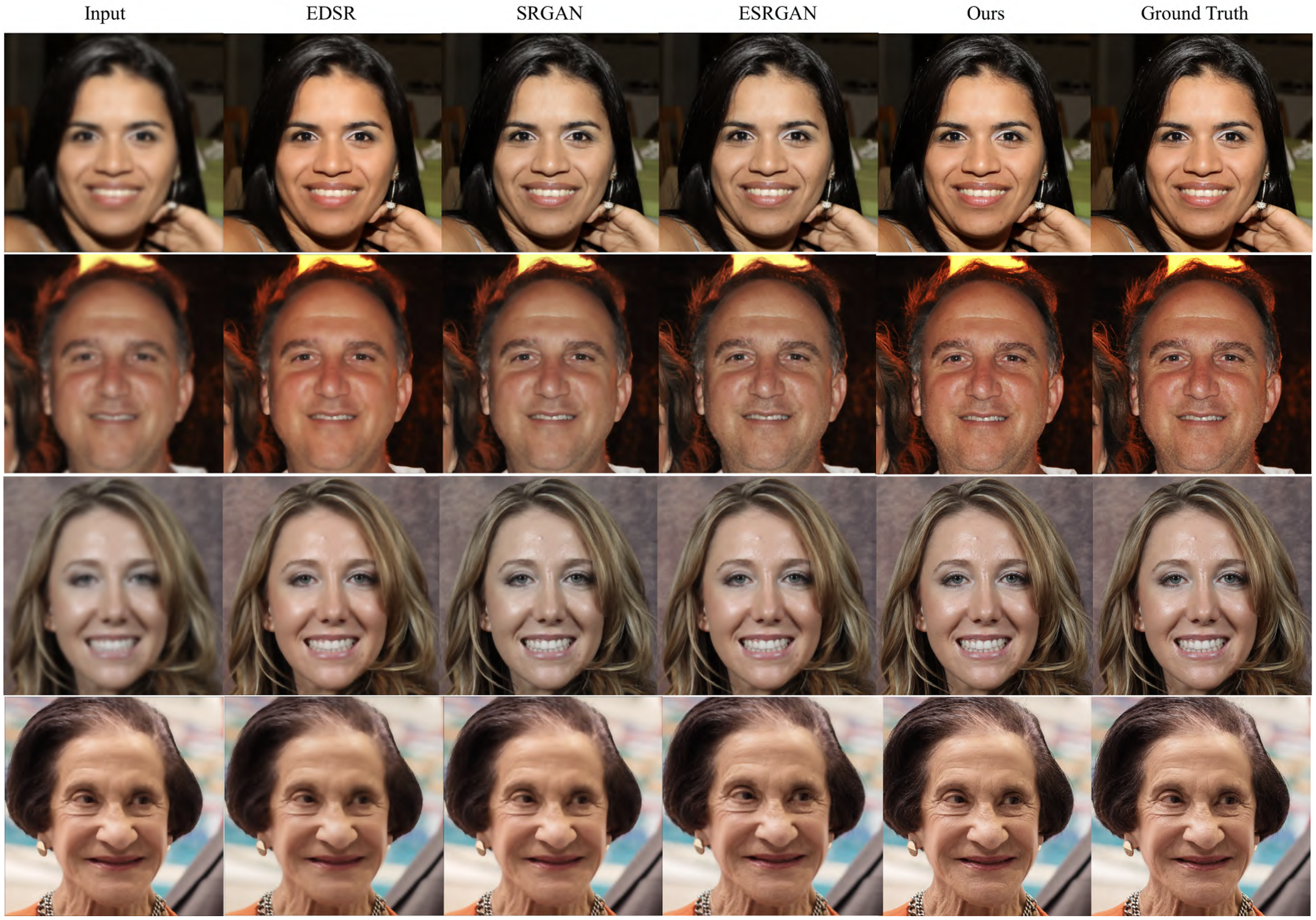

Figure 4: Quantitative comparisons on the state-of-the-art methods on skin enhancement task. (The resolution of image is 1024×1024, please zoom in for better visual quality.)

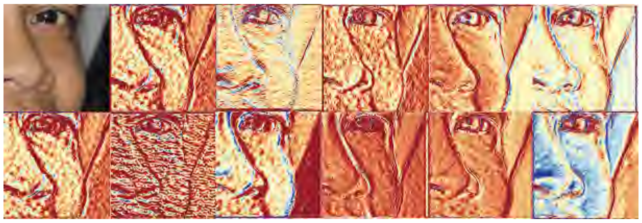

Figure 5: Features learned by Blue-noise Gabor Module. Left up coner is the low-quality input face image.

components, including eyes, forehead and around mouth, which still lacks detail high frequencies features. The skin details of our results tend to be more appealing and clearer than those from other methods especially on the forehead, check, and wrinkles around eye, our method generate texture details more accurately.

Table 1: Quantitative comparisons of PSNR and SSIM on the state-of-the-art methods on skin enhancement task.

Methods	SRCNN	FSRCNN	EDSR	SRGAN	ESRGAN	Ours
PSNR	24.40	26.11	27.97	25.38	28.69	**34.32**
SSIM	0.704	0.795	0.737	0.612	0.803	**0.878**

Table 2: Ablation study of our proposed network with different setting. (*$Setting_1$* is for studying the effect of different types of noise used in the Module; *$Setting_2$* is for different loss.)

$Setting_1$	without noise	Random-noise	Blue-noise
PSNR	29.89	30.6	**34.32**
SSIM	0.794	0.802	**0.878**
$Setting_2$	$\mathcal{L}_{adv}$, $\mathcal{L}_s$	$\mathcal{L}_{adv}$, $\mathcal{L}_f$	$\mathcal{L}_{adv}$, $\mathcal{L}_s$, $\mathcal{L}_f$
PSNR	30.47	29.45	**34.32**
SSIM	0.810	0.792	**0.878**

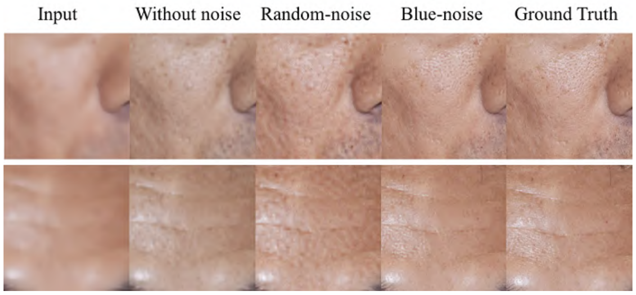

Figure 6: Ablation study on noise in Blue-Noise Gabor Module. (From left to right: Input; results without noise; random noise; blue noise; GT. Zoom in for better view of details.)

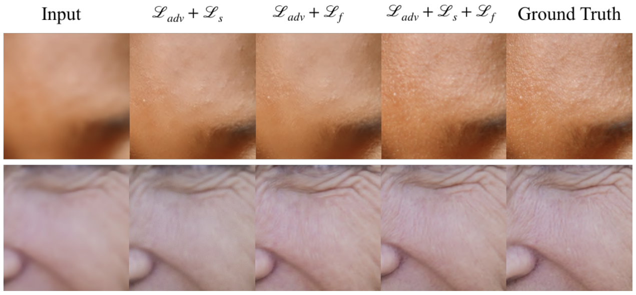
Input
$\mathscr{L}_{adv}+\mathscr{L}_s$
$\mathscr{L}_{adv}+\mathscr{L}_f$
$\mathscr{L}_{adv}+\mathscr{L}_s+\mathscr{L}_f$
Ground Truth

Figure 7: Ablation study on the loss function. (From left to right: input; results with $\mathcal{L}_{adv}$ + $\mathcal{L}_s$; $\mathcal{L}_{adv}$+$\mathcal{L}_f$; $\mathcal{L}_{adv}$+ $\mathcal{L}_s$+ $\mathcal{L}_f$; GT.)

4.3 Ablation Studies and Discussion

In this section, we conduct ablation study on the proposed Blue-Noise Gabor Module and spectral-spatial loss function.

4.3.1 Effectiveness of Blue-Noise Gabor Module. We conduct ablation experiments about the proposed Blue-Noise Gabor Module here. Table 2 illustrates the quantitative results about the ablation study of different types of noise used in the Blue-noise Gabor module. The module using Blue-noise shows superior performance compared with that using Random-noise or without using noise, i.e., outperforming the second best with a large margin of 3.72 dB in PSNR and 0.076 in SSIM. Figure 6 demonstrates some qualitative results,it shows that the network produces blurry results without using any noises, this is because the network lacks the ability of modeling the high frequency skin texture details with the absent of the noise. The results with random noise used still has less details compared with the result using Blue-noise, due to the fact that random noise may has clusters at some local regions, and is less effective in modeling the texture details.

Figure 5 demonstrates some feature maps learned by our Blue-noise Gabor Module, it shows that our Gabor layer learns the textures of skin surface with different values denoted by different colors, scales and orientations. It could even extract small dotted and thin wrinkles of skin surface, which shows the detail textures with different frequency and magnitude could be learned from our proposed blue-noise Gabor module.

4.3.2 Effectiveness of Loss Function. We also conduct ablation experiments about the proposed spectral-spatial loss. Table 2 illustrates the influences of different losses on the performance of generating high quality face images. It shows that only using the spectral loss $\mathcal{L}_f$ or the spatial loss $\mathcal{L}_s$ will lead to much lower performance on PSNR and SSIM compared with our spectral-spatial loss. Figure 7 demonstrates some qualitative results. Obviously, the results produced by using full set of losses looks much clearer and have more realistic skin texture details (e.g. pores and tiny wrinkle, with different magnitude and orientation), compared with the results produced by using only spectral or spatial loss which seem to be blurry and have unrealistic artifacts.

The pixel-wise L1 loss does not take image quality into account, which results lack high-frequency details and are perceptually unsatisfying with over smooth textures. Thus the feature-level perceptual loss improves the visual quality, and the spectral loss is target to maintain the high frequency textures, together with the discriminate loss makes the faces sharper and more realistic. Therefore, by both using $\mathcal{L}_s$ and $\mathcal{L}_f$, our model creates much more realistic textures and produces visually more satisfactory results.

5 CONCLUSION

In this paper, we propose a Facial skin enhancement method. We design a novel deep generative model, SkinNet to enhance the detail features of facial skin, which contains and identity content enhancement branch and a texture details generation branch. A novel Blue-Noise Gabor Module is proposed aiming at the enhancement and synthesis of higher frequencies detail textures, e.g. pores and tiny wrinkle, with different magnitute and orientation of gaber kernel and could be learned during training. Our spectral-spatial loss is proposed which enforces the similarity of both high-frequency details together with low-frequency content features encoded by perceptual network. Extensive experimental results have shown that our method greatly can generate high-quality realistic skin texture details from the low-quality input face image and outperform the state-of-the-arts methods.

REFERENCES

[1] Mathieu Aubry, Sylvain Paris, Samuel W Hasinoff, Jan Kautz, and Frédo Durand. 2014. Fast local laplacian filters: Theory and applications. *ACM Transactions on Graphics (TOG)* 33, 5 (2014), 1–14.

[2] Jean Baptiste Joseph baron Fourier. 1822. *Théorie analytique de la chaleur.* F. Didot.

[3] Robert Bridson. 2007. Fast Poisson disk sampling in arbitrary dimensions. *SIGGRAPH sketches* 10 (2007), 1278780–1278807.

[4] Adrian Bulat and Georgios Tzimiropoulos. 2018. Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 109–117.

[5] Liang Chen, Faming Fang, Tingting Wang, and Guixu Zhang. 2019. Blind image deblurring with local maximum gradient prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1742–1750.

[6] Qifeng Chen, Jia Xu, and Vladlen Koltun. 2017. Fast image processing with fully-convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision.* 2497–2506.

[7] Yu Chen, Ying Tai, Xiaoming Liu, Chunhua Shen, and Jian Yang. 2018. Fsrnet: End-to-end learning face super-resolution with facial priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2492–2501.

[8] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. 2018. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 6306–6314.

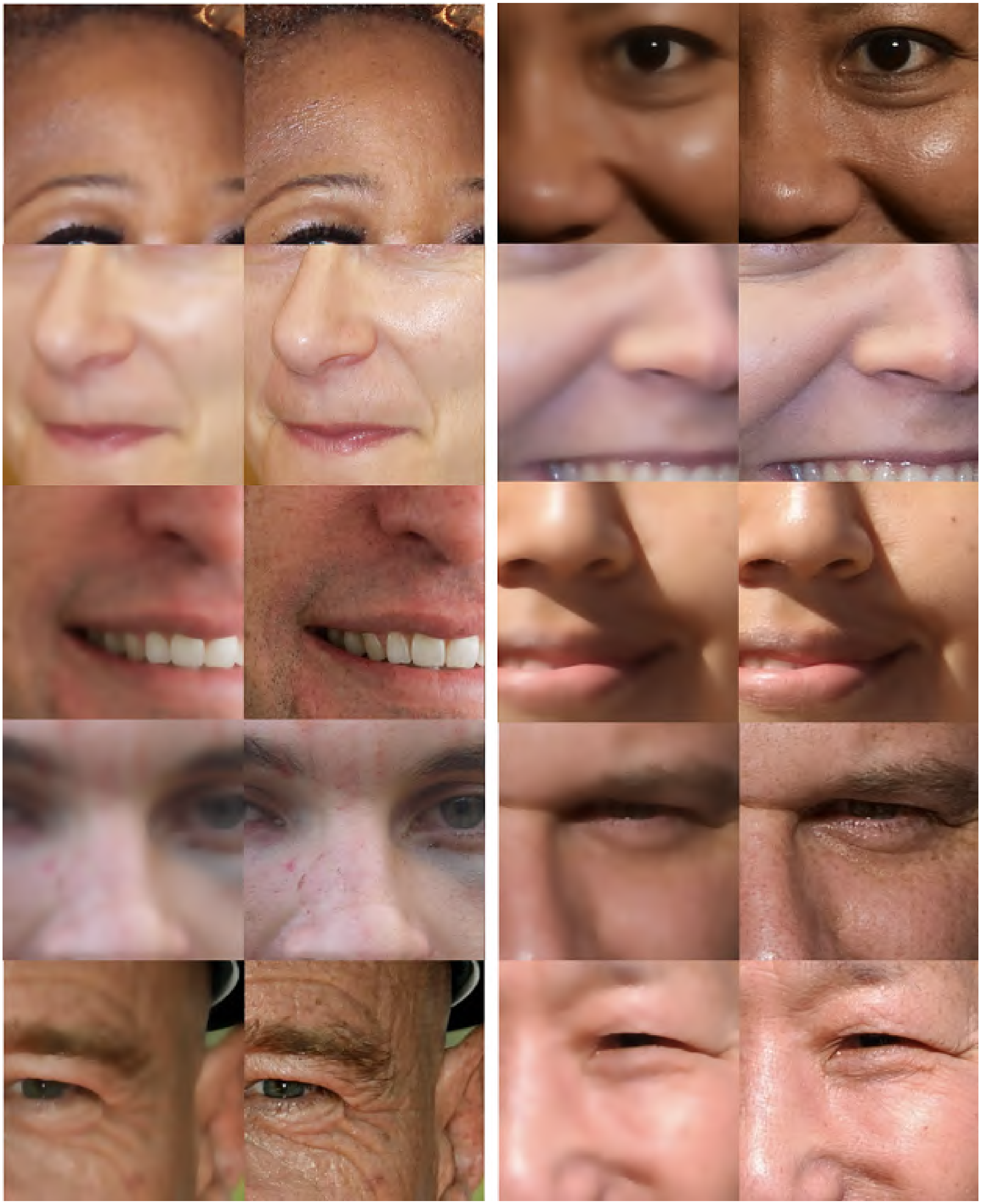

Figure 8: Skin enhance results generated from our model. For each result, left image is the input low-quality image; right is the result after skin enhance.

[9] Robert L Cook and Tony DeRose. 2005. Wavelet noise. *ACM Transactions on Graphics (TOG)* 24, 3 (2005), 803–811.

[10] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2014. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*. Springer, 184–199.

[11] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. 2017. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1–12.

[12] Alexander Goldberg, Matthias Zwicker, and Frédo Durand. 2008. Anisotropic noise. *ACM Transactions on Graphics (TOG)* 27, 3 (2008), 1–8.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

[14] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2017. DSLR-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*. 3277–3285.

[15] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*. Springer, 694–711.

[16] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. 2017. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196* (2017).

[17] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4401–4410.

[18] Ares Lagae, Sylvain Lefebvre, George Drettakis, and Philip Dutré. 2009. Procedural noise using sparse Gabor convolution. *ACM Transactions on Graphics (TOG)* 28, 3 (2009), 1–10.

[19] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4681–4690.

[20] John-Peter Lewis. 1989. Algorithms for solid noise synthesis. In *Proceedings of the 16th annual conference on Computer graphics and interactive techniques*. 263–270.

[21] Lerenhan Li, Jinshan Pan, Wei-Sheng Lai, Changxin Gao, Nong Sang, and Ming-Hsuan Yang. 2018. Learning a discriminative prior for blind image deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6616–6625.

[22] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. 2017. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 136–144.

[23] Jianxin Lin, Tiankuang Zhou, and Zhibo Chen. 2018. Multi-scale face restoration with sequential gating ensemble network. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[24] Shangzhen Luan, Chen Chen, Baochang Zhang, Jungong Han, and Jianzhuang Liu. 2018. Gabor convolutional networks. *IEEE Transactions on Image Processing* 27, 9 (2018), 4357–4366.

[25] Xiang Ma, Junping Zhang, and Chun Qi. 2010. Hallucinating face by position-patch. *Pattern Recognition* 43, 6 (2010), 2224–2236.

[26] Nazeer Muhammad, Nargis Bibi, Adnan Jahangir, and Zahid Mahmood. 2018. Image denoising with norm weighted fusion estimators. *Pattern Analysis and Applications* 21, 4 (2018), 1013–1022.

[27] Darwyn R Peachey. 1985. Solid texturing of complex surfaces. In *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*. 279–286.

[28] Ken Perlin. 2002. Improving noise. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*. 681–682.

[29] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[30] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing* 22, 9 (2013), 3538–3548.

[31] Xiaogang Wang and Xiaoou Tang. 2005. Hallucinating face by eigentransformation. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 35, 3 (2005), 425–434.

[32] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2018. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 0–0.

[33] Zhongyuan Wang, Ruimin Hu, Shizheng Wang, and Junjun Jiang. 2013. Face hallucination via weighted adaptive sparse regularization. *IEEE Transactions on Circuits and Systems for video Technology* 24, 5 (2013), 802–813.

[34] Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu. 2016. Automatic photo adjustment using deep neural networks. *ACM Transactions on Graphics (TOG)* 35, 2 (2016), 1–15.

[35] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. 2018. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2472–2481.