Legacy Photo Editing with Learned Noise Prior

Yuzhi Zhao[1(✉)] Lai-Man Po[1] Tingyu Lin[1] Xuehui Wang[2] Kangcheng Liu[3]
Yujia Zhang[1] Wing-Yin Yu[1] Pengfei Xian[1] Jingjing Xiong[1]
yzzhao2-c@my.cityu.edu.hk
[1]City University of Hong Kong [2]Sun Yat-sen University [3]The Chinese University of Hong Kong

Abstract

There are quite a number of photographs captured under undesirable conditions in the last century. Thus, they are often noisy, regionally incomplete, and grayscale formatted. Conventional approaches mainly focus on one point so that those restoration results are not perceptually sharp or clean enough. To solve these problems, we propose a noise prior learner NEGAN to simulate the noise distribution of real legacy photos using unpaired images. It mainly focuses on matching high-frequency parts of noisy images through discrete wavelet transform (DWT) since they include most of noise statistics. We also create a large legacy photo dataset for learning noise prior. Using learned noise prior, we can easily build valid training pairs by degrading clean images. Then, we propose an IEGAN framework performing image editing including joint denoising, inpainting and colorization based on the estimated noise prior. We evaluate the proposed system and compare it with state-of-the-art image enhancement methods. The experimental results demonstrate that it achieves the best perceptual quality. Please see the webpage https://github.com/zhaoyuzhi/Legacy-Photo-Editing-with-Learned-Noise-Prior for the codes and the proposed LP dataset.

1. Introduction

Restricted by the imaging technology, it remains incomplete parts and noise in legacy grayscale photos. It is highly challenging to restore them due to the great information loss of real world. Also, there is high demand for high-quality and colorful legacy photos. Recently, as deep learning techniques have been demonstrated to successfully applied to many low-level computer vision tasks, the legacy photo enhancement becomes possible. In this paper, we would first discover the representation of blind noise from legacy images as a prior, and then perform image editing based on the estimated noise prior.

Editing legacy photos is highly challenging since there are multiple degradation types in legacy photos. Firstly, there exist noises with unknown distribution and intensity. The noises may be caused by many reasons such as sensor noise, camera distortion, jpeg compression, preservation technology, etc. However, most of current denoisers [53, 41, 54, 29, 13, 30] are trained with specific noise models such as Gaussian and Poisson distribution. Directly applying those denoisers to legacy photos cannot well enhance the images [1]. Secondly, it remains flaws or cracks in legacy photos, which are not global noise but regional artifacts. Moreover, the levels of the artifacts are different for distinct pixels, which are hard to estimate. Finally, the grayscale legacy photos lack of color. Thus, the colorization process is significant to attach vivid colors to them. In conclusion, the pipeline of legacy photo editing can be categorized into three parts: denoising, inpainting, and colorization.

To address the issues, we propose a system to implement the pipeline sequentially. Firstly for denoising, the noise distribution of legacy photos is always unknown. However, the current denoisers pre-define a fixed noise model. It is not practical to directly apply the denoisers to process legacy photos with blind noise. If denoised images are not clean enough, the following inpainting and colorization will also be affected. Moreover, there are no pairs of degraded and clean target legacy photos (i.e. legacy photos are normally noisy). There may be three approaches to address the issue such as estimating noise model [2, 47], unsupervised training [26, 3] and learning blind noise distribution [6, 52]. Since the camera settings are unknown (i.e. the ISP of old cameras is extremely hard to acquire), and the unsupervised training methods also assume a noise distribution, we alternatively propose to learn the noise prior on unpaired legacy photos and clean images by a NEGAN.

Based on the CycleGAN framework [12, 59], we proposed the NEGAN to estimate the blind noise model. Firstly, we notice that the noisy regions normally include more high-frequency components than common regions; whereas flatten (or noise-free) areas comprise the low-frequency components. Thus, we utilize discrete wavelet

Figure 1. The edited real legacy photo samples by proposed method (chosen from LP dataset, captured around 1950). The first, second and third row denote the real legacy photos, image enhancement results by [13, 51, 55] sequentially and the proposed pipeline, respectively.

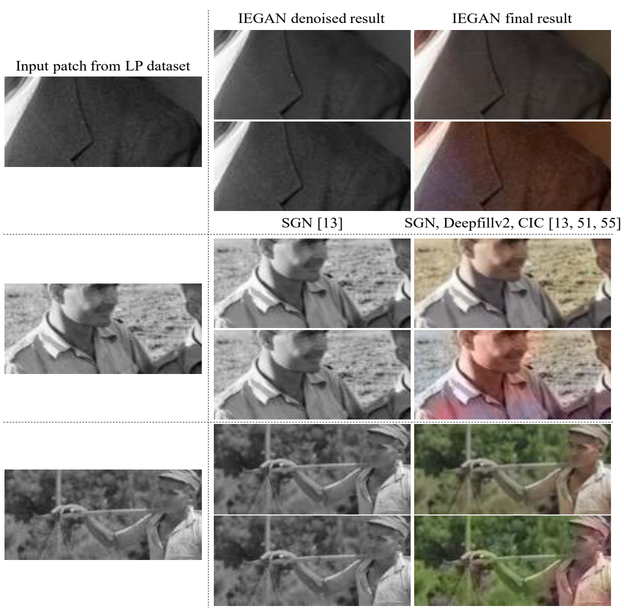

Figure 2. The details of edited real legacy photo by 3 samples. The left part includes the input patches. The right part includes denoised result, final result by the proposed system, denoised result by [13], and previous pipeline [13, 51, 55], respectively.

transform (DWT) to extract the high-frequency components of generated images and real noisy photos, which are used for computing the domain adversarial loss. Secondly, we randomly select patches rather than resizing whole images, in order to maintain the low-level statistics. In addition, we collect a legacy photo dataset (LP dataset), which contains more than 25000 old photos with different levels of noises.

If the NEGAN is well trained, we can obtain training pairs by manually degrading the clean images from a large-scale dataset, such as ImageNet [39]. The degraded images have similar statistics with legacy photos. Thus, the following inpainting and colorization processes are based on paired data. The solutions can be briefly categorized into reference-based and automatic. To improve the quality of generated images, we propose to perform the image editing by an IEGAN, i.e. reference-based inpainting and colorization. For the inpainting, it is hard to annotate the real cracks on each legacy photo. Thus, we alternatively collect some templates for modelling the cracks. By multiplying the cracks and clean images, we can obtain the masked images. While for colorization, we use the color scribbles as additional input guidance to enhance colorization reality. The adversarial losses used in IEGAN aim to improve perceptually quality of generated images.

We evaluate the proposed pipeline on ImageNet [39] validation set. Compared with previous pipelines (i.e. denoising with AWGN, inpainting, and colorization networks), the proposed system achieves the best perceptual performance. Also, we visualize some samples in LP dataset in Figure 2 (resolution 1760×1760) and details in Figure 2 (resolution 256×512). Since the images in LP dataset generally have little cracks, we manually add the masks to real legacy photos for better visualization.

The main contributions of this paper are as follows:

1) We propose a novel NEGAN for estimating blind noise of legacy photos using unpaired data;

2) We create a new legacy photo dataset (LP dataset) including different types of degradation of real legacy photos for learning noise prior;

3) We propose an IEGAN that jointly performs denoising, inpainting and colorization in a user-guided way based on the noise prior estimated by NEGAN.

2. Related Work

Image Denoising. Image denoising is a fundamental problem in low-level vision. Recently, researches have shown that deep learning technologies outperform traditional methods such as bilateral filtering, BM3D [9], non-local algorithm [4]. Mao et al. [33] designed U-Net shaped network to perform image denoising, which was improved by DnCNN [53] using residual learning and MemNet [41] using long memory. Using a tunable noise level map as the input, FFDNet [54] handled a wide range of noise levels and removed spatially variant noise. Considering both Gaussian-Poisson Model and in-camera processing pipeline, CBDNet [14] further improved the blind denoising ability by embedding a noise estimation network. To further improve the network architecture, MWCNN [29] utilized DWT to avoid down-sampling information loss. SGN [13] greatly decreased the memory consumption and runtime, while it was further improved by DSWN [30] using residual path and reconstruction path.

Image Inpainting. The image inpainting denotes the process of filling cracks of images. Normally, the masks of corresponding masked images are known. Pathak et al. [36] firstly adopted a conditional GAN [12] for context completion. It was enhanced by jointly utilizing global and local discriminators by Iizuka et al. [19] to strengthen sharpness for filled regions. Liu et al. [28] introduced a partial convolution with automatically updated status to deal with irregular input masks. It was improved by gated convolution [51]. It is the combination of vanilla convolution and gate state, which generalizes the partial convolution by a learnable dynamic feature selection mechanism. The EdgeConnect [35] proposed an edge generator and image completion network to minimize blurry effect. Xiong et al. [49] further enhanced it for foreground-aware image inpainting.

Image Colorization. The existing colorization methods can be briefly categorized into three classes: scribble-based [27, 50, 7, 56], example-based [20, 37, 48, 16, 17], and fully-automatic [8, 55, 18, 10]. The former two kinds of approaches are user-guided that learn a mapping function to propagate user hint to the grayscale image. Since grayscale images only include the edge information, the results are highly relevant to the reasonability of human hints. On the other hand, fully-automatic algorithms directly solve an end-to-end objective from grayscale images to corresponding color embeddings. Normally, these approaches are trained on a very large dataset, which is essential for the system to exploit necessary information from the large-scale database without any human intervention.

Generative Adversarial Network for Image Enhancement. The image enhancement is a general idea to improve the image quality. It is addressed by a list of subtasks including demosaicking [58, 5], deblurring [24, 25], super-resolution [45, 57, 44], etc. The performance of image enhancement has been greatly improved through the data-driven deep learning approaches. Generative adversarial network (GAN), developed by Goodfellow et al. [12], defines a minmax game between generator and discriminator. The goal of generator is producing convincing samples which fool discriminator, so as to distinguish generated samples from ground truth. The first well-known general GAN-based image enhancer is Pix2Pix [21] that translates the images from two different domains. It was improved by Wang et al. [43] for processing high-resolution images and Zhu et al. [59] for multimodal generation.

3. Methodology

3.1. Problem Formulation

Suppose the clean images are from the domain Z and legacy noisy images are in domain N. The target is to process the legacy photo $n \in N$ and obtain colorful clean image $z \in Z$. The images in both domain Z and N are totally different in terms of noise, content, and color.

However, the spatial pixels of clean image z and legacy photo n are not aligned. To constitute valid training pairs, we propose to decolorize z and add pseudo noise to clean image $x \in X$ to obtain image $\hat{x} \in N$, which exists similar low-level characteristics of $n \in N$. We utilize a neural network G to simulate the blind noise for clean image x. This transformation process can be formulated as:

$$\hat{x} = G(x). \tag{1}$$

In order to recover colorful clean image z from the artificial degraded image $\hat{x}$, we summarize the process as denoising, inpainting, and colorization, respectively. Similarly, they are implemented by neural networks due to the highly non-linear process. To simulate random mask of legacy photos, we use several binary mask samples m to process $\hat{x}$ and obtain masked degraded image $\ddot{x} = \hat{x} \times m$. In addition to input image, we provide the masks and color scribbles to edit final colorized image and obtain sound perceptual quality. It can be represented as:

$$z = col(inp(den(\ddot{x})), s), \tag{2}$$

where $col(*)$, $inp(*)$ and $den(*)$ represent the colorization, inpainting, and denoising operations, respectively. The s is the color scribble provided by user. In practice, we combine $den(*)$ and $inp(*)$ into one architecture C to accelerate inference, while $col(*)$ is implemented by network R.

3.2. Training and Testing Pipeline

Figure 3 shows the training of the proposed pipeline. Specifically, the left and right part of Figure 3 correspond to representations of equation 1 and 2, respectively. They are concluded by two architectures, i.e. NEGAN and IEGAN, where NEGAN comprises the sub-network F and G, IEGAN comprises the sub-network C and R. The NEGAN is trained firstly. Then, it is used to degrade clean images,

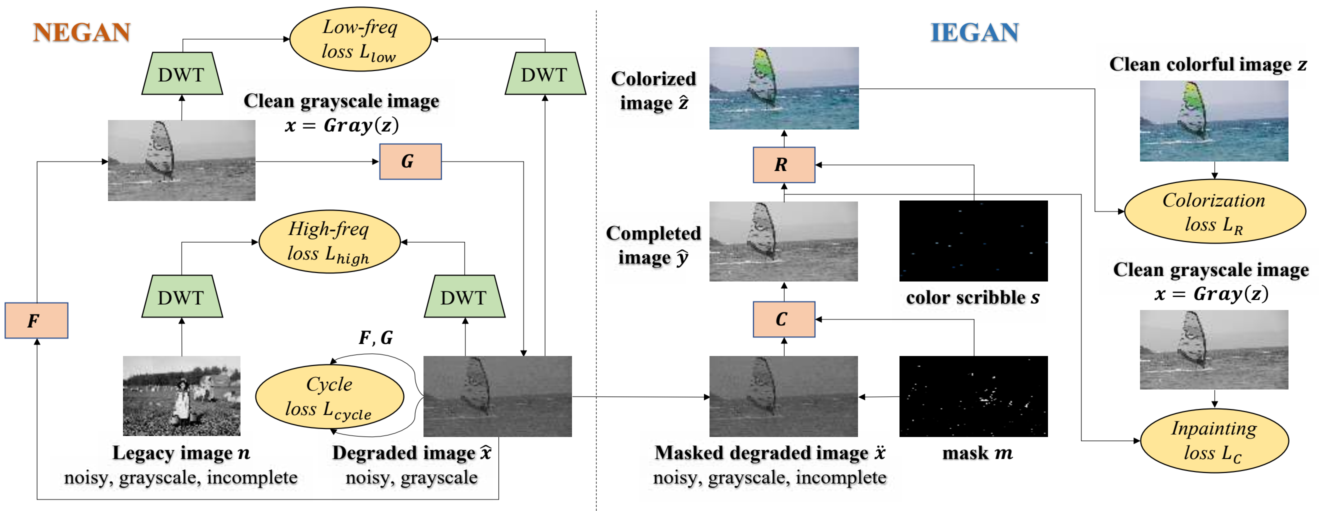

Figure 3. Illustration of training pipeline of proposed method. It contains four convolutional neural networks G, F, C and R. The left part (NEGAN) represents the process that learns noise prior. The right part (IEGAN) shows image editing procedure including the joint denoising, inpainting and scribble-based colorization.

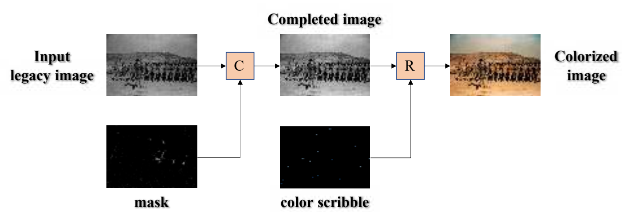

Figure 4. Illustration of testing stage of proposed system. The user-provided mask and color scribble map assist the system to produce photorealistic colorizations from legacy photos.

which are for IEGAN training. To stabilize training of IEGAN sub-networks, we propose to enforce a loss on C directly. Figure 4 shows the testing process, where only C, R with additional mask and color scribble are adopted to edit legacy photos. The network architectures and the training details will be presented in following paragraphs.

3.3. Noise Estimation GAN

The Noise Estimation GAN (NEGAN) including G and F aims to implement equation 1, i.e. NEGAN translates the images $x \in X$ to noisy image domain N. Since there is no evident noise model and paired training data in our application, the unique characteristics of noise become significant. A complete image is composed of low-frequency and high-frequency parts and we notice the noise occupies most high-frequency components of images. Based on the observation, we propose to utilize a clean image $x \in X$ and keep the original low-frequency part x_L. Then, we replace its high-frequency component x_H with statistics of noisy image from domain N to implement the translation. Therefore, we propose a Noise Estimation GAN (NEGAN) based on unpaired images to learn the implicit noise distribution, which is called noise prior in following text.

To divide the low-frequency and high-frequency parts, we need to map the images into frequency domain. The common way is to utilize low-pass and high-pass filters, e.g. Gaussian filter and its inverse. It can be defined as:

$$x = x_L + x_H = w * x + (\delta - w) * x, \tag{3}$$

where x, w, δ represent clean image, low-pass filter, and impulse function, respectively, while $(\delta - w)$ is viewed as high-pass filter since it is the reverse of filter w. The "$*$" is convolution operator. But kernel w is often set artificially, which cannot well separate different frequencies. To improve the functionality of the kernel, we introduce discrete wavelet transform (DWT) for frequency division and inverse discrete wavelet transform (IDWT) for image construction. Suppose two components x_L and x_H of input image x are derived from DWT, the whole learning losses for training NEGAN can be represented as:

$$\begin{aligned} L_{low}(G,F) = \mathbb{E}[||G(x)_L - x_L||_1] \\ + \mathbb{E}[||F(G(x))_L - x_L||_1], \end{aligned} \tag{4}$$

$$\begin{aligned} L_{high}(G, D_N, X, N) = \mathbb{E}_{x\sim X}[||(D_N(G(x)_H))^2||] \\ + \mathbb{E}_{n\sim N}[||(D_N(n_H) - 1)^2||], \end{aligned} \tag{5}$$

$$\begin{aligned} L_{cycle}(G,F) = \mathbb{E}_{x\sim X}[||F(G(x)) - x||_1] \\ + \mathbb{E}_{n\sim N}[||G(F(n)) - n||_1], \end{aligned} \tag{6}$$

$$\begin{aligned} L_{NEGAN} = \lambda_{low} L_{low}(G,F) + \lambda_{cycle} L_{cycle}(G,F) \\ + L_{high}(G, D_N, X, N) + L_{high}(F, D_X, X, N), \end{aligned} \tag{7}$$

where G, F, D_X, D_N denote generator from domain X to N, generator from domain N to X, and their corresponding discriminators, respectively. The x and n are random samples from both domains. The L_{high} utilizes the LSGAN loss term [32]. The L_{high} only matches the low-frequency part of images, which is different from CycleGAN. Also, the discriminators distinguish between fake and real noisy images by matching only the high-frequency part.

3.4. Image Editing GAN

The second step of proposed method is to recover a high-quality image from the pseudo noisy image by an Image Editing GAN (IEGAN). The inference of IEGAN is divided into two sub-networks: inpainting network (C) and colorization network (R). The C generates a complete grayscale image and the R colorizes the output of C. As shown in Figure 3, the proposed IEGAN framework receives pseudo noisy grayscale image with additional mask and color map guidances.

We utilize L1 loss for both sub-networks C and R. The losses for them share same representations. It is defined as:

$$L_1 = \mathbb{E}[||t_1 - t_2||_1], \tag{8}$$

where the two variables t_1 and t_2 equal to $\hat{y}$ and x for C, meanwhile they equal to $\hat{z}$ and z for R. The input $\ddot{x} = \hat{x} \odot m$ is a masked grayscale image with an additional Gaussian noise added. The outputs $\hat{y} = C(\ddot{x}, m)$ and $\hat{z} = R(y, s)$. The definitions can be found in Figure 3.

To boost perceptual quality of generated images, we adopt perceptual loss [22], which is defined as:

$$L_{percep} = \mathbb{E}[||\phi_l(t_1) - \phi_l(t_2)||_1], \tag{9}$$

where $\phi_l(*)$ represents the features of the l-th layer of the pre-trained CNN. In our experiment, we use the $conv_{4\_3}$ layer of VGG-16 [40] network, which is pre-trained on ImageNet [39] dataset.

Instead of traditional GAN training method [12], we utilize the PatchGAN [21] with LSGAN critic [32] to minimize the Pearson χ^2 divergence between the generated samples and ground truth. It is defined as:

$$L_G = \frac{1}{2}\mathbb{E}[(D(t_1) - 1)^2], \tag{10}$$

$$L_D = \frac{1}{2}\mathbb{E}[(D(t_2) - 1)^2] + \frac{1}{2}\mathbb{E}[(D(t_1))^2]. \tag{11}$$

The total loss functions of IEGAN can be defined as:

$$L_{IEGAN} = L_{1C} + L_{1R} + \lambda_{percep} L_{percepR} + \lambda_G L_{GR}, \tag{12}$$

where inpainting network C only adopts L1 loss term L_{1C}. The colorization network R utilizes all three loss terms L_{1R}, $L_{percepR}$, and L_{GR}. The definitions of the loss terms can also be found in Figure 3.

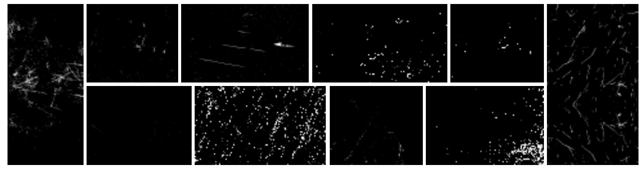

Figure 5. Illustration of mask templates used in this paper.

4. Experiment

4.1. Implementation Details

Dataset. We use LP dataset to include enough modes of noisy image domain N, for NEGAN training. There are over 25000 grayscale legacy photos with different resolutions in the dataset. Also, we choose ImageNet [39] (1.3 million images) for clean image domain X. It contains 1000 categories, which is general and robust for learning the mapping. At training, we randomly select unpaired sample $n \in N$ and $x \in X$. The images are randomly cropped to 256×256 local patches and normalized to range of [0, 1]. Moreover, the binary mask samples m is randomly cropped from templates, as shown in Figure 5.

Network Architecture. For NEGAN architecture, the generators adopt 8 residual blocks [15] as transformer with residual connection between input and output. There are no downsampling and upsampling operations since they may affect the low-level details. The discriminators adopt 16×16 PatchGAN architecture and all layers are spectral normalized [34]. The pre-trained NEGAN produces corresponding degraded images from input while maintains the low-frequency parts. For IEGAN architecture, the generator C and R adopt U-Net structure [38]. The convolutional layer of C is replaced by gated convolution [51] to learn adaptive inpainting. The discriminator D_C and D_R adopt convolution part of a VGG-16 architecture while the final output is one channel. The networks are instanced normalized [42]. Each layer is LeakyReLU activated [31].

Optimization. At first stage, the parameters of all networks are initialized using Xavier method [11] and the learning rate is initialized as 1×10^{-4}. The NEGAN and two sub-networks of IEGAN are trained independently for 20 epochs. At second stage, all networks are optimized jointly. The learning rate is fixed to 5×10^{-5} while the system is trained for another 20 epochs. The learning rate is fixed in both stages. We use Adam optimizer [23] with β_1 = 0.5, β_2 = 0.999 and batch size of 32. Moreover, we randomly select 0 - 30 color scribbles as hint for network R. The hyperparameters λ_{percep}, λ_G equal to 10 and 0.1, respectively. We implement our system with PyTorch framework and train it on 4 NVIDIA Titan Xp GPUs. It takes approximately 2 weeks to complete the whole training process.

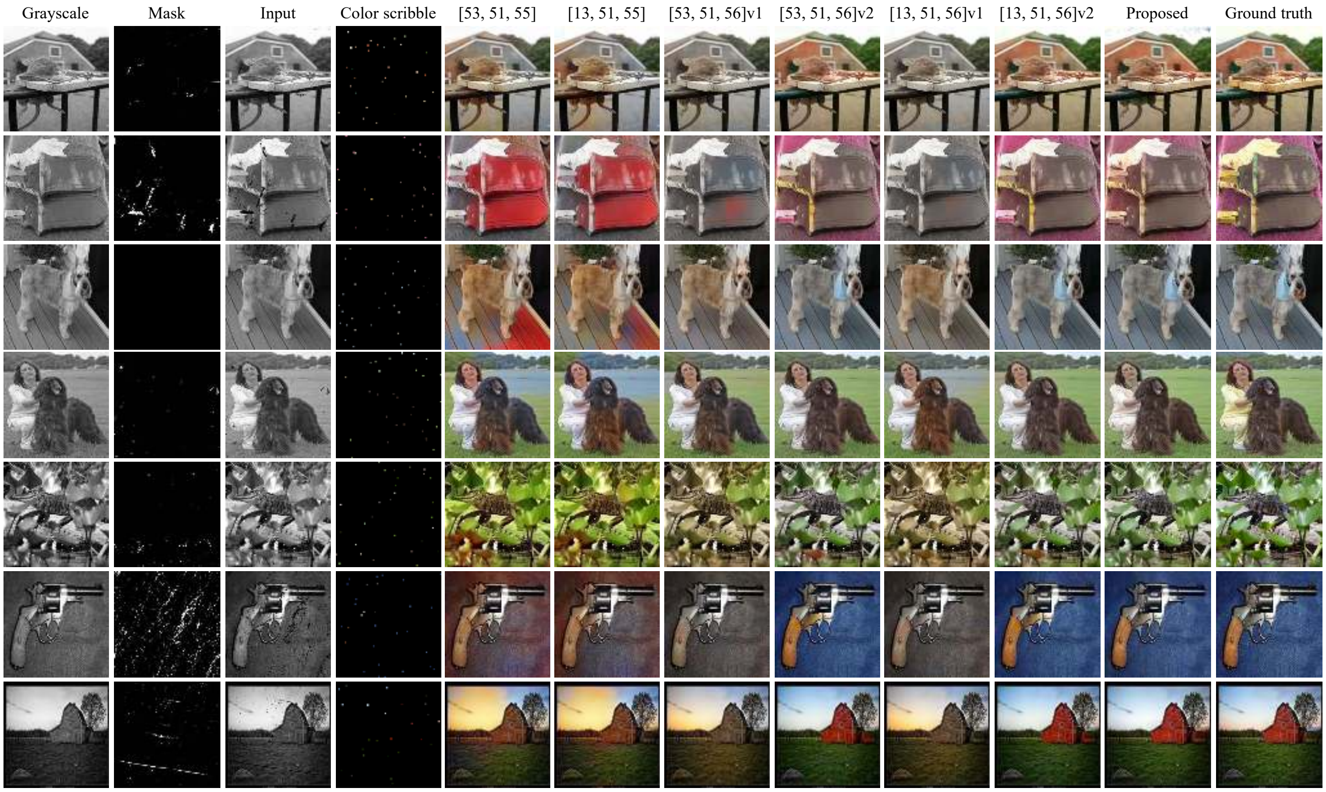

Figure 6. Illustration of image editing results. The input masked images are obtained by multiplication operation of grayscale images and masks. Different columns represent different samples edited by methods in experiment. They are randomly selected from validation set.

Table 1. Comparison results of the proposed pipeline and other 6 state-of-the-art pipelines. The grayscale images (clean) are obtained from ground truth colorful images. In "Reference" item, the "mask" and "color" denote the additional mask and color scribble input. Also, the [53, 51, 55] represents using [53], [51], [55] for inference sequentially.

Method	Reference	PSNR	SSIM
Grayscale (clean)	/	23.24	0.9443
[53, 51, 55]	mask	21.26	0.8865
[13, 51, 55]	mask	21.18	0.8865
[53, 51, 56]v1	mask	23.62	0.9059
[53, 51, 56]v2	mask, color	27.51	0.9233
[13, 51, 56]v1	mask	23.50	0.9024
[13, 51, 56]v2	mask, color	27.34	0.9194
Proposed	mask, color	**28.02**	**0.9408**

4.2. Validation on Image Editing Quality

In this section, we quantitatively evaluate the image enhancement quality of the proposed system. Since there is no ground truth for legacy photos, we alternatively adopt the ImageNet validation 50000 images. We convert the images to grayscale and rescale them to 256×256. Each validation image is added a pseudo mask and an additive Gaussian noise with standard deviation of 0.05 to simulate a legacy image, which is similar to training process. At inference stage, only IEGAN is used since the noise prior modelled by NEGAN is implied in C at training. We utilize different combinations of denoisers [53, 13], inpainting network [51], and colorization networks [55, 56] as pipelines and there are overall 6 combinations. All aforementioned algorithms are trained on ImageNet training dataset. Specifically, the denoisers are trained on the same noise level (i.e. AWGN) as validation data, whereas IEGAN is trained on blind noise learned from noise prior. The method [56] is a scribble-based colorization algorithm while [55] is fully-automatic colorization method. Color scribbles are used in both IEGAN and method [56]; therefore all the approaches in experiment adopt reference information. There are 30 scribbles used for IEGAN and for [56] at test.

The comparison results are summarized in Table 1 and illustrated in Figure 6. The methods using color scribbles achieve the PSNR higher than 27 and obviously outperform others since they have precise color prior. Note that, the proposed method achieves the highest PSNR and SSIM since the two parts of IEGAN are trained collaboratively. Therefore, the colorized images are more natural and realistic than other methods.

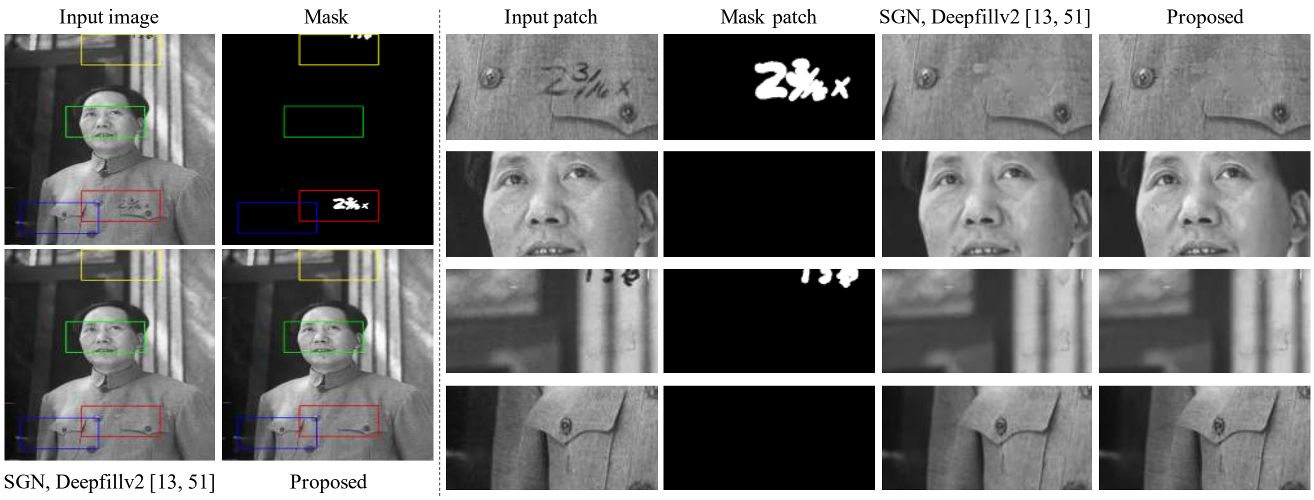

Figure 7. Comparison of legacy photo enhancement results of the proposed and previous [13, 51] pipeline. The left part and right part include the full resolution legacy photos and local patches, respectively. The colorful rectangles denote the locations of selected patches.

Table 2. Comparison results for ablation study.

Methods	Setting	PSNR	SSIM
w/o DWT-based losses	1)	27.91	0.9397
w/o perceptual loss	2)	27.78	0.9369
w/o GAN loss	2)	27.83	0.9396
w/o both losses	2)	26.18	0.9334
10 color scribbles	3)	26.78	0.9354
20 color scribbles	3)	27.59	0.9390
w/o joint training	4)	27.45	0.9358
Proposed	/	**28.02**	**0.9408**

4.3. Ablation Study

In order to demonstrate the effectiveness of NEGAN and IEGAN losses, we set up 4 ablation study settings. We use 50000 ImageNet validation data for validation. All images are added unknown noise by pre-trained NEGAN to simulate legacy photos. The settings are show as:

1) Drop the DWT-based loss terms that NEGAN noise prior learner retrogrades to a CycleGAN [59];

2) Drop the perceptual loss or GAN loss or both loss terms of IEGAN to compare their effectiveness, while the NEGAN remains unchanged;

3) Decrease the number of color scribbles to 20 or 10;

4) Train two sub-networks of IEGAN framework separately in order to evaluate joint training scheme.

As shown in Table 2, the full system reaches the best performance on PSNR and SSIM [46]. If the DWT-based loss terms are dropped, the system is hard to handle the "real noise" generated by the NEGAN. Also, each loss term or joint training contributes to better performance. In conclusion, all components of proposed method and significant.

4.4. Validation on Legacy Photo Enhancement

In this section, we assess the denoising and inpainting ability of the proposed system, i.e. network C of IEGAN. The state-of-the-art denoising and inpainting methods [13, 51] are used for comparison. For the denoising, the results of proposed approach are more sharper than [13, 51]. For instance, the eyebrows, cheeks and beard generated by the proposed method are more clear, as shown in the second patch in Figure 7. For inpainting, the patches produced by the proposed method are also realistic. For instance, the color of filled regions are closer to clothes, as shown in the first patch. Also, the patch of proposed model in third row is much more smoother than [13, 51]. Since the NEGAN better estimates the noise model, the generated results are cleaner and sharper. Moreover, the inpainted regions are more plausible due to better denoising ability.

4.5. Validation on Legacy Photo Colorization

In this section, we assess the editing quality of the proposed system on real legacy photos. We utilize the state-of-the-art pipelines, i.e. [13, 51, 56] for comparison and 15 color scribbles are adopted for both methods, as shown in Figure 8. The samples are selected from the proposed LP dataset. The proposed method produces more plausible colors than compared method since the it utilizes joint training scheme for image denoising, inpaiting and colorization. Moreover, the proposed method learns noise prior well, thus it produces high-quality images.

5. Failure Cases

Our system can predict relatively reasonable colorizations in many cases; however, there are still some common

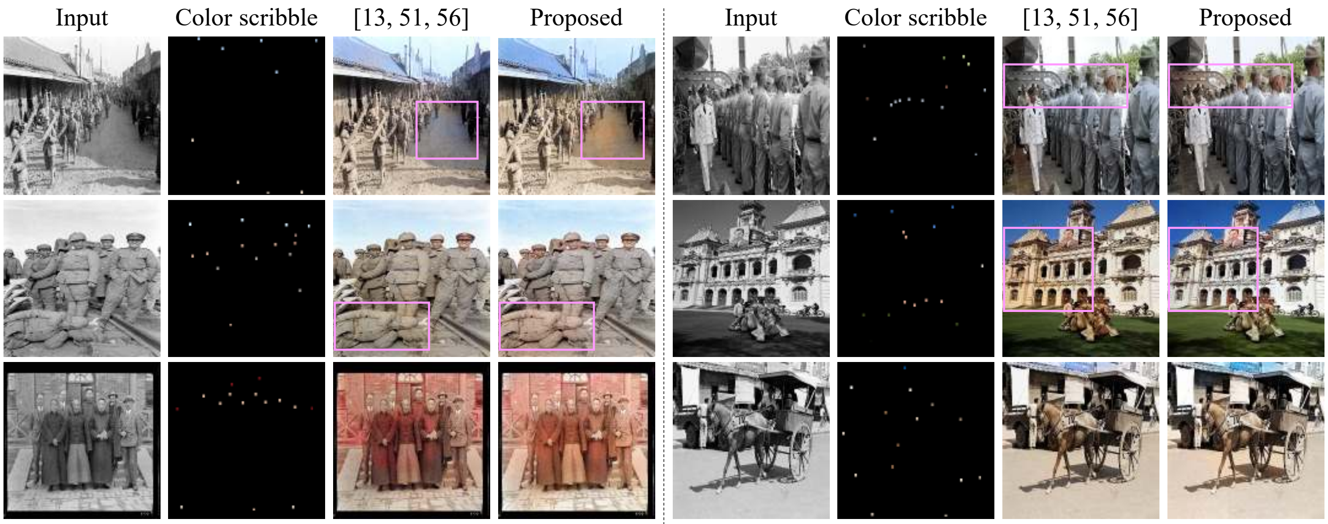

Figure 8. Comparison of the proposed and previous [13, 51, 56] pipeline on real legacy photos. The rectangles denote the highlighted areas.

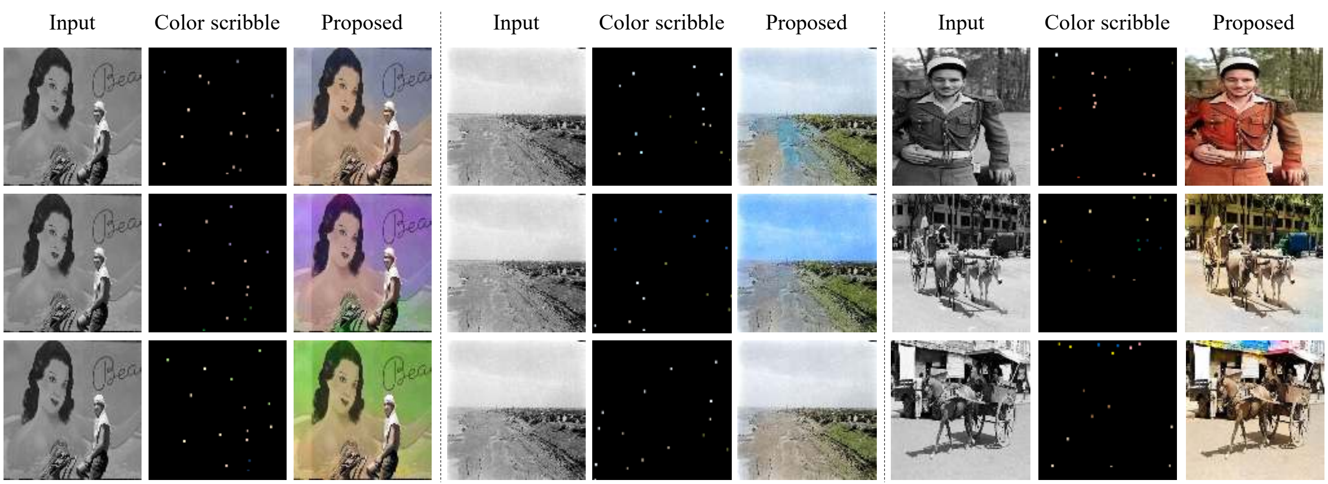

Figure 9. Illustration of some failure image editing cases of proposed method, including color bleeding, artifacts and inconsistent colors.

failure cases, shown in Figure 9. For left part (left 3 rows in the figure), there exists slight color bleeding effect when given "not reasonable color scribbles". Since the degradation degree is shifted for many legacy photos, the output images of center part still contain artifacts. Finally, as color scribbles provided by users are not reasonable enough, the results are also not very plausible, as right part shows. We will enhance the design of the proposed framework and. Moreover, we will add semantic information into our framework to guide inpainting and colorization in the future.

6. Conclusion

In this paper, we present a novel framework for editing legacy photos in an end-to-end manner. Since the legacy photographs are captured by old cameras, they are corrupted with undesirable noise, artifacts and saved in grayscale format. The noise is often blind, thus it is difficult to use a specific distribution for modelling. Thus, we propose a NEGAN to simulate noise prior learned from real legacy photos based on unpaired data training. We enforce the NEGAN to focus more on noisy parts (i.e. high-frequency components) of images by introducing DWT-based loss functions. Moreover, we collect a large-scale legacy photo dataset (LP dataset) including more than 25000 real photographs in different scenes for training NEGAN. Moreover, to remove the artifacts and colorize legacy photos, we propose an IEGAN that performs joint denoising, inpainting and scribble-based colorization sequentially, based on estimated noise prior. At test phase, users can edit the legacy photo by providing masks and color scribbles. Experimental results show that the proposed framework has better performance than the state-of-the-art pipelines.

References

[1] Abdelrahman Abdelhamed, Mahmoud Afifi, Radu Timofte, Michael S Brown, et al. Ntire 2020 challenge on real image denoising: Dataset, methods and results. In *Proc. CVPRW*, pages 496–497, 2020.

[2] Abdelrahman Abdelhamed, Marcus A Brubaker, and Michael S Brown. Noise flow: Noise modeling with conditional normalizing flows. In *Proc. ICCV*, pages 3165–3173, 2019.

[3] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *Proc. ICML*, pages 524–533, 2019.

[4] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *Proc. CVPR*, volume 2, pages 60–65, 2005.

[5] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proc. CVPR*, pages 3291–3300, 2018.

[6] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. Image blind denoising with generative adversarial network based noise modeling. In *Proc. CVPR*, pages 3155–3164, 2018.

[7] Xiaowu Chen, Dongqing Zou, Qinping Zhao, and Ping Tan. Manifold preserving edit propagation. *ACM Trans. on Graphics*, 31(6):1–7, 2012.

[8] Zezhou Cheng, Qingxiong Yang, and Bin Sheng. Deep colorization. In *Proc. ICCV*, pages 415–423, 2015.

[9] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, 2007.

[10] Aditya Deshpande, Jiajun Lu, Mao-Chuang Yeh, Min Jin Chong, and David Forsyth. Learning diverse image colorization. In *Proc. CVPR*, pages 6837–6845, 2017.

[11] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proc. AISTAT*, pages 249–256, 2010.

[12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Proc. NeurIPS*, pages 2672–2680, 2014.

[13] Shuhang Gu, Yawei Li, Luc Van Gool, and Radu Timofte. Self-guided network for fast image denoising. In *Proc. ICCV*, pages 2511–2520, 2019.

[14] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proc. CVPR*, pages 1712–1722, 2019.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. CVPR*, pages 770–778, 2016.

[16] Mingming He, Dongdong Chen, Jing Liao, Pedro V Sander, and Lu Yuan. Deep exemplar-based colorization. *ACM Trans. on Graphics*, 37(4):47, 2018.

[17] Satoshi Iizuka and Edgar Simo-Serra. Deepremaster: temporal source-reference attention networks for comprehensive video enhancement. *ACM Trans. on Graphics*, 38(6):1–13, 2019.

[18] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Trans. on Graphics*, 35(4):110, 2016.

[19] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion. *ACM Trans. on Graphics*, 36(4):1–14, 2017.

[20] Revital Ironi, Daniel Cohen-Or, and Dani Lischinski. Colorization by example. In *Rendering Techniques*, pages 201–210, 2005.

[21] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proc. CVPR*, pages 1125–1134, 2017.

[22] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Proc. ECCV*, pages 694–711, 2016.

[23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proc. ICLR*, 2014.

[24] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proc. CVPR*, pages 8183–8192, 2018.

[25] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proc. ICCV*, pages 8878–8887, 2019.

[26] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2noise: Learning image restoration without clean data. In *Proc. ICML*, pages 2965–2974, 2018.

[27] Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. In *ACM Trans. on Graphics*, volume 23, pages 689–694, 2004.

[28] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proc. ECCV*, pages 85–100, 2018.

[29] Pengju Liu, Hongzhi Zhang, Kai Zhang, Liang Lin, and Wangmeng Zuo. Multi-level wavelet-cnn for image restoration. In *Proc. CVPRW*, pages 773–782, 2018.

[30] Wei Liu, Qiong Yan, and Yuzhi Zhao. Densely self-guided wavelet network for image denoising. In *Proc. CVPRW*, pages 432–433, 2020.

[31] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, volume 30, page 3, 2013.

[32] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proc. ICCV*, pages 2794–2802, 2017.

[33] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In *Proc. NeurIPS*, pages 2802–2810, 2016.

[34] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *Proc. ICLR*, 2018.

[35] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Qureshi, and Mehran Ebrahimi. Edgeconnect: Structure guided image inpainting using edge prediction. In *Proc. ICCVW*, 2019.

[36] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros. Context encoders: Feature learning by inpainting. In *Proc. CVPR*, pages 2536–2544, 2016.

[37] Erik Reinhard, Michael Adhikhmin, Bruce Gooch, and Peter Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21(5):34–41, 2001.

[38] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proc. MICCAI*, pages 234–241, 2015.

[39] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.*, 115(3):211–252, 2015.

[40] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *Proc. ICLR*, 2014.

[41] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proc. ICCV*, pages 4539–4547, 2017.

[42] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[43] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proc. CVPR*, pages 8798–8807, 2018.

[44] Xuehui Wang, Qing Wang, Yuzhi Zhao, Junchi Yan, Lei Fan, and Long Chen. Lightweight single-image super-resolution network with attentive auxiliary feature learning. In *Proc. ACCV*, 2020.

[45] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proc. ECCVW*, pages 0–0, 2018.

[46] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004.

[47] Kaixuan Wei, Ying Fu, Jiaolong Yang, and Hua Huang. A physics-based noise formation model for extreme low-light raw denoising. In *Proc. CVPR*, pages 2758–2767, 2020.

[48] Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller. Transferring color to greyscale images. In *Proc. of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pages 277–280, 2002.

[49] Wei Xiong, Jiahui Yu, Zhe Lin, Jimei Yang, Xin Lu, Connelly Barnes, and Jiebo Luo. Foreground-aware image inpainting. In *Proc. CVPR*, pages 5840–5848, 2019.

[50] Li Xu, Qiong Yan, and Jiaya Jia. A sparse control model for image and video editing. *ACM Trans. on Graphics*, 32(6):1–10, 2013.

[51] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, and Thomas S Huang. Free-form image inpainting with gated convolution. In *Proc. ICCV*, pages 4471–4480, 2019.

[52] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *Proc. CVPR*, pages 2696–2705, 2020.

[53] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.*, 26(7):3142–3155, 2017.

[54] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Trans. Image Process.*, 27(9):4608–4622, 2018.

[55] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *Proc. ECCV*, pages 649–666, 2016.

[56] Richard Zhang, Jun-Yan Zhu, Phillip Isola, Xinyang Geng, Angela S Lin, Tianhe Yu, and Alexei A Efros. Real-time user-guided image colorization with learned deep priors. *ACM Trans. on Graphics*, 9(4), 2017.

[57] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *Proc. ICCV*, pages 3096–3105, 2019.

[58] Yuzhi Zhao, Lai-Man Po, Tiantian Zhang, Zongbang Liao, Xiang Shi, Yujia Zhang, Weifeng Ou, Pengfei Xian, Jingjing Xiong, Chang Zhou, et al. Saliency map-aided generative adversarial network for raw to rgb mapping. In *Proc. ICCVW*, pages 3449–3457, 2019.

[59] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proc. ICCV*, pages 2223–2232, 2017.