Multi-scale Generative Adversarial Deblurring Network with Gradient Guidance

Jinxiu Zhu[1], Xue Xu[1], Chang Choi[2], Xin Su[1]*

[1] College of IoT Engineering, Hohai University, China
[2] Department of Computer Engineering, Gachon University, Republic of Korea
19941489@hhu.edu.cn, xx_flyimg@163.com, enduranceaura@gmail.com, leosu8622@163.com

Abstract

With regards to the lack of crisp edges and a poor recovery of high frequency information such as details in deblurred motion pictures, this research proposes a multi-scale adversarial deblurring network with gradient guidance (MADN). The algorithm uses the classical generative adversarial network (GAN) framework, consisting of a generator and a discriminator. The generator includes a multi-scale convolutional network and a gradient feature extraction network. The multi-scale convolutional network extracts image features at different scales with a nested connection residual codec structure to improve the image edge structure recovery and to increase the perceptual field. This gradient network incorporates with intermediate scale features to extract the gradient features of blurred images to obtain their high frequency information. The generator combines the gradient and multiscale features to recover the remaining high-frequency information in a deblurred image. The loss function of MADN is formed in this research combining adversarial loss, pixel L2-norm loss and mean absolute error. Compared to those experimental results obtained from current deblurring algorithms, our experimental results indicate visually clearer images retaining more information such as edges and details. This MADN algorithm enhances the peak signal-to-noise ratio by an average of 3.32dB and the structural similarity by an average of 0.053.

Keywords: Motion deblurring, Multi-scale network, Nested residual connection, GAN, Gradient feature extraction

1 Introduction

Image deblurring technique is getting an increasing popularity in recent years in fields like transportation, medicine and remote sensing, even in artificial intelligence. Image deblurring algorithms improve image quality and recover lost features in blurred images. Image deblurring technology also lays the foundation for subsequent computer vision tasks, such as target detection [1-2]. General motion blurs occur when images are captured in daily life with objects in relative motions. Camera shaking and scene depth changes come next.

Image deblurring aims at a sharper image over the blurred one. Current recovery methods fall within two categories, namely, non-blind and blind deblurring, depending on whether the blur kernel is known. The majority of early researches [3-4] concentrate on non-blind deblurring, predicting the blur kernel and implementing the deblurring through a deconvolution operation. Non-blind methods primarily introduce a prior knowledge as a regularization term to solve the mathematical model, and iteratively estimate the blurring kernel. Despite good recovery results, non-blind methods mostly occur in uniform blurred images, and are computationally intensive and prone to ringing on other blurred images [5]. Most real-world blurred images are non-uniform with multiple scales of blur kernels. Chakrabarti [6] uses a Convolutional Neural Network (CNN) to estimate the fuzzy kernel by predicting their fourier coefficients in the frequency domain, which functions well only for uniform motion fuzzy image recovery and is time-consuming. On account of the fuzzy kernel's unknown nature, blind image recovery approaches have been more widely used [7-8]. Besides, traditional blind recovery approaches focus on blurs caused by simple target motion, camera translation, and other factors. Blurs in real-world dynamic scenes are so complex that traditional blind recovery methods have a very limited role to play here. Kotera et al. [9] introduce a super-Laplacian prior as well as L0 parametric constraints, but these models do not fully express the edge sparsity characteristics of natural images, resulting in inaccurate estimations of blur kernels.

With its ever-increasing computational power of neural networks, deep learning extends applications to image processing tasks, including image debluring and classification [10]. Furthermore, image datasets have progressed from manual syntheses to real-world images, closer to the genesis blurring origin, so datasets also drive the advancement of image deblurring algorithms. Blind image deblurring algorithms are mainly classified into CNNs, GANs and regularization-based approaches. Most CNN-based image blind deblurring algorithms [11-12] only extract image features at a single scale, and insufficiently make use of multi-scale image features. With growing artificial intelligence, GAN [13] finds an increasingly wider range of applications in image processing, from super-resolution reconstructions [14] to image deblurring tasks.

Since the solution of the blind recovery model is not unique, certain prior conditions are required to constrain the model. The regularization-based blind image recovery algo-

*Corresponding Author: Chang Choi; E-mail: enduranceaura@gmail.com
DOI: 10.53106/160792642023032402003

rithm uses regular terms to constrain the prior information and further constructs regularization models. Natural images contain rich details and features, so selecting suitable regular terms to constrain these features can help recover these details and features. Currently, a variety of priors are proposed, such as gradient sparse priors [15] and low-rank prior [16], among which algorithms using gradient prior have a wide range of applications.

A MADN algorithm is proposed in this research, based on gradient guidance for a partial feature loss in motion image deblurring tasks caused by multiscale networks. A gradient feature extraction network is created, drawing on gradient sparse prior to help recover features such as details. The work covers the following four points:

(1) A multi-scale network (SN) is proposed based on the characteristics of non-uniform motion blurred images. In order to capture effective multi-scale features, SN adds cross-scale [17] connections to share network weights.
(2) A gradient feature extraction network (GN) is added to the generator, considering the fact that sharing weights may cause high-frequency information loss, unfavorable for image detail recovery. Next, the scale features are coupled with GN to reconstruct the gradient map, avoiding over-reliance on blurred images when extracting features.
(3) A nested residual-connected network is constructed in the codec structure to prevent problems like gradient disappearance. This approach makes full use of the underlying features to accelerate the network convergence while reducing the artifacts brought by feature stacking.
(4) The proposed loss function's effectiveness is verified through ablation experiments. MADN is trained and tested on the GoPro dataset, in comparison with current deblurring ones.

2 Related Work

The image degradation model is introduced first in this section, followed by a brief review to different learning-based methods for blind image deblurring, mostly in two categories: CNN methods, and GAN methods. Also investigated are methods related to gradient prior. MADN research is accordingly conducted with a reference to the above-mentioned methods.

2.1 Degradation Model

The fuzzy image degradation process is obtained through a convolution operation of a clear image with a fuzzy kernel plus noise. The mathematical model is expressed as follows:

$$f = k * g + n. \tag{1}$$

Where f is the blurred image, k is the fuzzy kernel, also known as the point spread function, g is the clear image, n is the additive noise, and * is the convolution operation. Two-dimensional filter matrices, namely, the convolution kernel and the CNN fuzzy kernel extract image features through layer-by-layer convolution changes, resulting in a spatial mapping of blurred images to sharp images.

2.2 Learning-based Methods

CNN used for computer vision tasks has been extensively studied with the development of deep learning. Due to the remarkable feature extraction ability of CNN, its application has even expanded to the field of integrated Internet of things in recent years [18]. Blind deblurring methods for motion images take advantage of the local weights sharing to process high-dimension data in a convolutional operation. Sun et al. [19] adopt CNN to remove non-uniform blurring on real blurred images with a poor recovery, as the dataset is artificially synthesized uniform blurred images. For this reason, Nah et al. [20] propose a dataset with sharp blurred image pairs-GoPro. This approach follows the multi-scale mechanism in traditional methods for blurring in dynamic scenes and obtained better results. However, independent parameters at each scale in the network result in a complex structure and slow convergence. To address the problem, a scale recurrent network model is proposed with shared parameters to achieve a desirable effect by Tao et al. [21]. With fewer parameters, this model is simpler to implement. Based on Tao et al., Zhang et al. [22] propose a deep hierarchical multi-patch network, greatly improving the running speed.

Parameter sharing causes significant scale feature losses, unfavorable to detail recovery. Thus, CNN alone in unfit to eliminate motion blurs and recover details, besides the artifacts resulted from ablation trials. Sharp images are known to be rich in textural and structural features, hence image recovery tasks necessitate a network with advanced feature extraction capabilities. Based on game theory principles, the GAN-based image motion deblurring method improves parameter fitting accuracy, restoring more realistic outcomes.

GAN is first used by Radford et al. [23] for image generation to construct deep convolutional generative adversarial network models in 2010s. Richer details and higher subjective visual quality are achieved in DeblurGAN algorithm based on paired datasets, but the texture structure recovery is less effective [24]. To solve the above problem, DeblurGAN-v2 [25] is proposed to better recover texture structure, taking spatial pyramids as the backbone network. Lately, a self-referential deblurring GAN is designed by Gong et al. [26], which is better at recovering locally blurry images.

In brief, being data-driven, GAN is difficult to obtain sharp-blurry image pairs of the same real scenes, causing self-adaptability deficiency with regards to non-uniform blurs. That is why the generator in this research combines scale features and shares weights across distinct scales to increase the capacity of handling non-uniform blurry feature.

2.3 Gradient-based Methods

Most neural network methods integrated with traditional image prior fail to recover details and textures. Researches [27-29] decompose blurry images to extract features from each layer separately, and guide the blurry image recovery with gradient low-rank prior. Gradient feature has been utilized in previous work [30-31].

It has been demonstrated in the field of image super-resolution that using gradients as a prior to guide image res-

toration better preserve edge features and generate sharper high-resolution images.

Under the guidance of the edges extracted by off-the-shelf edge detector, Yang et al. [32] introduce a recurrent residual network to reconstruct fine details, aiming to recover high-frequency components. Structure-Preserving Super Resolution Network (SPSR) proposed by Ma et al. [33] experimentally verifies the feasibility of generating rich texture and detail in an adversarial training manner guided by gradient features. However, when paired with the incoming features from the super-resolution encoder, the gradient network in SPSR relies heavily on the specific blurry image, resulting in ineffective gradient maps. In summary, those models combined with gradient prior are able to help to restore sharper edges and smoother details. Thus, an introduction of a gradient feature extraction network becomes the focus for high frequency information recovery such as details.

Based on the above discussion, a multi-scale adversarial deblurring network is designed with gradient guidance, according to motion blur image characteristics. The generator (G) consists of a Scale-Network (SN) and a Gradient Network (GN) (See 3.2). Gradient features are combined with the multiscale features to recover a sharp image with high frequency information. What's more, a loss function consisting of content loss and adversarial loss is designed to extract effective features.

3 The Proposed Method

This section first presents the formula transformation and the architectural design of the proposed model. It then separately elaborates on the detailed compositions of SN and GN.

3.1 Transformed Formula

According to Zhu's modeling idea [34], we have derived the following formula. The convolution operation is described by matrix multiplication, then Equation 2 is written as follows:

$$\begin{aligned} F &= K \cdot H + N \\ H &= \frac{1}{K} \cdot F + \frac{-N}{K}. \end{aligned} \tag{2}$$

Where F, K, H, N are represented as a fuzzy image, a fuzzy kernel, a sharp image and noise in matrix form respectively. The operation in Equation 2 is then expressed as a convolution operation, which is written as Equation 3:

$$g = W * f + B. \tag{3}$$

Where W denotes the weight parameter of the neural network and B denotes the network bias. W can be defined by the fuzzy kernel matrix, B is related to the noise:

$$\begin{cases} W \triangleq 1/K \\ B \propto N \end{cases}.$$

The fuzzy kernel k is converted into the weights W that the neural network needs to learn.

G takes the blurry image X_b as input and generates the deblurred image $\hat{X}_b$ based on the sharp image X_t. In this research, the image deblurring problem is converted into an image generation problem. G is trained to learn the mapping $g(\cdot)$ from the blurry image to the clear one without estimating the blurring kernel. $g(\cdot)$ is obtained according to Equation 3:

$$\hat{X}_b = g(X_b) = W * X_b + B. \tag{4}$$

In order to learn $g(\cdot)$, G needs to update its own parameter θ_g during the training process, thus, Equation 4 can be re-written as:

$$\hat{X}_b = g(X_b; \theta_g). \tag{5}$$

Regularization theory [35] is a widely-used algorithm in image blind recovery, whose basic concept is to select a suitable regularization term to constrain the prior information. The traditional image restoration method utilizes a gradient sparse prior to extract image edges, combined with a matrix low-rank prior to suppress noise interference [36]. Equation 6 represents the gradient sparse prior, and Equation 7 adds a matrix low-rank prior to Equation 6 for the noise suppression:

$$\arg\min_{\hat{X}_b, k} \left\| \hat{X}_b * k - X_b \right\|_2^2 + \delta \|k\|_2^2 + \rho \left\| X_b \right\|_0. \tag{6}$$

$$\begin{aligned} (\hat{X}_b^*, k^*) = \arg\min_{\hat{X}_b, k} & \left\| X_b * k - X_b \right\|_2^2 + \delta \|k\|_2^2 \\ & + \rho \left\| \hat{X}_b \right\|_0 + \zeta \sum\nolimits_i \left\| \nabla X_{b_i} \right\|_{w,*}. \end{aligned} \tag{7}$$

Here, δ, ρ, ζ are regularization parameters, $\|\cdot\|_{w,*}$ is weighted nuclear norm. Since bias B is related to noise n, which, in analogy to traditional low-rank prior, acts as a noise suppressor in neural networks.

Since image deblurring is an ill-posed problem, the solution process for the optimal solution θ_g^* is combined with the natural image priors described above, and the following model is proposed:

$$\theta_g^* = \arg\min_{\theta_g} \left\| g(X_b; \theta_g) - X_t \right\|_1 + \mu L_{grad}. \tag{8}$$

L_{grad} is a regularization term based on gradient sparse prior, while μ is the regularization parameter. In this research, Sobel operator is adopted to extract the image gradient [37]. In this way, the above formula derivation manifests the general MADN design principle.

3.2 Overall Network Structure

Motivated by Equation 8, the architecture of MADN is designed as shown in Figure 1. A generative network is designed containing SN and GN. Where G learns the mapping from the blurry image to the recovered one and generates a sharp result, visually closer to the target. D discriminates between the generated results and the real-world images.

Here, $\hat{X}_g$ denotes the gradient reconstruction map output by G, $M(X_t)$ denotes the sharp gradient map, I_i denotes the scale features, $grad\_f$ denotes the gradient features, while L_{pixel}, $L_{content}^{GB}$, $L_{content}^{G}$, L_{adv} and L_d stand for pixel loss, content loss of GN, content loss of G, adversarial loss of G and discriminant loss respectively. The solid black line indicates the feature transmission while the dashed line indicates the back propagation of the loss function's gradient.

The specific deblurring process is to input the blurry images into G in batches, after which they are fed straight into the SN network, whose gradient map serves as the input to GN. The scale features are connected to GN in a hierarchical manner to guide the gradient image reconstruction. Finally, gradient features are merged with scale features for an up-sampling operation to recover the final deblurred images. During the training procedure, G and D are optimized according to their own loss functions in an alternative manner until the loss function values reach equilibrium.

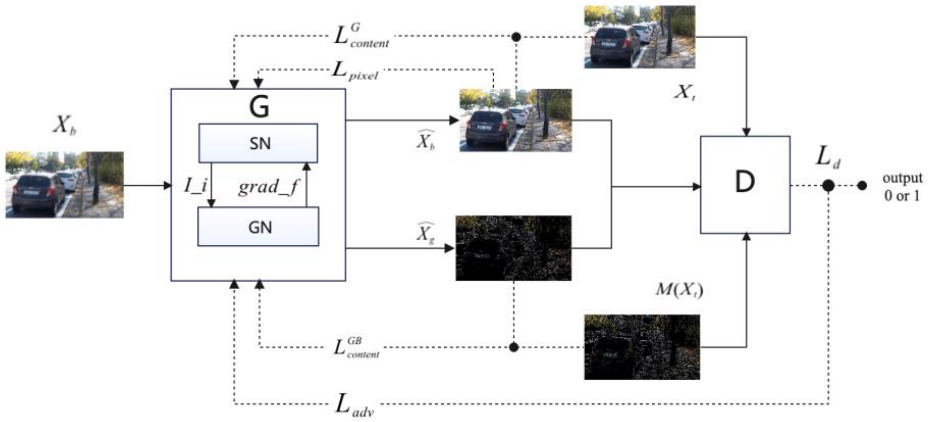


Figure 1. The architecture of the proposed MADN

3.3 Design of the Generator

Inspired by Tao et al. [21] and Ma et al. [33], the generator model G is composed of SN and GN (Figure 2). GN first feeds the acquired multi-scale features into its own encoder in a hierarchical manner, then it combines the gradient features with scale features to up-sample the final deblurred image (output).

3.3.1 Design of Multi-scale Structure

In this project, SN is designed with the traditional coarse-to-fine concept image deblurring to obtain features at different scales. In order to reduce parameter numbers in SN, a recurrent connection is added, that is, the previous layer's output is adopted as the next layer's input to increase the training speed in a weight-sharing way. The specific steps are to input the minimum scale blurry image blur_3 for deblurring. A sharp latent image I_3 is obtained, then I_3 is up-sampled to the same scale as blur_2. They are fed into the next scale together to obtain I_2. Similarly, the final sharp latent image I_1 is obtained from SN.

The same codec structure is utilized at each scale [38]. The encoder maps the input image to the low-dimensional space through a down-sampling operation, while the decoder maps the compressed image to the output layer through an up-sampling operation. Skip connection is also added to solve those problems, such as the loss of characteristics during decoding. That is, the encoder passes the image features to the decoder through a cross-layer skip connection, so that the features at the lower layer are combined with the features at the higher layer.

The specific parameters in the model are set as follows. As shown in Figure 2, each SN level consists of one input block (InBlock), two encoding blocks (EBlock), two decoding blocks (DBlock) and one output block (OutBlock). InBlock and EBlock contain one convolutional layer and four residual blocks (ResBlock), while DBlock and OutBlock contain four residual blocks and one deconvolutional layer sequentially. The convolutional kernel size used in the image restoration method is usually 5×5 or 7×7. Considering the training cost and the parameter number, the convolutional kernel size is set to 3×3 in this research. It's widely accepted that two 3×3 convolutional kernels are used instead of one 5×5 convolutional kernel to ensure the same perceptual field. The activation functions of all layers are rectified linear units (ReLU). The blurred image is first input to the encoder, which outputs a compressed feature map after down-sampling operation with a step size of 2. The decoder performs up-sampling operation on the feature map to gradually recover the image to its original size. The final convolution operation uses a convolution kernel of size 1×1 to achieve the fusion of features obtained from different channels. According to the relevant experimental verification [19-20], the recovery effect is better when the scale is taken as 3.

3.3.2 Design of the Gradient Network

The gradient network is based on the U-Net structure [39] with skip connection, as shown in Figure 2, where gradient features are extracted through sampling operations. The basic module composition in GN remains the same as in SN. The input of GN is a fuzzy gradient map. A concat layer is added after each EBlock to superimpose the corresponding scale features from the SN output. The decoding section outputs the reconstructed gradient image though up-sampling. Finally, the gradient features are combined with the multi-scale features, and then input to the up-sampling block for generating the final restored image.

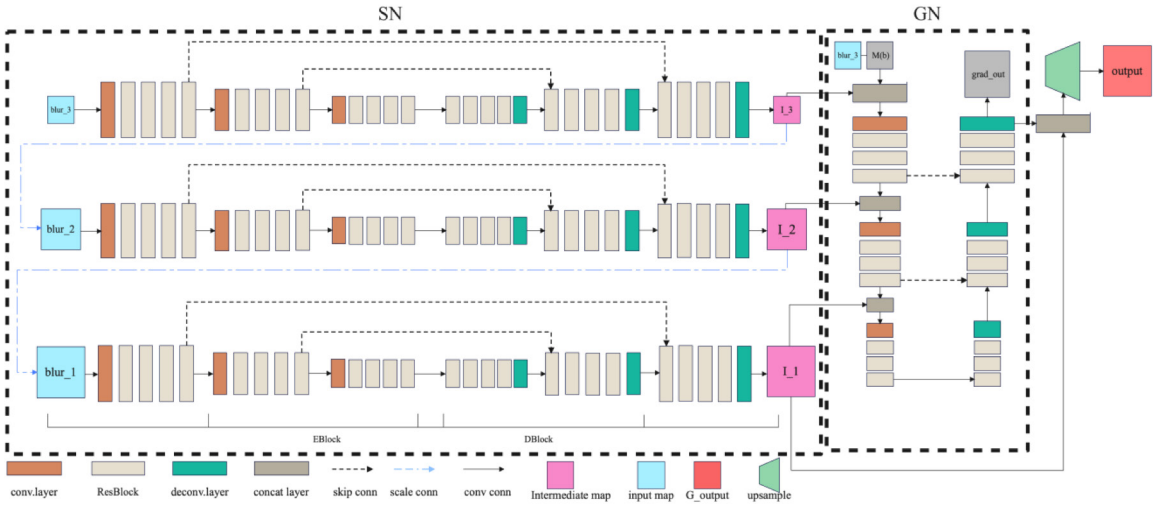


Figure 2. The architecture of the designed Generator

3.3.3 Design of the Residual Blocks

To take full advantage of the features of different network layers and expand the perceptual field, this research extends the residual block connection [40] to a nested connection, as shown in Figure 3.

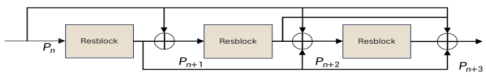


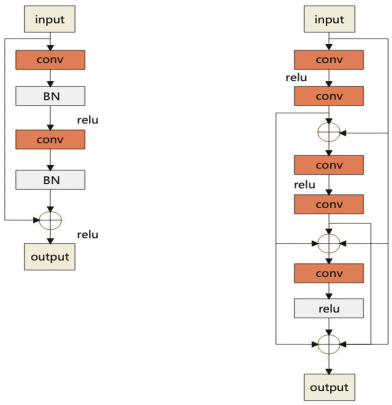
Figure 3. The method of nested residual linking

The residual network structure is expressed as follows:

$$P_{n+3} = P_n + R_1(P_n) + R_2(P_n + R_1(P_n)). \quad (9)$$

Where P_n denotes the input to the current nth layer of the network, where $R(\ )$ denotes the residual connection, $R_1(P_n) = P_{n+1}$. Residual network's input enters the interior of the ResBlock on one side and is connected to the convolutional layer of the current ResBlock as the next ResBlock's input on the other side [41]. The output of the current ResBlock is used as the next ResBlock's input through nested connections. This kind of nested connection can effectively exploit the features between different ResBlocks, which not only reduces parameter numbers, and prevents gradient disappearance as well.

Since the multi-scale structure in this research uses a scaling factor of size 0.5, the BN layer is removed from the ResBlock to increase the network's flexibility during training. In addition, removing the rectified linear unit before the final layer can speed up the model's convergence. A comparison of internal structure with the original residual block is shown in Figure 4's (a), (b).



(a) Original Resblock (b) Designed Resblock

Figure 4. Comparison of different Resblock structures

3.4 Design of the Discriminator

The image source is discriminated with the discriminator. In order to recover richer details, D in this research abandons pooling layers. Each convolution layer uses a Leaky Rectified Linear Unit (LeakyReLU) with a slope of 0.2 as the activation function. The structure is shown in Figure 5, where g_out denotes the generated result and gt denotes the ground truth, i.e. the real sharp image.

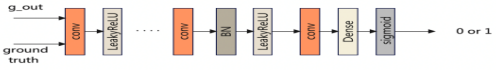


Figure 5. The architecture of the designed Discriminator

The input is compressed by down-sampling, then the channel is compressed by two dimensionally invariant 1×1 convolutions, followed by a Dense Layer. The output is mapped between 0 and 1, using sigmoid activation function. The output is judged to be sharp when it is greater than 0.5, and fuzzy otherwise. The specific network hierarchy and parameter settings are shown in Table 1.

Table 1. Parameter settings

Layer	Kernel size	Channel	Strides
Conv_1~2	3x3	64	2
Conv_3~4	3x3	64	2
Conv_5~6	3x3	128	2
Conv_7~8	3x3	256	2
Conv_9~10	3x3	512	2
Conv_11~12	3x3	1	2
Dense_1	3x3	1024	1
Dense_2	-	1	-

4 The Proposed Loss Function

A loss function is used to guide the training direction of the model. In this research, the total loss function consists of content loss and adversarial loss. The content loss consists of pixel loss and the mean absolute error (MAE). MAE is constrained between the generated result and the baseline image in both image and gradient domains.

To optimize SN, pixel loss is used based on L2 norm for the output and the sharp image, which helps to correct the color and texture features. Theoretically, blurred images' edges are irregular and sharp images' edges are prominent, gradient L1 loss is introduced to enrich the edge features and structural features.

In addition, Lim et al. [42] show that for many image restoration tasks, MAE loss outperforms mean squared error (MSE) loss, which smooths out edges and local details in the reconstructed image and influences the feature accuracies obtained by the gradient network. Therefore, for the gradient domain, this research employs a MAE-based content loss.

4.1 The Loss Function of the Generator

The target of G is to deceive D as much as possible so that D cannot distinguish the image sources, i.e. the training goal of G is to maximize the probability that D will discriminate the generated output as a real image. This research uses the joint content loss and adversarial loss as the loss function of G:

$$L_G = L_{content} + \lambda L_{adv}. \tag{10}$$

Where $L_{content}$ is the content loss, L_{adv} is the adversarial loss and λ is the weighting factor for the adversaria loss.

Content Loss: To ensure the generated image's authenticity, this research uses content loss constraints in both the gradient and image domains. $L_{content}$ includes pixel loss L_{pixel} and mean absolute error L_{MAE} (Equation 11):

$$L_{content} = \alpha \cdot L_{pixel} + L_{MAE}. \tag{11}$$

Where α is the weighting factor, the specific expression for L_{pixel} is shown in Equation 12:

$$L_{pixel} = \sum_{i=1}^{k} \frac{w^i}{2N^i} \left\| I_i^{out} - I_i^{true} \right\|_2^2. \tag{12}$$

Where i is the scale level, w^i is the weight at scale, N^i is the number of elements normalized at scale i, I_i^{out} and I_i^{true} are the output of the model at scale i and the real image respectively. L_{MAE} is expressed as:

$$\begin{aligned} L_{MAE} &= \phi_1 \cdot L_{content}^{GB} + \phi_2 \cdot L_{content}^{G} \\ &= \phi_1 \cdot \left\| M(X_t) - G_{\theta_{gb}}(X_b) \right\|_1 \\ &+ \phi_2 \cdot (\beta_1 \left\| M(X_t) - M(G_{\theta_g}(X_b)) \right\|_1 \\ &+ \beta_2 \left\| G_{\theta_g}(X_b) - X_t \right\|_1). \end{aligned} \tag{13}$$

Where $L_{content}^{GB}$ constrains between the gradient reconstruction map and the sharp baseline gradient map with weight coefficient ϕ_1, $L_{content}^{G}$ constrains the final deblurred image and its gradient map with weight coefficient ϕ_2, β_1 and β_2 are the weight coefficients of the L1 loss. M denotes the operation of extracting image gradients, θ_{gb} denote the parameters of GN, while $G_{\theta_g}(X_b)$ and $G_{\theta_{gb}}(X_b)$ respectively denote the final generated deblurred image and the reconstructed gradient map.

This research uses bicubic interpolation to avoid jaggedness when changing the image size. The deblurred image has smoother edges and higher quality of detail recovery.

Adversarial Loss: The adversarial loss L_{adv} is used to constrain the output of G to a sharp image that is discriminated as true by D. The expression for the adversarial loss is shown in Equation 14, where L_{adv}^{G} (Equation 15) denotes the adversarial loss in the image domain, represented by, and L_{adv}^{GB} (Equation 16) denotes the adversarial loss in the gradient domain.

$$L_{adv} = L_{adv}^{G} + L_{adv}^{GB}. \tag{14}$$

$$L_{adv}^{G} = -E_{X_b \sim P_{blur}}[\log(D(G_{\theta_g}(X_b)))]. \tag{15}$$

$$L_{adv}^{GB} = -E_{X_b \sim P_{blur}}[\log(D(G_{\theta_{gb}}(M(X_b))))]. \tag{16}$$

E denotes the mathematical expectation, P_{blur} denotes the data distribution of blurred images in the training set, $D(G_{\theta_g}(X_b))$ denotes the discriminatory result on the deblurred image, and $D(G_{\theta_{gb}}(M(X_b)))$ denotes the discrimina-

tory result on the gradient reconstructed image. The adversarial loss minimized entails a maximized probability that D discriminates the generator's output as true. In summary, generator's loss function can be written as the following equation:

$$\begin{aligned} L_G &= L_{content} + \lambda L_{adv} \\ &= \alpha \cdot L_{pixel} + L_{MAE} + L_{adv}^{G} + L_{adv}^{GB} \\ &= \alpha \cdot L_{pixel} + \phi_1 + L_{content}^{GB} \\ &\quad + \phi_2 \cdot L_{content}^{G} + L_{adv}^{G} + L_{adv}^{GB}. \end{aligned} \tag{17}$$

4.2 The Loss Function of the Discriminator

D aims to distinguish between the real image and the generated image. D is optimized in parallel with G. The task of D is to keep the discriminant close to 1 for the sharp image and close to 0 for the generated image. The internal parameters of D are continuously optimized according to its loss L_d, denoted by Equation 18. L_d^G for the deblurred image is denoted by Equation 19. L_d^{GB} for the gradient reconstruction map is denoted by Equation 20.

$$L_d = L_d^G + L_d^{GB}. \tag{18}$$

$$\begin{aligned} L_d^G &= -E_{X_t \sim P_{gt}}[\log(D_{\theta_d}(X_t))] \\ &\quad -E_{\hat{X}_b \sim P_g}[\log(1 - D_{\theta_d}(\hat{X}_b))]. \end{aligned} \tag{19}$$

$$\begin{aligned} L_d^{GB} &= -E_{X_t \sim P_{gt}}[\log(D_{\theta_d}(M(X_t)))] \\ &\quad -E_{\hat{X}_b \sim Pg}[\log(1 - D_{\theta_d}(\hat{X}_g))]. \end{aligned} \tag{20}$$

Where P_{gt} stands for the data distribution of real sharp images in the training set, P_g denotes the data distribution learned by G and θ_d denotes D's parameter. The image domain $E_{X_t \sim P_{gt}}[\log(D_{\theta_d}(X_t))]$ denotes the discriminant expectation of D on the real image, and $E_{\hat{X}_b \sim P_g}[\log(1 - D_{\theta_d}(\hat{X}_b))]$ denotes the discriminant expectation of D on the generator result, while $\hat{X}_b = G_{\theta_g}(X_b)$ is the generated image. The gradient domain, similarly, $\hat{X}_g = G_{\theta_{gb}}(X_b)$ is the reconstructed gradient map.

4.3 Training Algorithm

The general training procedure aims to minimize the generator's loss L_G and the discriminator's loss L_d. To optimize the G and D, Adaptive Momentum (Adam) [43] is adopted with parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\in = 10^{-8}$. Algorithm 1 shows the specific process of Adam's algorithm to update the internal parameters of G and D.

Algorithm 1. Adam-based training algorithm

Require: t: updated steps
ε: learning rate
Require: $f_1(\theta_g)$: Equation 10; $f_2(\theta_d)$: Equation 18
Require: θ_0: initialized parameter vector
θ_{g_t} : G's parameter at t step
θ_{d_t} : D's parameter at t step
$t \leftarrow 0$

Input: training_set $\left\{X_b^i, X_t^i\right\}_{i=1}^{N}$,
hyperparameters:
$\beta_1 = 0.9,\ \beta_2 = 0.999,\ \varepsilon = 10^{-8}$

while θ_{g_t}, θ_{d_t} not converged **do**

step 1: sample batch $\left\{X_b^i, X_t^i\right\}_{i=1}^{M}$

step 2: calculate the gradient of G and D

$$g_t \leftarrow \nabla_{\theta_g} f_1(\theta_{g_{t-1}})$$

$$d_t \leftarrow \nabla_{\theta_d} f_2(\theta_{d_{t-1}})$$

step 3: calculate step 2's first order moments

$$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$$

$$p_t \leftarrow \beta_1 \cdot p_{t-1} + (1 - \beta_1) \cdot d_t$$

step 4: calculate step2's second order moments

$$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g^2{}_t$$

$$q_t \leftarrow \beta_2 \cdot q_{t-1} + (1 - \beta_2) \cdot d^2{}_t$$

step 5: calculate step 2's offset correction

$$\hat{m}_t \leftarrow m_t / (1 - \beta^t{}_1) p_t \leftarrow p_t / (1 - \beta^t{}_1)$$

$$\hat{v}_t \leftarrow v_t / (1 - \beta^t{}_2) v_t \leftarrow v_t / (1 - \beta^t{}_2)$$

step 6: update parameters

$$\theta_{g_t} \leftarrow \theta_{g_{t-1}} - \varepsilon \cdot \hat{m}_t / (\sqrt{v_t} + \beta^t{}_1)$$

$$\theta_{d_t} \leftarrow \theta_{d_{t-1}} - \varepsilon \cdot \hat{p}_t / (\sqrt{q_t} + \beta^t{}_1$$

Return $\theta_{g_t}, \theta_{d_t}$

5 Experiments

In this section, the experiment environment configuration is presented first, followed by research tests and result analyses to two different datasets.

5.1 Implementation Details

The basic experimental parameters are set as follows: batchsize 16, initial learning rate of 0.0001 for G and 0.0004 for D. The learning rate is reduced to 0 after 2000 epochs applying a polynomial decay method. Since SN involves recurrent connections, gradient clipping is employed to prevent the gradient disappearance, setting the minimum value to 1e-5 and the maximum value to 2. Table 2 shows the experimental environment configuration.

After repeated experiments, an order of magnitude balance is achieved between the values of the different loss

terms when the weighting factor takes the value: $\lambda = 0.05$, $\phi_1 = 0.5$, $\phi_2 = 1$, $\alpha = 0.01$, $\beta_1 = 0.01$, $\beta_2 = 1$.

Table 2. Experimental environment configuration

Hardware environment	Single NVIDIA 3090Ti GPU Ubuntu 21.10 video memory 24G
Software environment	TensorFlow 2.4 Python 3.6 Opencv-python 3.3.0.10 Numpy 1.16.0

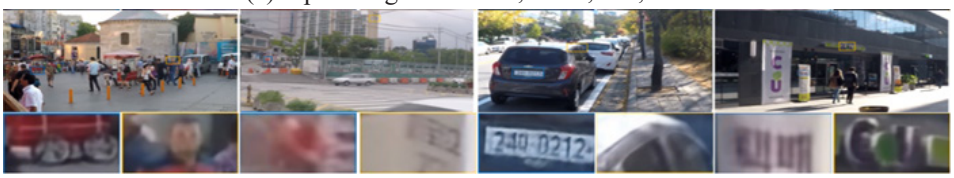
5.2 GoPro Dataset

GoPro dataset is used to train the model, which contains 3214 pairs of blurred and sharp images. Of these, 2103 pairs are used for training and 1111 pairs are used to evaluate image performance, with all images having a resolution of 1280×720. In order to meet the network's input requirement, RGB images are randomly cropped to 256×256 for training and testing. The input and output resolutions at each scale are 64×64, 128×128, 256×256. MADN is first compared with other currently used deblurring algorithms in both quantitative and qualitative terms. Then ablation experiments are conducted to verify the effectiveness of the designed loss function based on GoPro dataset.

5.2.1 Qualitative Analysis

Since our model deals with general camera shake and object motion, it is reasonable to compare with traditional non-uniform deblurring methods. The method of Whyte et al. [44] is selected as a representative traditional non-uniform method. So, MADN is first compared with it. As is seen from the images, Whyte's model fails to resolve the blurring, even worse than the input. This is because the traditional model is solely based on the sparse edge characteristics of natural images combined with a super Laplacian prior modeling. It does not accurately model the real blur kernel, generating a model that is not suitable for dynamic deblurring tasks. Sun's method [19] first applies CNN to image deblurring. Nah [20] first proposes a multi-scale framework to remove blur and the effect is good. MADN's main framework is a multiscale framework based on convolutional neural networks, so the above algorithms are selected for comparison. Sun's model is trained on a synthetic training set with a slightly clearer image over the input. Results are somewhat better in Nah's method, but minute blurring still exists in edge outlines. Benefiting from a well-designed structure and a suitable loss function, results in our model are closer to a real sharp one with good texture details for visual perception. The ingenious use of gradient features makes our algorithm perform better in recovering digital text patterns. A comparison of the different methods is shown in Figure 6.

(a) Input images: Crowds, Road, Car, Store



(b) Whyte's [44]

(c) Sun's [19]

(d) Nah's [20]

(e) Ours

Figure 6. Experimental results for each algorithm on the GoPro dataset

5.2.2 Quantitative Analysis

Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [45] are taken as algorithm performance measures in this research. Based on the four sample images in Figure 6, metric values are computed for each method separately, as shown in Table 3. The histograms in Figure 7 and Figure 8 more intuitively reflect the proposed algorithm's effectiveness. Then the average metric values are measured on GoPro's test set, as shown in the Table 4. Sun's MRF CNN model slightly improves on the SSIM metric compared to Whyte's traditional approach [44]. Compared to the previous two, Nah's Multi-scale CNN has a significant enhancement in metric values, as it uses a multi-scale approach to refine the image features layer by layer. However, Nah's model suffers from too large parameters and a high training time cost. With fewer parameters than in Nah's model, our research method improves on PSNR and SSIM by 1.1dB and 0.0223 respectively.

Table 3. PSNRs and SSIMs of the results in Figure 6

Methods	Evaluation	Crowds	Road	Car	Store	Average
Whyte's [44]	PSNR (dB)	27.296	21.560	25.419	25.913	25.047
	SSIM	0.8924	0.6595	0.7067	0.8781	0.7842
Sun's [19]	PSNR (dB)	27.388	21.718	25.678	26.442	25.307
	SSIM	0.9030	0.6678	0.7414	0.8727	0.7962
Nah's [20]	PSNR (dB)	30.438	22.960	27.717	28.247	27.341
	SSIM	0.9041	0.7869	0.7526	0.8887	0.8331
Ours	PSNR (dB)	31.498	24.361	28.780	29.972	**28.653**
	SSIM	0.9052	0.8170	0.8542	0.8897	**0.8665**

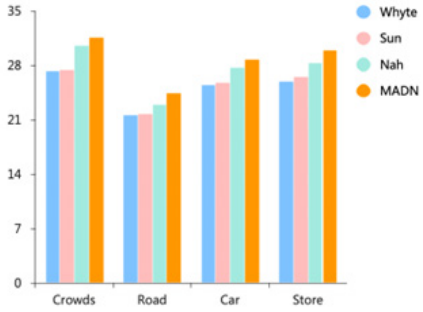


Figure 7. PSNR of different methods

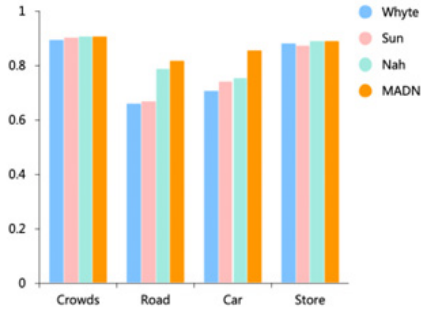


Figure 8. SSIM of different methods

Table 4. GoPro dataset evaluation results

Method	PSNR (dB)	SSIM
Whyte's [44]	24.53	0.8458
Sun's [19]	24.64	0.8429
Nah's [20]	29.08	0.9135
Ours	30.18	0.9312

5.2.3 Investigation of Different Loss Function Terms

Comparative experiments are carried out in this section, using different combinations of loss function terms (Figure 9). SN is able to complete basic deblurring with pixel loss and content loss, but the deblurred results are not sufficiently detailed and even distorted, as shown in sample picture 2, where color distortion emerges in the background buildings. Adding the adversial loss improves the the deblurred results in detail and solves the chromatic aberration. Both the content loss and the adversial loss are restricted in the gradient domain, and the results have richer details and clearer edge contours, which are in line with human eye observation characteristics. In summary, the loss function designed in this research outperforms on the deblurring task.

5.3 Lai Dataset

Lai dataset [46] is composed of real-world blur images, completely representing the real-world motion blur complexity. Test results on face2 in Lai dataset can only be observed visually, for its lack of a corresponding sharp image as a comparison. In order to meet the network's requirement, RGB images are also randomly cropped to 256×256 for testing. Our method noticeably excels other ones, especially in the eyes and necklace areas (Figure 10). The girl's eyes restored by our algorithm are clearer, and the edges of the necklace are sharper than other methods without distortion and artifact.

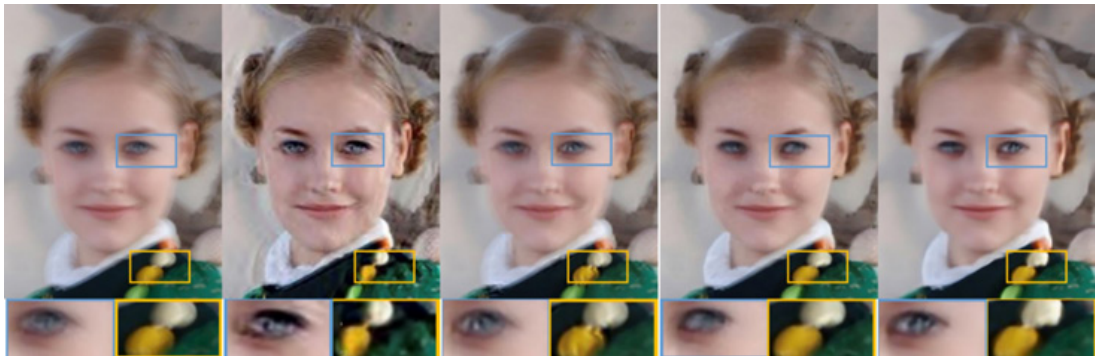
6 Conclusion

This research proposes a multi-scale adversarial deblurring network with gradient guidance (MADN) algorithm to address the lack of sharp edges and unsatisfactory detail recovery in motion deblurred images. The generator in the MADN consists of a scale network SN and a gradient feature extraction network GN. By utilizing image features of different scales, MADN reduces the deblurred image artifacts and

preserves the edge structures. The SN intermediate features are connected to the GN encoding network for guiding the gradient image reconstruction, recovering high frequency information such as details. Experimental results on GoPro and Lai dataset show that this proposed model has a powerful deblurring capability, and its visual quality and evaluation metrics are significantly better than those of other current learning-based deblurring models. However, MADN's parameters have reached more than 29 million, making it difficult to train. The background area of the blurred image is not recovered well enough.

Our future researches would focus on new algorithm explorations to design a relatively lightweight structure with attention mechanism to recover higher quality images from a global feature extraction perspective.

Figure 9. Results of different loss function terms



Blurry input Whyte's [44] Sun's [19] Nah's [20] Ours

Figure 10. Experimental results from each algorithm on Lai dataset

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government. (MSIT) (2021R1A2B5B02087169).

References

[1] L. Zhang, Z. Wang, L. Wang, Z. Zhang, X. Chen, L. Meng, Machine Learning-Based Real-time Visible Fatigue Crack Growth Detection, *Digital Communications and Networks*, Vol. 7, No. 4, pp. 551-558, November, 2021.

[2] Y.-T. Chang, W. K. T. M. Gunarathne, T.-K Shih, Deep Learning Approaches for Dynamic Object Understanding and Defect Detection, *Journal of Internet Technology*, Vol. 21, No. 3, pp. 783-790, May, 2020.

[3] S. Cho, S. Lee, Fast Motion Deblurring, *ACM Transactions on Graphics*, Vol. 28, No. 5, pp. 89-97, December, 2009.

[4] A. Goldstein, R. Fattal, Blur-kernel Estimation from Spectral Irregularities, *European Conference on Computer Vision*, Florence, Italy, 2012, pp. 622-635.

[5] Q.- T. Nguyen, J.-X. Sun, Y. Sun, H.-X. Liu, L.-R. Zhao, G.-W. Liu, Disposing of Outliers in Camera-Shake Blurred Images Restoration, *Journal of Image and Graphics*, Vol. 19, No. 5, pp. 677-682, May, 2014.

[6] A. Chakrabarti, A Neural Approach to Blind Motion Deblurring, *European Conference on Computer Vision*, Amsterdam, The Netherlands, 2016, pp. 221-235.

[7] R.-W. Liu, Y. Wei, S. Xiong, S. Peng, Lo-Regularized Hybrid Gradient Sparsity Priors for Robust Single-Image Blind Deblurring, *IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, Canada, 2018, pp. 1348-1352.

[8] Z. Dou, K. Gao, X. Zhang, H. Wang, Fast Blind Image Deblurring Using Smoothing-Enhancing Regularizer, *IEEE Access*, Vol. 7, pp. 90904-90915, July, 2019.

[9] J. Kotera, F. Šroubek, P. Milanfar, Blind Deconvolution Using Alternating Maximum A Posteriori Estimation with Heavy-Tailed Priors, *Computer Analysis of Images and Patterns*, York, UK, 2013, pp. 59-66.

[10] K. Manning, X. Zhai, W. Yu, Image Analysis and Machine Learning-Based Malaria Assessment System, *Digital Communications and Networks*, Vol. 8, No. 2, pp. 132-142, April, 2022.

[11] J. Li, K. Li, B. Yan, Scale-aware Deep Network with Hole Convolution for Blind Motion Deblurring, *IEEE International Conference on Multimedia and Expo* (ICME), Shanghai, China, 2019, pp. 658-663.

[12] K.-H. Liu, C.-H. Yeh, J. W. Chung, C.-Y. Chang, A Motion Deblur Method based on Multi-Scale High Frequency Residual Image Learning, *IEEE Access*, Vol. 8, pp. 66025-66036, April, 2020.

[13] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative Adversarial Networks, *Conference and Workshop on Neural Information Processing Systems (NIPS)*, Montreal, Canada, 2014, pp. 2672-2680.

[14] K. Ning, Z. Su, Z. Zhang, G.-J. Kim, An Image Reconstruction Algorithm Based on Frequency Domain for Deep Subcooling of Melt Drops, *Journal of Internet Technology*, Vol. 22, No. 6, pp. 1273-1285, November, 2021.

[15] Y. Zhang, Z. Yang, J. Hu, S. Zou, Y. Fu, MRI Denoising Using Low Rank Prior and Sparse Gradient Prior, *IEEE Access*, Vol. 7, pp. 45858-45865, April, 2019.

[16] C. Li, C. Liu, Z. Liu, R. Yang, Y. Huang, Fabric Defect Detection Method based on Cascaded Low-Rank Decomposition, *International Journal of Clothing Science and Technology*, Vol. 32, No. 4, pp. 483-498, July, 2020.

[17] J. Ye, S. Xue, A. Jiang, Attention-based Spatio-temporal Graph Convolutional Network Considering External Factors for Multi-step Traffic Flow Prediction, *Digital Communications and* Networks, Vol. 8, No. 3, pp. 343-350, June, 2022.

[18] K. Yan, X. Zhou, Chiller Faults Detection and Diagnosis with Sensor Network and Adaptive 1d CNN, *Digital Communications and Networks*, Vol. 8, No. 4, pp. 531-539, August, 2022.

[19] J. Sun, W. Cao, Z. Xu, J. Ponce, Learning A Convolutional Neural Network for Non-Uniform Motion Blur Removal, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 769-777.

[20] S. Nah, T.-H. Kim, K.-M. Lee, Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 257-265.

[21] X. Tao, H. Gao, X. Shen, J. Wang, J. Jia, Scale-Recurrent Network for Deep Image Deblurring, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 8174-8182.

[22] H. Zhang, Y. Dai, H. Li, P. Koniusz, Deep Stacked Hierarchical Multi-Patch Network for Image Deblurring, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 5971-5979.

[23] A. Radford, L. Metz, S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, January, 2015, https://arxiv.org/abs/1511.06434.

[24] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, J. Matas, DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 8183-8192.

[25] O. Kupyn, T. Martyniuk, J. Wu, Z. Wang, DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better, *IEEE/CVF International Conference on Computer Vision (CVPR)*, Seoul, Korea (South), 2019, pp. 8877-8886.

[26] G. Gong, K. Zhang, Local Blurred Natural Image Restoration based on Self-Reference Deblurring Generative Adversarial Networks, *IEEE International Conference on Signal and Image Processing*

Applications (ICSIPA), Kuala Lumpur, Malaysia, 2019, pp. 231-235.
[27] L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, M.-H. Yang, Blind Image Deblurring via Deep Discriminative Priors, *International Journal of Computer Vision*, Vol. 127, No. 8, pp. 1025-1043, August, 2019.
[28] X. Yu, W. Xie, Blur Kernel Estimation Method based on Deep Encoder-Decoder Network, *Control Theory & Applications*, Vol. 37, No. 4, pp.731-738, May, 2020.
[29] C. Chen, J. Xu, C. Wei, Q. Li, Multi-scale Image Blind Deblurring based on Salient Intensity and A Priori Gradient, *Laser & Optoelectronics Progress,* Vol. 57, No. 4, pp. 1-7, February, 2020.
[30] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, L.-V. Gool, Night-to-Day Image Translation for Retrieval-based Localization, *International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada, 2019, pp. 5958-5964.
[31] F. Luan, S. Paris, E. Shechtman, K. Bala, Deep Photo Style Transfer, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 6997-7005.
[32] W. Yang, J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, S. Yan, Deep Edge Guided Recurrent Residual Learning for Image Super-Resolution, *IEEE Transactions on Image Processing*, Vol. 26, No. 12, pp. 5895-5907, December, 2017.
[33] C. Ma, Y. Rao, Y. Cheng, C. Chen, J. Lu, J. Zhou, Structure-Preserving Super Resolution with Gradient Guidance, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Online, pp. 7766-7775.
[34] J. Zhu, L. Meng, W. Wu, D. Choi, J. Ni, Generative Adversarial Network-based Atmospheric Scattering Model for Image Dehazing, *Digital Communications and Networks*, Vol. 7, No. 2, pp. 178-186, May, 2021.
[35] S. Viswanath, M. Ghulyani, S.-D. Beco, M. Dahan, M. Arigovindan, Image Restoration by Combined Order Regularization with Optimal Spatial Adaptation, *IEEE Transactions on Image Processing*, Vol. 29, pp. 6315-6329, April, 2020.
[36] N. Yair, T. Michaeli, Multi-scale Weighted Nuclear Norm Image Restoration, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 3165-3174.
[37] Y. Du, A Crop Image Segmentation and Extraction Algorithm based on Mask RCNN. *Entropy*, Vol. 23, No. 9, pp.1160, September, 2021.
[38] X. Mao, C. Shen, Y.-B. Yang, Image Restoration Using Very Deep Convolutional Encoder-decoder networks with Symmetric Skip Connections, *Conference and Workshop on Neural Information Processing Systems (NIPS)*, Barcelona, Spain, 2016, pp. 2810-2818.
[39] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional Networks for Biomedical Image Segmentation, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, 2015, pp. 234-241.
[40] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, 2016, pp. 770-778.
[41] H. Prasetyo, A. W. H. Prayuda, C.-H. Hsia, M. A. Wisnu, Integrating Companding and Deep Learning on Bandwidth-Limited Image Transmission, *Journal of Internet Technology*, Vol. 23 No. 3, pp. 467-473, May, 2022.
[42] B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, Enhanced Deep Residual Networks for Single Image Super-Resolution, *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, 2017, pp. 1132-1140.
[43] K. S. Kim, Y. S. Choi, HyAdamC: A New Adam-Based Hybrid Optimization Algorithm for Convolution Neural Networks, *Sensors*, Vol. 21, No. 12, Article No. 4054, June, 2021.
[44] O. Whyte, J. Sivic, A. Zisserman, J. Ponce, Non-uniform Deblurring for Shaken Images, *International Journal of Computer Vision*, Vol. 98, No. 2, pp. 168-186, June, 2012.
[45] A. Horé, D. Ziou, Image Quality Metrics: PSNR vs. SSIM, *International Conference on Pattern Recognition (ICPR)*, Istanbul, Turkey, 2010, pp. 2366-2369.
[46] W.-S. Lai, J.-B. Huang, Z. Hu, N. Ahuja, M.-H. Yang, A Comparative Study for Single Image Blind Deblurring, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, 2016, pp. 1701-1709.

Biographies

Jinxiu Zhu is a professor and master's supervisor of Hohai University. Her main research interests include neural network, edge/fog computing, intelligent image processing, etc.

Xue Xu is a master at the College ofIoT Engineering, Hohai University, China. Her research interests include artificial intelligence, deep learning-based image restoration algorithm.

Chang Choi is Assistant Professor with Gachon University, Rep. Of Koreaa. His research interests include intelligent information processing, semantic web, the smart IoT systems, and intelligent system security.

Xin Su is a professor at the College of IOT Engineering, Hohai University, China. He recently focuses the research topics on mobile communication, 5/6G systems, edge computing/fog computing, IOT applications, and smart ocean.