# Multitask Learning Mechanism for Remote Sensing Image Motion Deblurring

Jie Fang, Xiaoqian Cao, Dianwei Wang, and Shengjun Xu

***Abstract*—As a fundamental preprocessing technique, remote sensing image motion deblurring is important for visual understanding tasks. Most conventional approaches formulate the image motion deblurring task as a kernel estimation. Because the kernel estimation is a highly ill-posed problem, many priors have been applied to model the images and kernels. Even though these methods have obtained relatively better performances, they are usually time-consuming and not robust for different conditions. To address this problem, we propose a multitask learning mechanism for remote sensing image motion deblurring in this article, which contains an image restoration subtask and an image texture complexity recognition one. First, we consider the image motion deblurring problem as a domain transformation problem, from the blurred domain to a clear one. Specifically, the blurred domain represents the data space consisted of blurring images, and the definition of clear domain is similar. Second, we design a novel weighted attention map loss to enhance the reconstruction capability of the restoration subbranch for difficult local regions. Third, based on the restoration subbranch, a recognition subbranch is incorporated into the framework to guide the deblurring process, which provides the auxiliary texture complexity information to help the optimization of restoration subbranch. Additionally, in order to optimize the proposed network, we construct three large-scale datasets, and each sample in the dataset contains a clear image, a blurred image, and its texture label obtained by corresponding texture complexity. Finally, the experimental results on three constructed datasets demonstrate the robustness and the effectiveness of the proposed method.**

***Index Terms*—Domain transformation, image deblurring, multitask learning mechanism.**

Manuscript received November 10, 2020; revised December 18, 2020; accepted December 23, 2020. Date of publication December 28, 2020; date of current version February 5, 2021. This work was supported in part by the Fund of Shaanxi Key Laboratory of Intelligent Processing for Big Energy Data, Yanan University under Grant IPBED6, in part by the Special Project of Strengthening Police with Science and Technology of Ministry of Public Security under Grant 2019GABJC42, in part by the Special Scientific Research Project of Education Department of Shaanxi Province under Grant 19JK0140, and in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2020JM472, Grant 2020JM473, and Grant 2019JQ760. *(Corresponding author: Jie Fang.)*

Jie Fang and Dianwei Wang are with the Center for Image and Information Processing, School of Communications and Information Engineering & School of Artificial Intelligence, Xi'an University of Posts & Telecommunications, Xi'an 710121, P. R. China (e-mail: 2443952262@qq.com; wangdianwei@xupt.edu.cn).

Xiaoqian Cao is with the School of Electric and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, P. R. China (e-mail: caoxiaoqian@sust.edu.cn).

Shengjun Xu is with the School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, P. R. China, and also with the AI & DE, Guangzhou 510320, P. R. China (e-mail: duplin@sina.com).

Digital Object Identifier 10.1109/JSTARS.2020.3047636

## I. Introduction

With the rapid development of satellite imaging technology, remote sensing imagery interpretation has attracted much attention during the last decade since its vital scientific research value and wide range of practical applications. Actually, motion blur of images can be caused by several factors such as camera shakes, object motions, and depth variations [1]–[4], so its inverse restoration process is of difficulty. Remote sensing image motion deblurring aims to restore sharp image from a blurred one, which is an important preprocessing technique for many tasks, and some examples are shown in Fig. 1. Through an in-depth investigation, we found that most existing methods are based on the conventional blur model [5], which is defined as follows:

$$\mathbf{I}_b = \mathbf{K}_b \mathbf{I}_s + \mathbf{N} \tag{1}$$

where $\mathbf{I}_b$ denotes the blurred image, $\mathbf{K}_b$ denotes the blur kernel, $\mathbf{I}_s$ denotes the latent sharp image, and $\mathbf{N}$ denotes the uncontrolled noise. Our goal is to exploit an available method to get the best latent sharp image $\mathbf{I}_s$. According to (1), the most difficult problem of image deblurring is the blur kernel estimation, which is highly relevant to the quality of the restored image. In practice, both blur kernel and latent sharp image are unknown. As a result, blind deblurring approaches need to infer blur kernel $\mathbf{K}_b$ and latent sharp image $\mathbf{I}_s$ simultaneously.

At the beginning, image deblurring methods can be divided into two main branches, one is based on variational Bayesian inference, and the other is based on maximum *a posteriori* (MAP) estimation. Fergus *et al.* [6] proposed an approach using a mixture of Gaussians to learn an image gradient prior through variational Bayesian inference. Levin *et al.* [7] analyzed the method based on variational Bayesian inference further. However, the optimization process of variational Bayesian inference is computationally expensive. Afterwards, MAP equipped with different likelihood functions and image priors [8], is widely used for image deblurring task.

Recently, benefited by the rapid progress of computational load and storage resource, convolutional neural networks (CNNs) have achieved competitive performances on many computer vision tasks, such as semantic segmentation [9] and object detection [10]. As for image motion blurring task, Chakrabarti [11] proposed to use a network to estimate the sharp image from blurred one by an unknown motion kernel. Chen *et al.* [12] proposed to use the popular generative adversarial network (GAN) equipped with inceptionresdensenet

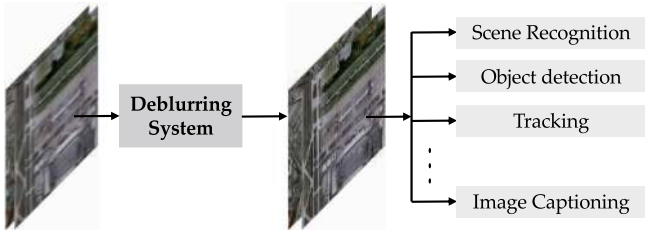


Fig. 1. Remote sensing image deblurring is a fundamental preprocessing technique, which can improve the performances of several image understanding tasks such as scene recognition, object detection, tracking, and image captioning.

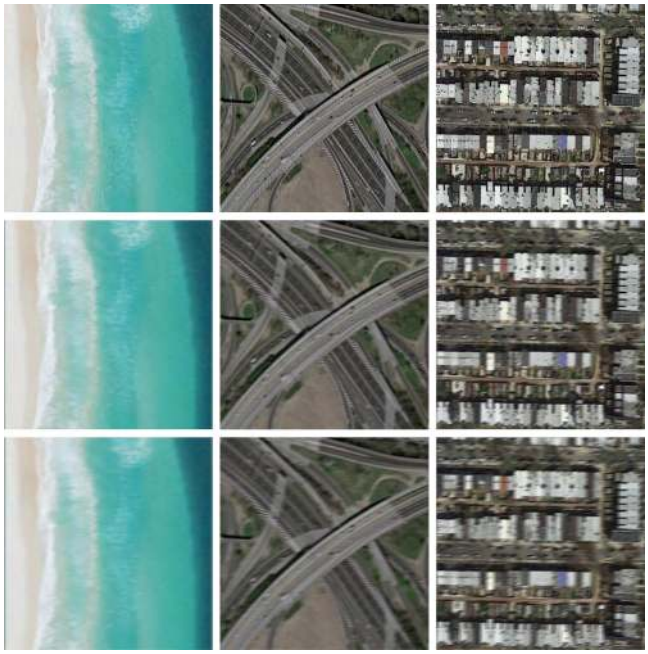


Fig. 2. Visualized results when apply different blur kernels to images with different texture complexities. Specifically, the texture complexity level of image is gradually increased from left to right, and the blur factor applied on the third row is more severe than that of the second row. In addition, the D-score represents the sharpness differences between images in the second row and corresponding ones in the third row, it can be calculated with Brenner gradient function and Vollath function. From which we can find that, images with higher complexity texture are more sensitive to the blur factor especially the severe ones.

strategy to obtain more accurate results. Even though these CNN-based methods have achieved relatively satisfactory performances compared to traditional machine learning-based ones, they often ignore the effect of image complexity on deblurring performances. This negligence may lead to the significantly different deblurring performances on images with different texture complexities, specifically, well performance on simple images but poor performance on complex ones [13]. The reason is that, mapping relationships of paired images with simple textures are much easier to learn, compared to ones with complex textures [14]. In addition, pixel-wise norm loss can not explicitly depict this point effectively [12], and hence, the trained model tend to fit well for simple samples. Actually, when the same motion kernel act on images with different textures complexities, the degradations are different. As shown in Fig. 2, images with more complicated texture structures are more sensitive for blur factors, even if these factors are the same. In addition, the increment of the blur factor leads to more severe influence to the images with complex textures, and we use the differential score of texture information loss under blur factors with different degrees (D-score) to demonstrate this point, which is defined as follows:

$$D-\text{Score}=\frac{\text{Bre}(\mathbf{I}_{b1})-\text{Bre}(\mathbf{I}_{b2})}{2\text{Bre}(\mathbf{I}_{c})}+\frac{\text{Vol}(\mathbf{I}_{b1})-\text{Vol}(\mathbf{I}_{b2})}{2\text{Vol}(\mathbf{I}_{c})}, \tag{2}$$

where $\mathbf{I}_c$ denotes the clear image. $\mathbf{I}_{b1}$ and $\mathbf{I}_{b2}$ denote the corresponding slight blurred and severe blurred image, respectively. $\text{Bre}(\cdot)$ and $\text{Vol}(\cdot)$ denote two conventional image quality assessment functions, Brenner gradient and Vollath, and they are formulated by (3) and (4), respectively

$$Bre(\mathbf{I})=\sum_{w=1}^{W}\sum_{h=1}^{H}\left|\mathbf{I}_{w+2,h}-\mathbf{I}_{w,h}\right|^2 \tag{3}$$

$$Vol(\mathbf{I})=\sum_{w=1}^{W}\sum_{h=1}^{H}\mathbf{I}_{w,h}\bigodot\mathbf{I}_{w+1,h}-\mu^2$$

$$\text{s.t. }\ \mu=\frac{1}{WH}\sum_{w=1}^{W}\sum_{h=1}^{H}\mathbf{I}_{w,h} \tag{4}$$

where $W$ and $H$ denote the width and height of the image $\mathbf{I}$, respectively, and $\bigodot$ represents the element-wise multiplication.

In particular, we define the texture structure complexity of image according to the richness of its edge information. Particularly, texture complexities of images with more high-contrast edges are increasingly higher than those with fewer ones. Abstractly, image motion blurring is a filtering process [15], which is to represent each pixel by all ones in its neighbourhood. Based on this point, the differences of high-contrast pixels or regions in clear images and corresponding blurred ones are much significant than those of flat regions. In other words, images with high texture complexities are influenced more severely by motion blur kernels, and the conventional models without considering the texture complexities can not fit well for these samples. In these cases, incorporate the texture complexity information into the existing method as a prior seems like a reasonable choice to improve their deblurring capability for complex images.

According to the aforementioned analysis, we propose a multitask learning-based mechanism for remote sensing image motion deblurring, which contains an image restoration subtask and an image texture complexity recognition subtask. The proposed mechanism considers the motion blur kernels and texture complexity of image simultaneously. Specifically, the first subtask considers the remote sensing image deblurring problem as a domain transformation one, and restore sharp images from blurred ones through a U-Net-based architecture. Additionally, based on the mse loss, we design a novel weighted attention map loss term to enhance the reconstruction capability of the image restoration subbranch for difficult local regions. The second subtask considers the relationships of image texture complexity and debluring complexity, which provides the auxiliary information to guide the optimization of the first subtask. Finally, to optimize

the proposed network, we build three large-scale datasets for remote sensing image debluring, each sample in the dataset contains a clear image, a blurred image, and its corresponding texture complexity label. Summarily, the contributions of this article can be listed as follows.

1) We propose a multitask learning mechanism for image deblurring, which contains a texture complexity classification subtask and an image restoration subtask.
2) We find the relationship of blur kernel sensitivity and texture complexity for different images, more complicated more sensitive.
3) We design a weighted attention map loss for the restoration subbranch, which enhances the reconstruction capability for difficult local regions.
4) We construct three large scale datasets for remote sensing image motion deblurring task, each sample in the dataset contains a clear image, a blurred image, and its texture label.

The remainder of this article is organized as follows. Section II reviews some related works about image deblurring. Section III details the proposed method. Section IV provides the dataset. Section V reports and analyzes the experimental results. Section VI concludes the article.

## II. Related Works

Image debluring is an important fundamental problem in visual understanding field, especially for mobile imaging process equipments [16]–[19]. During the imaging process, because of the air noises, camera shakes, and other unpredicted influence factors, photographed pictures might be blurred in different extents [20]–[22]. For humans, the quality-dropped images are not a feast for the eyes, but humans still can analyze their contents, due to experimental knowledge, such as imagination, reasoning ability [23]–[25], etc. However, for machines or computers, all knowedge of the model are learned from the very training dataset [26]. In this case, the interferences of the training data are severe for the performance of the models. For instance, models trained with blurred images are not sensitive for clear images and the performance indicators must be decreased.

According to aforementioned reasons, image deblurring plays a vital role in imagery understanding field. During the last decade, many comparative approaches were proposed for this task, mainly including two subbranches, nonblind image deblurring and blind image deblurring. Nonblind image deblurring [27] has been studied for many years, which aims to restore latent sharp image from blurred one with known blur kernels. Nonblind image deblurring task is an ill-posed problem because the inverse is unstable to interference and noise, even a small amount of noise may lead to severe distortions. Classic methods to tackle this problem mainly consist of Wiener filter [28] and Richardson–Lucy algorithm [29], [30]. Nonblind image is considered as an optimization problem. Many classical priors, such as sparse priors [31], [32], have been incorporated for regularization. Blind image deblurring [33], [34] aims to restore latent sharp image from blurred one without knowledge of blur kernels. Traditional gradient-based priors of natural images tend to fail [34] because they often favor blurry images with most low frequencies in the Fourier domain. With the development of regularization and optimization, some sophisticated priors have also been used to address this problem. For instance, $l_0$-norm-based prior [8], [35], framelet-based prior [36], dark channel prior [37], sparse coding-based prior [38], and low rank prior [39]. Even though these methods have achieved obvious improvements for image deblurring task, each one has its own limits. $l_0$-norm-based prior is nonconvex, and its convex relaxation, $l_1$ norm, is parameter-sensitive and thus not robust [40]. Framelet-based prior depends on hand-crafted wavelet functions, since they are not general enough to tackle real-world blurred images [41]. Sparse coding prior is based on an assumption that, the statistical distributions between training set and target set are similar, but this may not be real in practice [41]. Additionally, low-rank prior and dark channel prior often suffer from high computation complexity [40].

Recently, deep neural networks especially CNNs have achieved significant performances in many tasks such as image recognition [42], [43], object detection [44], semantic segmentation [45], etc. As for image deblurring, it is actually an image projection task, from blurred domain to clear one. Generative adversarial networks (GAN) [46], [47] especially conditional generative adversarial network (cGAN) have achieved significant performances on image transformation, due to the discriminative mechanism being incorporated into the generative models. Specifically, pix2pix GAN [46] and ResGAN [47] have achieved satisfactory performances on image transformation task. Based on the aforementioned superiorities of GAN, Ramakrishnan *et al.* proposes deep generative filter (DGF) [48] for motion deblurring. DGF integrates global skip connection and dense architecture into the network to tackle the ill-posed nature problem in this challenging task. Even though GAN-based methods surpasses the most existing ones, its training process is not stable and the generative ability is usually limited. Additionally, remote sensing images have abundant texture information, which makes it difficult to train a robust GAN model. In these cases, we propose weighted attention map loss, the replacement of the adversarial loss, to encode the detailed information of the image into the representation and obtain more clear deblurred images.

## III. Proposed Method

### A. Overview

As is shown in Fig. 3, the proposed remote sensing image deblurring approach mainly contains of two components: 1) U-Net-based image restoration sub-branch; and 2) multilayer perception-based image texture complexity recognition subbranch. For 1), we consider the image deblurring task as an image-to-image transformation one, from blurred domain to deblurred domain. Specifically, the deblurring subbranch is based on U-Net, which can propagate sufficient spatial details from shallow layers to deep ones through the skip connection strategy. For 2), we find the texture complexity is an important factor for image deblurring, and incorporate image texture complexity recognition into the deblurring framework. Specifically, a conventional classification network is utilized to

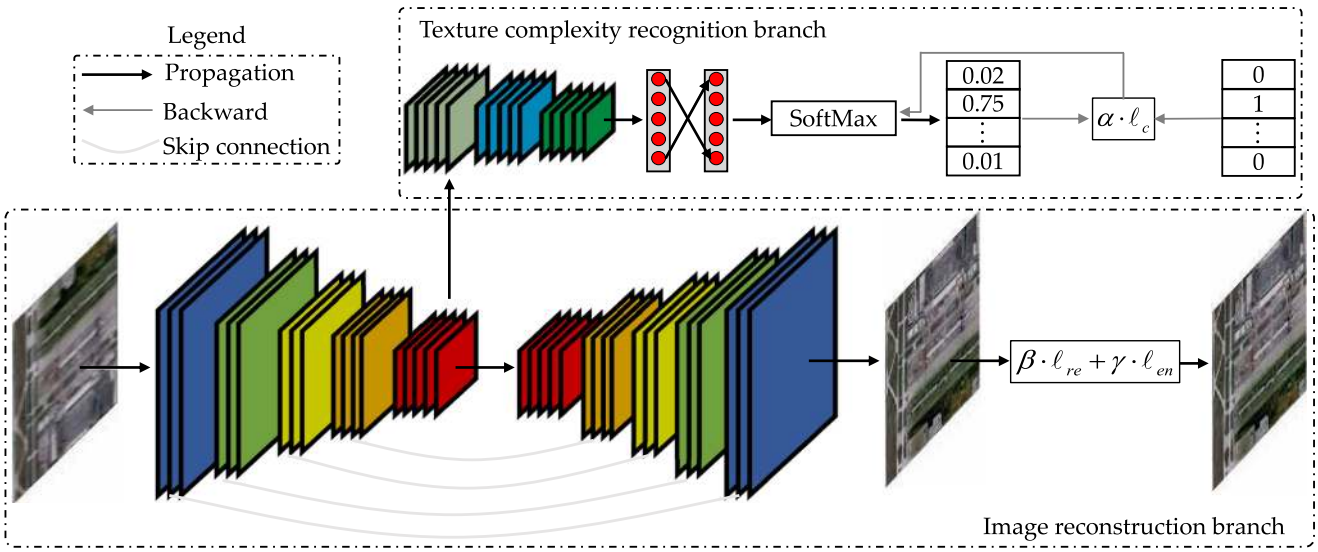


Fig. 3. Multitask learning mechanism for image quality improvement, which mainly contains two important components, reconstruction branch, and texture complexity recognition branch. Specifically, the reconstruction is used to finalize the projection of blur domain and clear domain, and the texture complexity recognition branch is to guide the optimization of the reconstruction, which gives images different attention according to theirs texture information. Besides, $\ell_{re}$ is the classical reconstruction loss. $\ell_{en}$ is the proposed weighted attention map loss, which can assist the system to improve its sensitivity for the difficult regions in the image. $\ell_c$ is the classification-based texture complexity recognition branch, which can provide auxiliary global priors to the reconstruction branch. In addition, $\alpha$, $\beta$, and $\gamma$ are three hyperparameters to balance the relative importance of the three loss terms.

recognize the texture complexity of the input image, which takes the output of the third convolution layer of image restoration subbranch as its input.

### B. Network Architecture

This subsection details the network architecture of the proposed method, which mainly contains two components: Image deblurring network and image texture complexity recognition network. Specifically, the former is based on U-Net, a deconvolution network with satisfactory performance on structured prediction task, such as boundary detection and semantic segmentation, etc. The latter is a multilayer perception network, which realizes the secondary feature coding and image texture complexity recognition simultaneously.

*1) Image Deblurring Network Architecture:* Because of the multilayer nonlinear mappings and weights shared strategy, CNNs can obtain more discriminative representations than handcrafted feature descriptors [49], [50], and further achieve satisfactory performances on many recognition tasks in computer vision field. However, large-receptive-field convolution kernels and pooling operations often lead to the fragmented outputs of CNNs, which limits their applications in structured prediction tasks.

Different from simple traditional recognition task, image-to-image transformation pay more attention to detailed texture information but not the discriminative ones, especially for the image deblurring task. In these cases, a structure-preserving network is very important for images deblurring, which must preserve sufficient detailed texture information from the original image to predicted image. Recently, U-Net have shown satisfactory performance in this point because of the skip connection mechanism [51], [52], and hence we use it as the backbone network of our restoration branch.

*2) Image Texture Complexity Recognition Network Architecture:* Besides the image reconstruction branch, a multilayer perception-based classification branch is used to recognize texture complexity level of the image. As is demonstrated in Section I, images with different texture complexity have different sensitivity level when they are interfered by the same motion blur kernel. Specifically, image with complex boundary texture structure and detailed information are more sensitive to the blur interference, compared to image full of flat regions. Because of the aforementioned reasons, we insist that texture complexity level is an important factor that influence the deblurring quality. In these cases, we incorporate a texture complexity level recognition branch into the framework. Through joint learning, feature maps from reconstruction branch are more representative, because the loss function contains image texture complexity explicitly and makes the features more texture-complexity-specific. The inner architecture of contexture complexity recognition branch is shown in Table I.

In this article, we use a small-scale CNN to predict the image texture complexity. Specifically, the input to this recognition network is the middle layer output from the image reconstruction branch. The reason why we use the middle layer output of the image reconstruction layer is that, feature maps with small size often can depict the global information of the image well. Additionally, in order to improve the representation capability of the texture complexity recognition branch, we design a three-layer perception network but not only a classifier to predict the texture information of the image.

TABLE I
TEXTURE COMPLEXITY RECOGNITION BRANCH

| **Layer** | K/S/N | **Input** | Output Size |
|---|---|---|---|
| $conv1$ | $3 \times 3 \times 512$ | $drop5$ | $32 \times 32 \times 512$ |
| $pool1$ | 2 | $conv1$ | $16 \times 16 \times 512$ |
| $conv2$ | $3 \times 3 \times 256$ | $pool1$ | $16 \times 16 \times 256$ |
| $pool2$ | 2 | $conv2$ | $8 \times 8 \times 256$ |
| $conv3$ | $3 \times 3 \times 128$ | $pool2$ | $8 \times 8 \times 128$ |
| $GAP$ | / | $conv3$ | $1 \times 128$ |
| $fc1$ | 256 | $GAP$ | $1 \times 256$ |
| $softmax$ | $N_c$ | $fc1$ | $1 \times N_c$ |

### C. Loss Function

Besides the network architecture, loss function to optimize the model is also important for the proposed method. Specifically, similar to the network architecture, loss function for the model also contains two important components: Reconstruction loss term and texture complexity level recognition loss term.

*1) Reconstruction Loss Term:* Reconstruction branch is mainly to finalize the domain transforming, from blurred image domain to clear image domain. In this case, the Frobenius norm is used as the basic loss to optimize the network, and the function is defined as follows:

$$\ell_{re} = \frac{1}{N}\sum_{n=1}^{N}\left\|\mathbf{I}_{out}^{n} - \mathbf{I}_{gt}^{n}\right\|_{F}^{2} \tag{5}$$

where $N$ is the total number of the samples, $\mathbf{I}_{\text{out}}^{n}$ is the output of the $n$th image, $\mathbf{I}_{gt}^{n}$ is the groundtruth of the $n$th image. Additionally, $\|\cdot\|_F$ represents the conventional Frobenius norm function.

Even though the Frobenius norm can reflect the similarity of two images to a large certain extent, it gives same weights to each local region in the image. In other words, Frobenius norm can not classify the sample complexity well and give them different attentions. However, the reconstruction complexities for different local regions have large differences. For instance, regions with rigid boundary and other complex textures are more sensitive for blur interference, while the flat regions are more robust to the blur kernel. In this case, we propose a weighted attention map loss term to enhance the reconstruction capability of these regions. Specifically, the weighted attention reconstruction loss is defined as follows:

$$\ell_{en} = \frac{1}{N}\sum_{n=1}^{N}\left\|\left|\mathbf{I}_{in}^{n} - \mathbf{I}_{gt}^{n}\right| \bigodot \left|\mathbf{I}_{out}^{n} - \mathbf{I}_{gt}^{n}\right|\right\|_{F}^{2} \tag{6}$$

where $|\mathbf{I}_{in}^{n} - \mathbf{I}_{gt}^{n}|$ is the absolute difference between $n$th input image and corresponding groundtruth, which reflects the interference level of the blur kernel for the $n$th image. The bigger element in the matrix depicts the heavier blur interference, and the regions with heavier blur interference are more difficult to reconstruct. To obtain a robust model for image deblurring task, we use the absolute difference matrix as the weighted attention map to improve the reconstruction capability for difficult regions, some visualized parameter maps of the weights are shown in Fig. 4. This loss term is based on a reasonable and obvious assumption that, if the difficult regions are reconstructed well, other regions can also be reconstructed well.

Fig. 4. Visualized parameter maps of the attention weights. Images in the first column denote the blurred images, and ones in the second and third columns, respectively, denote the corresponding clear images and reconstruction attention weights.

*2) Texture Complexity Level Recognition Loss Term:* Texture complexity level recognition branch is to estimate the texture information of the image, and use it to guide the representation learning of the reconstruction branch. Specifically, we consider the texture complexity information estimation task as a classification one. In order to train this branch, we design a texture measurement approach and label setting rule for the dataset, the details are demonstrated clearly in Section IV. The loss function of texture complexity level recognition branch is a conventional softmax one, which is defined as follows:

$$\ell_{c} = -\frac{1}{N}\left[\sum_{n=1}^{N}\sum_{c=1}^{C}\mathrm{I}\left\{y^{n} = c\right\}\log\frac{e^{\theta_{c}^{T}x^{n}}}{\sum_{l=1}^{C}e^{\theta_{l}^{T}x^{n}}}\right] \tag{7}$$

where $N$ is the samples of the dataset, $C$ is the number of texture complexity levels of the dataset. $\mathrm{I}\{\cdot\}$ is the indictor function, which equals to 1 when the condition in $\{\}$ satisfies and 0 otherwise. $y^n$ is the predicted label of $n$th image, $c$ represents the real label of $n$th image. $\theta$ represents the parameters of the softmax layer, and $x_n$ is the $n$th input to the softmax layer.

In general, the overall loss of the framework contains the aforementioned three components, which is defined as follows:

$$\ell = \alpha \cdot \ell_{re} + \beta \cdot \ell_{en} + \gamma \cdot \ell_{c} + \tag{8}$$

where $\alpha$, $\beta$, and $\gamma$ are three hyperparameters, which are used to balance three loss components. The details of this point are demonstrated clearly in the Section V.

## IV. DATASET

To validate the superiority of our method, we construct three novel texture-label-based motion deblurring datasets. Specifically, we add motion blurs with different levels to three public and challenging remote sensing classification datasets, Sydney [53], UC-Merced [54], and AID [55]. Additionally, we give each image a specific texture label according to its complexity. These two points are detailed in Section IV-A and IV-B, respectively.

### A. Blurred Images

This subsection mainly introduces how to obtain the blurred image according to the clear image. Specifically, we apply different motion blur kernels to the images in the dataset to construct the clear-blurred pairs. To make it more concrete, the linear motions of the blur kernels are randomly distributed in the closed range [5, 25], while the angle degrees are randomly distributed in the closed range [2, 20]. Based on this construction strategy, deblurring complexities of the constructed datasets are dramatically increased because of the motion diversity, and hence the methods applied on them can be validated more radical.

### B. Texture Complexity Labels

To verify the assumption that, texture complexity can influence the deblurring quality to a certain extent, we construct three texture-label-based datasets, SydeneyBlurred dataset, UCMBlurred dataset, and AIDBlurred dataset. These three datasets are based on three classification ones, Sydney, UCM, and AID. The differences mainly contains two aspects, 1) different motion blurs are added to each image in the traditional classification datasets, which constructs a corresponding clear-blurred image recovery ones. 2) We attach a specific label to each image in the dataset according to its texture complexity, and the detailed definitions are shown as follows.

*1) Texture Complexity Calculation:* First of all, a reasonable texture complexity descriptor should be applied. In this article, we utilize the value of adjacent-pixel-difference to obtain the texture complexity of the image, which is defined as follows:

$$T = \frac{1}{HWC}\sum_{i=1}^{W}\sum_{j=1}^{H}\sum_{c=1}^{C}\left(\mathbf{T}_{\text{row}}(i,j,c) + \mathbf{T}_{col}(i,j,c)\right) \quad (9)$$

where *H*, *W*, and *C* represent the height, width, and channel of the image, respectively. Additionally, $\mathbf{T}_{\text{row}}$ and $\mathbf{T}_{\text{col}}$ represent the row-difference matrix and column-difference matrix, respectively. These two difference matrix are defined as (10) and (11), respectively

$$\mathbf{T}^{i}_{\text{row}} = \left|\mathbf{I}^{i}_{\text{row}} - \mathbf{I}^{i-1}_{\text{row}}\right| \quad (10)$$

$$\mathbf{T}^{j}_{\text{col}} = \left|\mathbf{I}^{j}_{\text{col}} - \mathbf{I}^{j-1}_{\text{col}}\right| \quad (11)$$

where $\mathbf{T}^{i}_{\text{row}}$ represents the $i$th row of $\mathbf{T}_{\text{row}}$, $\mathbf{T}^{j}_{\text{col}}$ represents the $j$th column of $\mathbf{T}_{\text{col}}$. Similarly, $\mathbf{I}^{i}_{\text{row}}$ and $\mathbf{I}^{j}_{\text{col}}$ are defined in the same way.

*2) Label Definition:* Through the aforementioned procedures, we can obtain the texture complexity of each image. Next, we should give each image a specific label according to its corresponding *T*. In this article, we utilize a simple truncation division strategy to generate the texture complexity label of each image in the datasets. The detailed definition regulars are demonstrated as follows. First, calculate the texture complexity of each image in the dataset. Second, set the truncation threshold of different complexity levels. Specifically, $T^{th}_{i} = T_{\min} + i \cdot \frac{T_{\max} - T_{\min}}{N_L}, i \leq 0 \leq N_L$. In which, $T^{th}_{i}$ is the $i$th threshold of the complexity, $T_{\min}$ and $T_{\max}$ are the minimum and maximum of the texture complexities of the image in the dataset, respectively. $N_L$ is the number of the complexity levels, and $i$ represents the index. Third, attach the specific label to each image according to the corresponding threshold, which is formulated as

$$L^{c}(\mathbf{I}_k) = j$$
$$\text{s.t.} T^{th}_{j-1} \leq T(\mathbf{I}_k) \leq T^{th}_{j} \quad (12)$$
$$1 \leq j \leq N_L$$

where $\mathbf{I}_k$ represent the $k$th image in the dataset, $L^{c}(\mathbf{I}_k)$ represents its label, and $T(\mathbf{I}_k)$ is the corresponding texture complexity value. Some examples of images with different texture complexity labels are shown in Fig. 5.

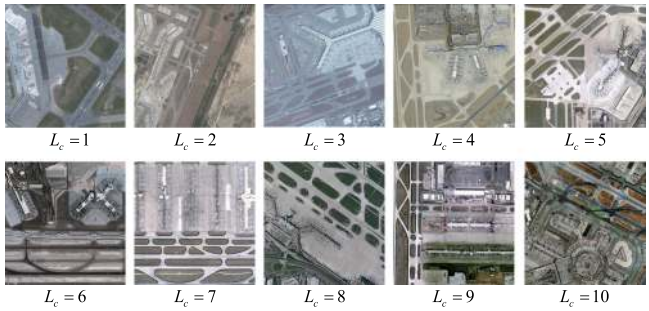


Fig. 5. Samples of images with different texture complexity label.

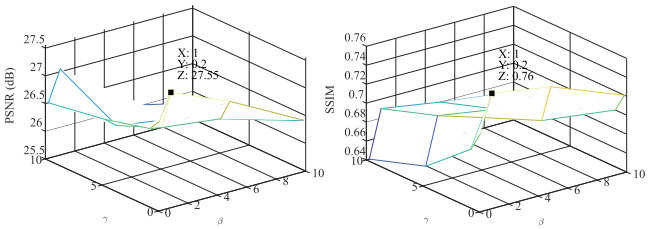


Fig. 6. Experimental results of UCMBlurred set under different hyperparameters.

## V. Experiments

This section details the experiments, including experiment settings, evaluation metrics, and experiment results and analysis.

### A. Experiment Settings

In this subsection, we will demonstrate the experimental settings, including data participation, hyperparameter settings, and data transformation.

*1) Data Partition:* In this article, we randomly divide each dataset into three parts, training set, validation set, and testing set. The proportions of these three parts are 60%, 20%, and 20%, respectively.

*2) Hyperparameter Settings:* Besides the participation proportion, there are still three extra hyperparameters to be set, $\alpha$, $\beta$, and $\gamma$ in 8, which are used to balance three loss terms. Considering the relative other than absolute importance of these three hyperparameters make sense, we fix $\alpha = 10.0$, and utilize the cross validation strategy to obtain corresponding optimal $\beta$ and $\gamma$. The quantitative results on UCMBlurred dataset are shown in Fig. 6. From which we can see that, when $\beta$ and $\gamma$

are, respectively, set to 1 and 0.2, our method achieves the best performance.

*3) Data Transformation:* This section introduces the details of the data transformation. First of all, the raw images are whitened. On one hand, which can avoid the interferes from noises. On the other hand, it can help us to control the relative importance of three terms in (8). Additionally, PSNR and SSIM are two metrics for grayscale images. In this case, we transform generated deblurred images and the raw clear images to grayscale ones when the performances of different methods are evaluated. The transformation formulation is defined as follows:

$$\text{Gray} = 0.2989 \cdot R + 0.5870 \cdot G + 0.1140 \cdot B \tag{13}$$

where R, G, and B represent three channel information of the image, respectively, and Gray represents the transformed grayscale image.

### B. Evaluation Metrics

In this article, we use peak signal-to-noise ratio (PSNR) and structural similarity index measurement (SSIM), which are widely used in many image quality assessment tasks [56], to evaluate the methods, which are defined as (14) and (16), respectively

$$PSNR\left(\mathbf{X},\mathbf{Y}\right) = 10\log_{10}\left(\frac{\left(2^N-1\right)^2}{\text{MSE}\left(\mathbf{X},\mathbf{Y}\right)}\right) \tag{14}$$

where MSE is the mean square error, which is formulated in (15). $N$ is the number of the bit, which is set to 8 in this article

$$\text{MSE}\left(\mathbf{X},\mathbf{Y}\right) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(\mathbf{X}(i,j)-\mathbf{Y}(i,j)\right)^2. \tag{15}$$

In (14) and (15), $\mathbf{X}$, $\mathbf{Y}$ are two grayscale images to be measured, which represent blurred image and clear image, respectively, in this article. Additionally, *H* represents the height and *W* represents the width of the image

$$SSIM\left(\mathbf{X},\mathbf{Y}\right) = l\left(\mathbf{X},\mathbf{Y}\right)\cdot c\left(\mathbf{X},\mathbf{Y}\right)\cdot s\left(\mathbf{X},\mathbf{Y}\right) \tag{16}$$

where $l(\mathbf{X},\mathbf{Y})$, $c(\mathbf{X},\mathbf{Y})$, and $s(\mathbf{X},\mathbf{Y})$ are defined as (17), (18), and (19), respectively.

$$l\left(\mathbf{X},\mathbf{Y}\right) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2+\mu_Y^2+C_1} \tag{17}$$

$$c\left(\mathbf{X},\mathbf{Y}\right) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2+\sigma_Y^2+C_2} \tag{18}$$

$$s\left(\mathbf{X},\mathbf{Y}\right) = \frac{\sigma_{XY}+C_3}{\sigma_X\sigma_Y+C_3}. \tag{19}$$

In which, $\mu_X$ and $\mu_Y$ represent the mean values of $\mathbf{X}$ and $\mathbf{Y}$, respectively. $\sigma_X$ and $\sigma_Y$ represent the variance of X and Y, respectively, and $\sigma_{XY}$ represents the covariance of $\mathbf{X}$ and $\mathbf{Y}$. Additionally, $C_1$, $C_2$, and $C_3$ are three constants to avoid divided by zero, which are set to 0.01, 0.03, and 0.015, respectively. In (18) to (19), the mean value, variance, and covariance can be calculated with (20) to (22), respectively

$$\mu_X = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W}\mathbf{X}(i,j) \tag{20}$$

$$\sigma_X^2 = \frac{1}{HW-1}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(\mathbf{X}(i,j)-\mu_X\right)^2 \tag{21}$$

$$\sigma_{XY} = \frac{1}{HW-1}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(\mathbf{X}(i,j)-\mu_X\right)\left(\mathbf{Y}(i,j)-\mu_Y\right). \tag{22}$$

### C. Contrasting Approaches

To demonstrate the superiority of our method, we compare it with five existing ones, including a traditional method and four deep neural network-based ones.

Graph-based blind image deblurring from a single photograph (GSP) [40] interprets an image patch as a signal on a weighted graph to strengthen its effectiveness.

Blind deconvolution for image deblurring based on edge enhancement and noise suppression (BDS) [57] improves its performance through emphasizing edge regions and suppressing noises.

Gated fusion network (GFN) [58] is a deep gated fusion CNN to generate a clear high-resolution frame from a single natural image with severe blur.

Deep generative filter (DGF) [48] is a deep architecture integrated with global skip connection and dense architecture for image deblurring.

Deblur generative adversarial network (DeblurGAN) [59] combines the content loss and adversarial framework to exploit the latent information of the image.

### D. Experimental Results and Analysis

This section depicts the experiments, mainly including hypothesis validation, ablation experiments, and contrasting experiments.

*1) Hypothesis Validation:* First, to verify the assumption that texture complexity can influence the image debluring quality to a certain extent, we design a direct contrasting experiment. Specifically, we apply an existing deblurring method [60] to images with different texture complexity, and then evaluate the performances of the same methods to different images. The experimental results are shown in Table II, where TCL is the abbreviation of texture complexity level.

From Table II we can see that, texture complexity is an important factor for remote sensing image deblurring. Specifically, images with complex textures are more difficult to restore, compared to those with simple textures. For instance, compared to the restored 1st-complexity-level image with *SSIM* 0.8154, *SSIM* of the restored image of 10th-complexity-level is only 0.5203. In this case, it is necessary for incorporating the texture complexity information of image into the deblurring architecture. Additionally, because the estimation of image texture complexity may

TABLE II
HYPOTHESIS VALIDATION RESULTS

| TCL | *PSNR*(dB) | *SSIM* |
|---|---|---|
| 1 | **28.5978** | **0.8154** |
| 2 | 28.2996 | 0.8345 |
| 3 | 27.4269 | 0.8062 |
| 4 | 24.9976 | 0.7519 |
| 5 | 21.7193 | 0.7265 |
| 6 | 21.6911 | 0.6988 |
| 7 | 20.4615 | 0.6786 |
| 8 | 20.4409 | 0.6225 |
| 9 | 19.8975 | 0.6079 |
| 10 | 18.7906 | 0.5203 |

The bold entities represent the best performance under the corresponding experimental settings and algorithms.

TABLE III
ABLATION EXPERIMENTS

| Methods | *PSNR*(dB) | *SSIM* |
|---|---|---|
| WithoutB | 25.0673 | 0.6845 |
| OnlywithT | 26.8255 | 0.7371 |
| OnlywithW | 26.7683 | 0.7424 |
| WithB | **27.0272** | **0.7782** |

The bold entities represent the best performance under the corresponding experimental settings and algorithms.

inaccurate, we utilize classification mode but not regression mode to depict its information, which has higher fault tolerance.

*2) Ablation Experiments:* Second, the proposed method mainly contains two aspects, texture complexity label information and weighted attention map loss term, both of them should be verified contributory to the image deblurring. In this case, we design a simple ablation experiment, train the network without, with one of them, or with both of them, and then evaluate the performances in different conditions. The experimental results are shown in Table III, where WithoutB means the model trained only by the traditional reconstruction loss, OnlywithT means the model trained by reconstruction loss term and texture complexity valuation loss term, OnlywithW means the model trained by reconstruction loss term and weighted attention map loss term, WithB means the model trained by reconstruction loss term, texture complexity valuation loss term, and weighted attention map loss.

From Table III, we can see that, both the texture complexity evaluation loss term and weighted attention loss term are helpful for the restore quality of remote sensing images. Specifically, when the texture complexity evaluation loss term is added to the traditional reconstruction Frobenius norm, 0.0526 improvement in terms of *SSIM* is obtained, which demonstrates that the texture information is very relative to the image deblurring task. Additionally, weighted attention loss term improves the performance further. Generally speaking, these two proposed terms are effective for deblurring.

TABLE IV
RESULTS ON THREE CONSTRUCTED DATASETS

| Dataset | Method | *PSNR*(dB) | *SSIM* |
|---|---|---|---|
| SydneyBlurred | GSP [40] | 24.7823 | 0.6551 |
| | BDS [57] | 25.9587 | 0.6995 |
| | GFN [58] | 26.4478 | 0.7031 |
| | DGF [48] | 26.6950 | 0.7226 |
| | DeblurGAN [59] | 27.0824 | 0.7364 |
| | MTLM | **27.7761** | **0.7583** |
| UCMBlurred | GSP [40] | 24.8235 | 0.6603 |
| | BDS [57] | 25.9042 | 0.6875 |
| | GFN [58] | 26.1564 | 0.6982 |
| | DGF [48] | 26.4237 | 0.7233 |
| | DeblurGAN [59] | 27.1098 | 0.7461 |
| | MTLM | **27.5526** | **0.7645** |
| AIDBlurred | GSP [40] | 24.5344 | 0.6437 |
| | BDS [57] | 25.8825 | 0.6716 |
| | GFN [58] | 26.2053 | 0.7038 |
| | DGF [48] | 26.6534 | 0.7315 |
| | DeblurGAN [59] | 26.8175 | 0.7542 |
| | MTLM | **27.0272** | **0.7782** |

The bold entities represent the best performance under the corresponding experimental settings and algorithms.

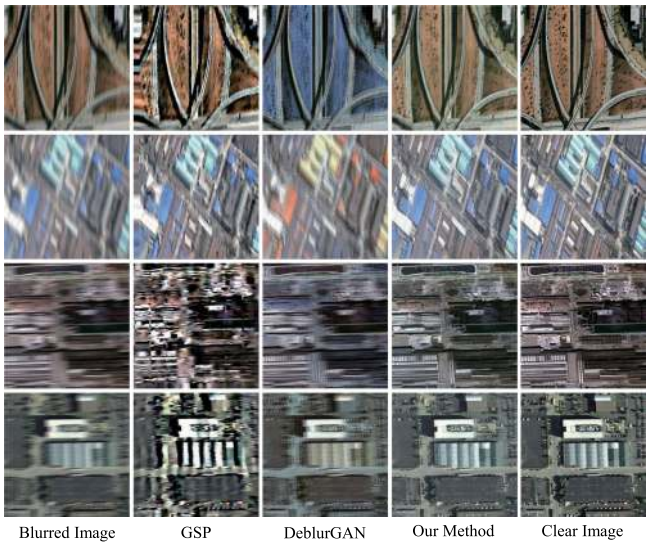


Fig. 7. Samples of visualized results of different methods. Images in the first column denote the blurred ones. Images in the second to fourth column denote the results of GSP, DeblurGAN, and our method, respectively. Images in the last column denote the corresponding clear images.

*3) Contrasting Experiments:* Third, to verify the superiority of our method, we apply the proposed method and the contrasting methods to the constructed datasets, respectively. The experimental results are shown in Table IV, and some samples of the visualized results are shown in Fig. 7.

From Table IV we can see that, the proposed multitask learning mechanism achieves the best performances on all three testing datasets. Specifically, on SydneyBlurred dataset, MTLM achieves 0.6937-dB improvements in terms *PSNR*, and 0.0219

improvement in terms of *SSIM* respectively, compared with DeblurGAN. On AIDBlurred dataset, MTLM gains 0.2097 dB improvement in terms of *PSNR* and 0.0258 increment in terms of *SSIM*, respectively. According to the experimental results, compared with other contrasting methods, especially DeblurGAN, the superiority of MTLM is unconspicuous on AIDBlurred dataset. The main reason is that, besides the architecture itself of the model, we utilize the proposed weighted attention loss term to depict the detailed structure of the image, while DGF uses a adversary loss term. As is known, adversary loss usually results in the unstable and even not-convergent phenomenon during the training process, especially when the scale of the training set is not sufficiently large. Additionally, when the training set scale is large enough, the adversary loss often can depict the details of images well. In these cases, the weighted attention loss term is superior than adversary loss in small scale training set, and so the results on AIDBlurred dataset is not so significant as we expected (the AIDBlurred dataset is a relatively larger one), compared with DeblurGAN.

Additionally, from Fig. 7 we can see that the visualized results of DeblurGAN are more significant than those of GSP, especially ones in the last two rows. This is because that the deep neural network can dig out more sufficient projection information between blurred domain and clear domain, compared to the tradition kernel estimation-based methods. Besides, the visualized results of the MTLM are more satisfactory than those of DeblurGAN, especially the example in the third row. The main reasons contain two aspects: 1) The adversarial model can not be trained well with limited samples, hence DeblurGAN is sensitive to even small uncontrollable noise, and further lead to the distortion. 2) The proposed method uses texture prior and attention prior to replace the adversarial mechanism to exploit the latent information of the image, which can ensure the robustness of the proposed method to a certain extent.

## VI. Conclusion

In this article, we propose a multitask learning mechanism-based method for remote sensing image motion deblurring. First, we find that the texture complexity is a relatively important factor for the image-deblurring task. In this case, we incorporate a texture complexity recognition branch into the traditional reconstruction network to improve the deblurring quality. Second, we design a weighted attention map loss term to enhance the deblurring capability of difficult regions. Finally, through a series of experiments on three constructed datasets, we verify both the proposed components can contribute to the performance of the deblurring to a certain extent. Generally speaking, compared to the conventional methods, the proposed method can effectively and efficiently realize single-step deblurring in time through adaptive matching filters.

Even though the proposed method achieves relatively satisfactory performance for remote sensing image motion deblurring task, there is still an important issue to be addressed: How to generalize the proposed method to other remote sensing image quality improve tasks such as denoising? In the future, we intend to incorporate the partition strategy into our network according to different quality reduction factors, and try to construct an unified architecture for remote sensing image quality improvement and improve its generalization capability further.

## References


[1] F. Zhang and Y. Tian, "Image motion deblurring based on salient structure selection and L0-2 norm kernel estimation," *J. Comput. Commun.*, vol. 5, no. 3, p. 24, 2017.

[2] M. Aittala and F. Durand, "Burst image deblurring using permutation invariant convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 731–747.

[3] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8174–8182.

[4] Q. Yuan, J. Li, L. Zhang, Z. Wu, and G. Liu, "Blind motion deblurring with cycle generative adversarial networks," 2019, *arXiv:1901.01641*.

[5] D. Wipf and H. Zhang, "Revisiting Bayesian blind deconvolution," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3595–3634, 2014.

[6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, "Removing camera shake from a single photograph," *ACM Trans. Graph.*, vol. 25, no. 3, 2006, pp. 787–794.

[7] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1964–1971.

[8] L. Xu, S. Zheng, and J. Jia, "Unnatural L0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 1107–1114.

[9] I. Donadello, "Semantic image interpretation - Integration of numerical data and logical knowledge for cognitive vision," Ph.D. dissertation, Univ. Trento, Trento, Italy, 2018. [Online]. Available: http://eprints-phd.biblio.unitn.it/2888/

[10] W. Cao, J. Yuan, Z. He, Z. Zhang, and Z. He, "Fast deep neural networks with knowledge guided training and predicted regions of interests for real-time video object detection," *IEEE Access*, vol. 6, pp. 8990–8999, 2018.

[11] A. Chakrabarti, "Neural approach to blind motion deblurring," in *Proc. Eur. Conf. Comput. Vis.,* 2016, pp. 221–235.

[12] Z. Chen and L. Chang, "Blind motion deblurring via inceptionresdensenet by using GAN model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Proc.*, May 2019, pp. 1463–1467.

[13] J. S. Lumentut, T. H. Kim, R. Ramamoorthi, and I. K. Park, "Fast and full-resolution light field deblurring using a deep neural network," 2019, *arXiv:1904.00352*.

[14] E. Peli, "Contrast in complex images," *J. Opt. Soc. Amer. A*, vol. 7, no. 10, pp. 2032–2040, 1990.

[15] B. Ham, M. Cho, and J. Ponce, "Robust guided image filtering using nonconvex potentials," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 192–207, Jan. 2018.

[16] S. K. Nayar and M. Ben-Ezra, "Motion-based motion deblurring," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 6, pp. 689–698, Jun. 2004.

[17] L. Xu and J. Jia, "Two-phase kernel estimation for robust motion deblurring," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 157–170.

[18] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless, "Single image deblurring using motion density functions," in Proc. *Eur. Conf. Comput. Vis.*, 2010, pp. 171–184.

[19] S. Su and W. Heidrich, "Rolling shutter motion deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1529–1537.

[20] J. Pan, Z. Hu, Z. Su, H.-Y. Lee, and M.-H. Yang, "Soft-segmentation guided object motion deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 459–468.

[21] P. Svoboda, M. Hradiš, L. Maršik, and P. Zemcìk, "CNN for license plate motion deblurring," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 3832–3836.

[22] L. Yang and H. Ji, "A variational EM framework with adaptive edge selection for blind motion deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 10 167–10 176.

[23] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao, "Image deblurring via extreme channels prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4003–4011.

[24] M. D. Kim and J. Ueda, "Real-time panoramic image generation and motion deblurring by using dynamics-based robotic vision," *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 3, pp. 1376–1387, Jun. 2016.

[25] Z. Shen *et al.*, "United estimation of blur distribution for space-variant rotation motion deblurring," *Proc. SPIE*, 2015, Art. no. 960005.

[26] C.-F. Chang, J.-L. Wu, and K.-J. Chen, "A hybrid motion deblurring strategy using patch based edge restoration and bilateral filter," *J. Math. Imag. Vis.*, vol. 60, no. 7, pp. 1081–1094, 2018.
[27] P. C. Hansen, J. G. Nagy, and D. P. O'leary, *Deblurring Images: Matrices, Spectra, and Filtering*. Philadelphia, PA, USA: Siam, 2006, vol. 3.
[28] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series-With Engineering Applications*." Cambridge, MA, USA: MIT press, 1949.
[29] W. H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, no. 1, pp. 55–59, 1972.
[30] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astronomical J.*, vol. 79, p. 745, 1974.
[31] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-Laplacian priors," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1033–1041.
[32] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, p. 70, 2007.
[33] M. S. Almeida and L. B. Almeida, "Blind and semi-blind deblurring of natural images," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 36–52, Jan. 2010.
[34] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, Dec. 2011.
[35] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "L_0-Regularized intensity and gradient prior for deblurring text images and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 342–355, Feb. 2017.
[36] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Framelet-based blind motion deblurring from a single image," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 562–572, Feb. 2012.
[37] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Blind image deblurring using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1628–1636.
[38] L. Sun, S. Cho, J. Wang, and J. Hays, "Edge-based blur kernel estimation using patch priors," in *Proc. IEEE Int. Conf. Comput. Photography*, 2013, pp. 1–8.
[39] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M.-H. Yang, "Image deblurring via enhanced low-rank prior," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3426–3437, Jul. 2016.
[40] Y. Bai, G. Cheung, X. Liu, and W. Gao, "Graph-based blind image deblurring from a single photograph," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1404–1418, Mar. 2019.
[41] B. Goyal, A. Dogra, S. Agrawal, B. Sohi, and A. Sharma, "Image denoising review: From classical to state-of-the-art approaches," *Inf. Fusion*, vol. 55, pp. 220–244, 2019.
[42] Y. Yuan, J. Fang, X. Lu, and Y. Feng, "Remote sensing image scene classification using rearranged local features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1779–1792, Mar. 2019.
[43] J. Fang, Y. Yuan, X. Lu, and Y. Feng, "Robust space-frequency joint representation for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7492–7502, Oct. 2019.
[44] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 265–278, Jan. 2019.
[45] J. Fang and X. Cao, "GAN and DCN based multi-step supervised learning for image semantic segmentation," in *Proc. Chinese Conf. Pattern Recognit. Comput. Vis.*, 2018, pp. 28–40.
[46] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1125–1134.
[47] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4681–4690.
[48] S. Ramakrishnan, S. Pachori, A. Gangopadhyay, and S. Raman, "Deep generative filter for motion deblurring," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2993–3000.
[49] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[50] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2005, pp. 886–893.
[51] L. Han, X. Li, and Y. Dong, "Convolutional edge constraint-based U-net for salient object detection," *IEEE Access*, pp. 48 890-48 900, 2019.
[52] S. Wu, J. Fang, X. Zheng, and X. Li, "Sample and structure-guided network for road crack detection," *IEEE Access*, vol. 7, pp. 130 032–130 043, 2019.
[53] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, Apr. 2015.
[54] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, 2010, pp. 270–279.
[55] G.-S. Xia *et al.*, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.
[56] T. Hu, W. Li, N. Liu, R. Tao, F. Zhang, and P. Scheunders, "Hyperspectral image restoration using adaptive anisotropy total variation and nuclear norms," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1516–1533, 2021.
[57] C. Cai, H. Meng, and Q. Zhu, "Blind deconvolution for image deblurring based on edge enhancement and noise suppression," *IEEE Access*, vol. 6, pp. 58 710–58 718, 2018.
[58] X. Zhang, H. Dong, Z. Hu, W.-S. Lai, F. Wang, and M.-H. Yang, "Gated fusion network for joint image deblurring and super-resolution," 2018, *arXiv:1807.10806*.
[59] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8183–8192.
[60] Q. Shan, J. Jia, and A. Agarwala, "High-quality motion deblurring from a single image," *ACM Trans. Graph. (tog)*, vol. 27, no. 3, p. 73, 2008.

**Jie Fang** received the B.S. degree from Xidian University, Xian, China, in 2015. He received the Ph.D. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2020.

He is currently an Associate Professor with the Center for Image and Information Processing, School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an, China. His research interests include artificial intelligence, machine learning, and video/audio analysis.

**Xiaoqian Cao** received the B.S. degree from School of Machine Engineering at Xi'an Jiaotong University, Xi'an, China, in 2009. She received the Ph.D. degree in signal and information processing techniques from the University of Chinese Academy of Sciences, Beijing, China, in 2014.

She is currently a Lecturer with the College of Electric and Control Engineering, Shaanxi University of Scinece and Technology, Xi'an, China. Her research interests include pattern recognition, visual understanding, and stereo matching.

**Dianwei Wang** received the Ph.D. degree in navigation, guidance, and control from Northwestern Polytechnical University, Xi'an, China, in 2010.

He is currently an Associate Professor with the Center for Image and Information Processing, School of Telecommunication and Information Engineering, Xian University of Posts and Telecommunications and School of Artificial Intelligence, Xian, China. His research interests include nonstationary signal processing and analysis, and application of artificial intelligence in the machine vision field.

**Shengjun Xu** received the master's degree in computer application from Xi'an University of Architecture and Technology, Xi'an, China, in 2003. He received the Ph.D. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, China, in 2014.

He is currently with the College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an, China. His research interests include image processing, pattern recognition, artificial intelligence, and automation.