BrGAN: Blur Resist Generative Adversarial Network With Multiple Joint Dilated Residual Convolutions for Chlorophyll Color Image Restoration

Ziyi Chen, *Member, IEEE*, Yuhua Luo, Yiping Chen, *Senior Member, IEEE*, Jing Wang,
Dilong Li, *Member, IEEE*, Kyle Gao, *Graduate Student Member, IEEE*,
Cheng Wang, *Senior Member, IEEE*, and Jonathan Li, *Fellow, IEEE*

Abstract—This article presents a blur resist generative adversarial network (BrGAN) with multiple joint dilated residual convolutions for chlorophyll image restoration of the Geostationary Ocean Color Imager (GOCI). First, a publicly available dataset was built to support this study. Second, a multiple attention perception mechanism and a multiple joint dilated residual convolution module were proposed to cope with the challenge of large missing areas in GOCI chlorophyll images. Third, a patch generative adversarial network (GAN)-based discrimination module was proposed to avoid the restored areas with generating mosaic and shadows. Our experimental results demonstrate that the BrGAN can reach 37.06 in the peak signal-to-noise ratio (PSNR) and 0.0485 in the Learned Perceptual Image Patch Similarity (LPIPS), respectively. The comparative study shows that the BrGAN achieves the highest effectiveness and advancement among other seven state-of-the-art (SOTA) methods.

Index Terms—Convolutional neural network (CNN), Geostationary Ocean Color Imager (GOCI) ocean chlorophyll images, image inpainting, image restoration.

Ziyi Chen, Yuhua Luo, Jing Wang, and Dilong Li are with the Department of Computer Science and Technology, Huaqiao University, Xiamen 361021, China (e-mail: chenziyihq@hqu.edu.cn; lyh3932855@163.com; wroaring@hqu.edu.cn; scholar.dll@hqu.edu.cn).

Yiping Chen is with the School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai 519082, China (e-mail: chenyp79@mail.sysu.edu.cn).

Kyle Gao and Jonathan Li are with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: y56gao@uwaterloo.ca; junli@uwaterloo.ca).

Cheng Wang is with the School of Information Science and Engineering, Xiamen University, Xiamen 361005, China (e-mail: cwang@xmu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2023.3324993

I. INTRODUCTION

EARTH observation images play an important role in providing crucial information assistance to maritime safety [1], [2], [3], ocean carbon cycling monitoring [4], [5], algae blooms detection [6], [7], current movement surveillance [8], [9], etc. Among the above applications, the large area current movement surveillance based on chlorophyll images has attracted great attention of researchers. However, the ocean chlorophyll images usually appear with incompleteness and serious noise due to the interference of clouds. How to conquer the above problem has a great significance to the further applications based on the ocean chlorophyll images.

Under our review, the current research of the denoising and inpainting restoration of ocean chlorophyll satellite images is rather rare, even there is no publicly available relevant datasets. Besides, although the image denoising and inpainting restoration has achieved great successes in natural scene images, most of the state-of-the-art (SOTA) denoising and inpainting restoration algorithms [10], [11], [12], [13], [14] are not suitable for the ocean chlorophyll satellite images and usually cannot achieve satisfactory performance. Most popular image denoising and inpainting restoration methods are convolutional neural network (CNN) [15] based and achieved great successes for face and building images [16]. However, due to the lack of considerations about the semantic information reconstruction mechanism to deal with the large areas missing problem in the ocean satellite chlorophyll images, those SOTA approaches failed to obtain good performances. Fig. 1 shows the two Geostationary Ocean Color Imager (GOCI) images with noises and large area missing caused by clouds. The noise and missing of large areas will cause great challenges to the follow-up relevant applications.

To overcome the above problems, this article mainly studies from three aspects. First, we constructed and publicly shared a GOCI chlorophyll image dataset for denoising and restoration. Our dataset consists of three parts: training, testing, and validation of images. As the GOCI chlorophyll images have no truth labels, we manually designed the missing areas
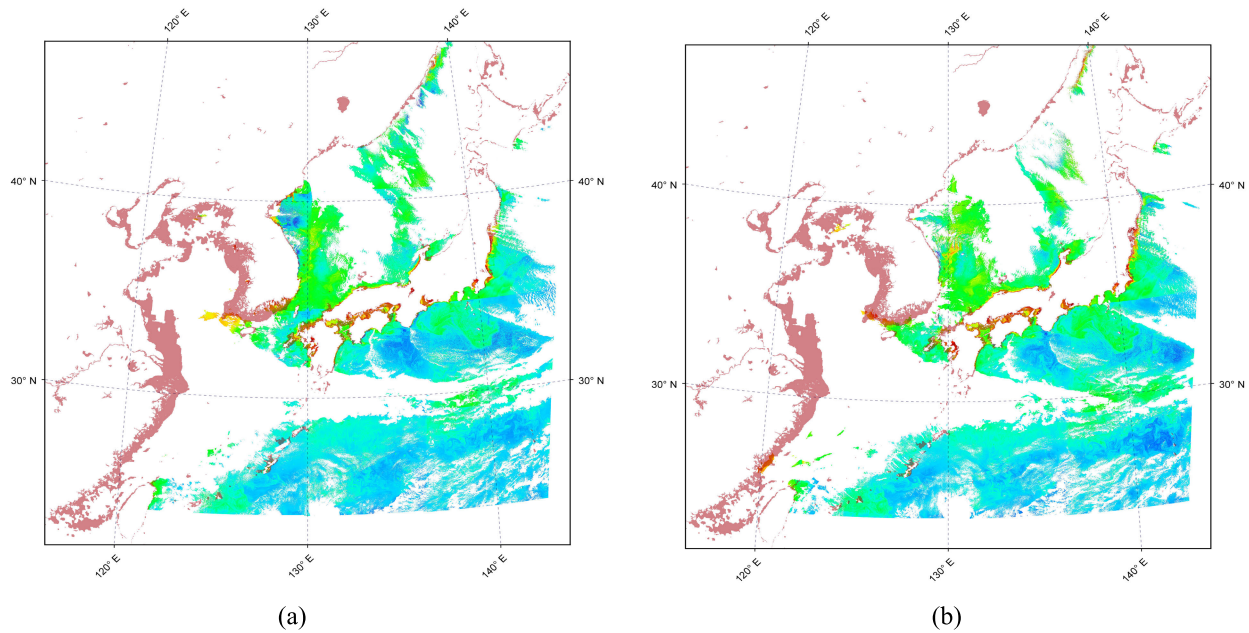
Fig. 1. Two sample GOCI chlorophyll images captured between (a) 1:00 pm to 2:00 pm and (b) 3:00 pm to 4:00 pm, January 29, 2014 (light coral color: the presence of shallow water along the coast).

of a complete image, thus, the original completed image can be the truth labels. Second, we proposed a Blur Resist Generative Adversarial Network (BrGAN) that utilizes a self-attention mechanism to reinforce the high-level semantic feature extraction, a union dilated convolutional module (UDCM) to reinforce the reconstruction ability and a union loss to control the model learning. Finally, to avoid generation images with shadows and mosaics, we proposed a patch generative adversarial network (GAN)-based discrimination module to guarantee the texture consistency of the final restored images.

Our contributions include the following two points.

1) A publicly available dataset of the GOCI chlorophyll images for denoising and restoration tasks. A novel BrGAN that is based on a self-attention mechanism, a UDCM and GAN framework for denoising and restoration of the GOCI chlorophyll images.

2) A union loss that is consisted of pixel-level losses, perceptual losses, style losses, and against loss, to control the trained network with excellent performances for denoising and restoration.

II. RELATED WORK

Image restoration has always been one of the vital tasks in computer vision. Briefly, the image restoration algorithm repair and inpainting the missed pixel values, texture features, and semantic information of an image. First, the image restoration algorithms were widely used in the restoration of the art drawings [17]. With the breakthroughs of deep learning, the completeness and the semantic information accuracy of the restored images have been greatly promoted, and achieved successes in the image inpainting of faces [18], buildings, and natural scenes [19], [20], [21].

In this section, we review the image restoration algorithms, including traditional manually designed, CNN-based, and GOCI chlorophyll image restoration.

A. Traditional Manually Designed Models

Most traditional manually designed models are diffusion-based algorithm and patch-based algorithm. The diffusion-based image restoration algorithms usually restore the image from complete area pixels to lost or damaged area pixels [22]. Recently, many researchers have focused on the study about those algorithms. An image restoration algorithm based on the structure information digging was proposed in [23]. To improve the completeness of structure information and guarantee the local consistency in the reconstructed images, they designed a new iteratively structure search approach. Similarly, it was found in [24] that when repairing the missing areas, the regional similarity of the image would be destroyed if the incomplete image was affected by noise, blur, and other factors. Therefore, they proposed a postprocess-based target removal method [24], which could further optimize the repaired image, and reduce repair trace in the missing image areas. The traditional image restoration methods include those ones based on image patch repair [17], [25], [26], [27], the main method is to mask or block the original image, then to repair the image. The drawing and filling method based on image Hankel structure matrix was first proposed in [28]. The method calculates the similarity degree between blocks through the Hankel structure matrix in image blocks, so as to repair missing blocks. A context-aware image repair method was proposed in [29] by cutting the image top-down into multiple variable patches to identify these areas.

B. Deep Learning-Based Methods

In the existing image repair method, partial convolution proposed in [10] will be used to repair the irregular holes in the image, and the original image will be patched according to the acquired missing areas through irregular mask. Similarly, the model of training with irregular masks has promoted the rapid development of image restoration.

Pluralistic image completion (PIC)-net adopts a probabilistic framework and two paths in parallel to generate multiple recovery results for a mask image, and introduces short-term and long-term memory mechanisms [30], so as to create a consistent connection between the encoder and the decoder, so that the recovery image of the lost image is as consistent as possible with the original image. The Shift-Net was proposed in [31] based on the U-Net structure presented in [32]. The main purpose is to restore the semantic information and texture structure of images. A new migration network is adopted to make the encoding part and decoding part have more precise connection.

In the process of inpainting missing areas, the discontinuity of local pixels in the missing area often occurs and the resolution of the restored image structure is low. For this reason, an image repair method was proposed in [33], in which the network focuses on coherent semantic information in the image. Network structure (coherent semantic attention (CSA)-net) divides the process of image restoration into two parts: rough restoration and high precision restoration, which are also based on U-Net structure [34]. The first part mainly makes a rough preliminary reasoning about the overall structure texture and pixel value of the missing area of the incomplete image, and introduces extra features of visual geometry group (VGG) layer in the first part. While the second part combines the input and output of the first part. Compared with the first part, this network structure has more depths. Moreover, the encoder and decoder learn the characteristics of the VGG layer through consistency loss to achieve a better prediction. Subsequently, a module based on recurrent feature reasoning (RFR) was proposed in [12], which is mainly divided into two parts, namely partial convolution layer and U-Net-based feature inference layer. The main function of partial convolution layer is image region recognition. The difference between the feature inference layer and the traditional U-Net structure layer is that a knowledge consistent attention (KCA) module is added between downsampling and upsampling to make the details in the image more easily to be expressed. The main process of this module is to first identify the region, and then input the results to the feature inference layer for prediction. The predicted result will be returned to the part of the convolution layer for the next stage of reasoning. This module makes use of the correlation between adjacent pixels to deepen the estimation of deeper pixels. A hypergraph convolution was used in [35] to generate image global semantics. However, there are two problems need to be solved in the restoration of GOCI chlorophyll color images by this model: one is that the loss area of the image to be repaired needs to be provided; the other is that it is difficult to remove the noise from the images.

Unlike the model proposed in [35], the BrGAN model proposed in this article for restoration of the GOCI chlorophyll color images simplifies the preprocessing in the restoration process. The network only needs to input the image in order to be repaired to output the restored image, and this model has achieved good results in noise removal, missing area restoration, and simultaneous implementation of two tasks. As shown in Fig. 2, when RFR-Net repairs images with both missing and noise points, pixel confusion appears in the repair effect from the perspective of vision, and the repair result of
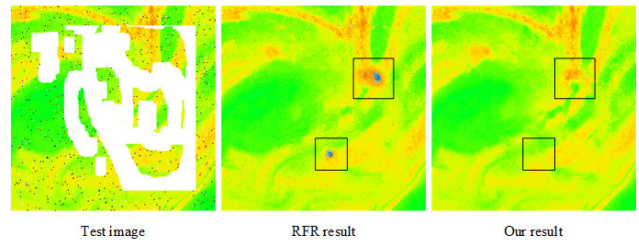

Fig. 2. Comparison of RFR model and our model in simultaneously removing noise and repairing missing areas (black box: the area in the image with considerable difference).

our model is smoother compared with it, and the probability of missing caused by noise points and cloud occlusion is very high in the GOCI chlorophyll color images.

III. METHOD

In this section, we will introduce our BrGAN from the aspects of the whole network structure, the generator part and includes the self-attention module, the multiple residual and dilated convolution module, the discriminator part, and the designed loss function.

A. Framework of BrGAN

In our BrGAN, two parts (generator and discriminator) are included. The generator module uses an image with missing pixels as input, then generates a repaired result. To evaluate the repaired result and guide the generator module to learn better network parameters, the generated result is then put into a discriminator module, and backward a discrimination loss. Finally, under the control of a fusion loss which includes four different kinds of losses, the BrGAN can be well trained in an end-to-end way.

B. Generator of BrGAN

As shown in the generator part in Fig. 3(a), the generator is divided into two parts, i.e., the encoder and the decoder. Encoders are divided into five groups from top to bottom. For the first group, it consists of the input layer, two convolution layers with 3×3 convolution kernels, Batch normalization (BN) [36], and the ReLU activation function, as shown in the gray square in Fig. 3. Besides the above layers, the residual strategy [37] is also used. In addition, the number of filters in the convolutional layer is 64. The other four groups are composed of 2×2 maximum pooling layer, two convolution layers with the size of 3×3 convolution kernel, a BN layer, and a ReLU layer, respectively. The residual mechanism is also used between the two adjacent convolution operations. The number of downsampling filters is 128, 256, 512, and 1024, respectively. The feature size is downsampled from 224×224 to 14×14 in stages. After that, the downsampled features are put into the self-attention module between the decoder and the encoder, obtaining more efficient high-level semantic information from the encoder and transferring to the encoder, making the decoder have more powerful recovery performance. The decoder structure can also be divided into five groups, as shown by the black arrows. For each group operation, the features in the encoder are spliced with the features in the corresponding decoder
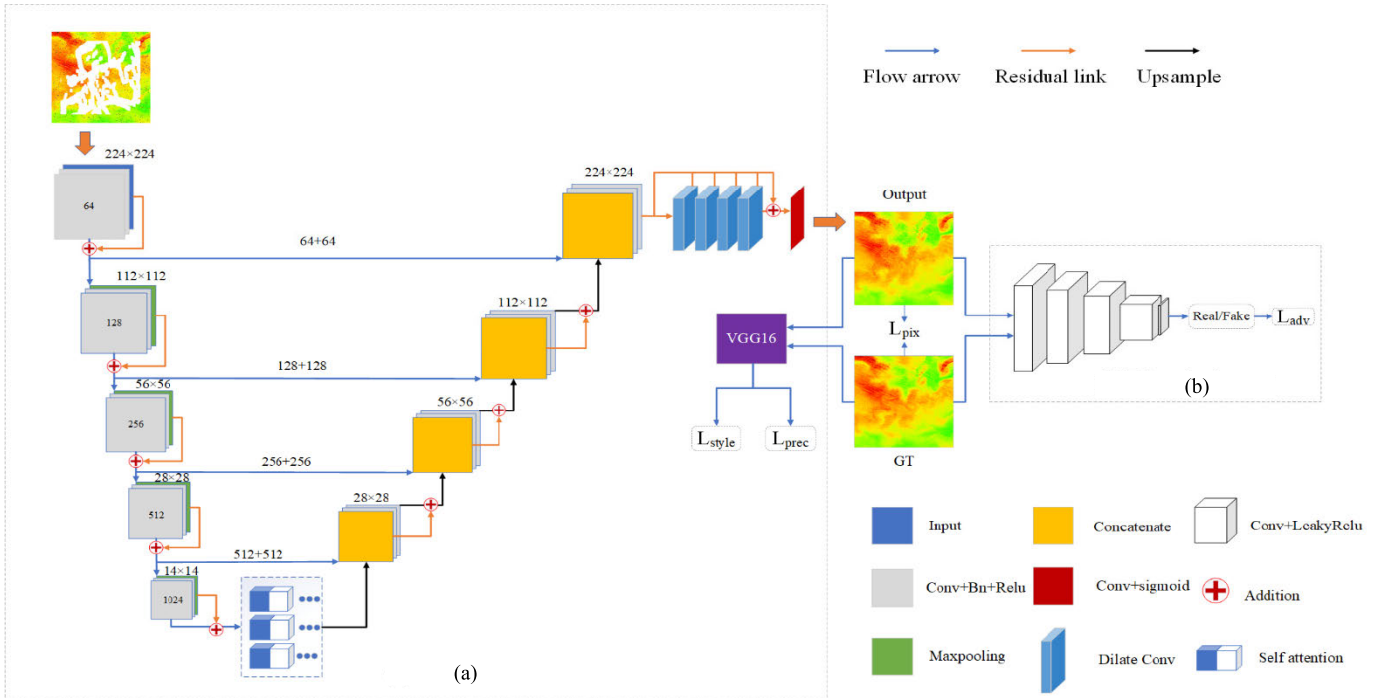
Fig. 3. Architecture of our BrGAN with two parts: (a) generator and (b) discriminator.
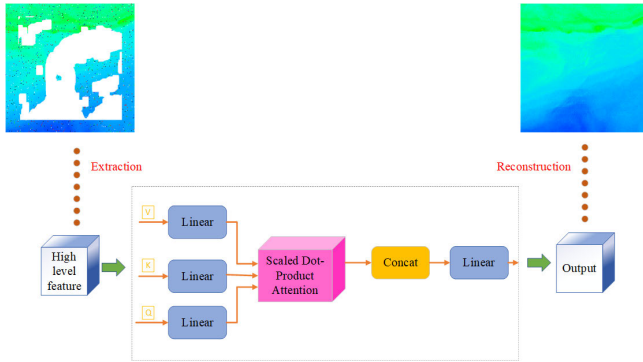

Fig. 4. Self-attention module architecture.

Fig. 5. Our designed UDCMs.

layer. Through the fusion of low-level and high-level features, the network can better retain high-resolution details in the high-level features. After that, two convolution layers, a BN layer and a ReLU layer are performed to learn effective high-resolution information in high-level features. After five groups of operations, the features are restored to the same size as the original image. Finally, a UDCM is used to enlarge the perceptual field of features and promote the integrity, the consistency, and authenticity of the restored image.

C. Self-Attention Module

As shown in Fig. 4, the self-attention module is to transfer high-level semantic information from the encoder to the decoder more effectively. Then, the decoder reconstructs from the image features and completes the final repair task. In our approach, we are inspired to apply self-attention based on channels. First, we reshaped the final output of the encoder to satisfy the input of the self-attention module and set four attention heads. Besides, three linear layers are set to satisfy the input requirements of the query feature, key feature, and
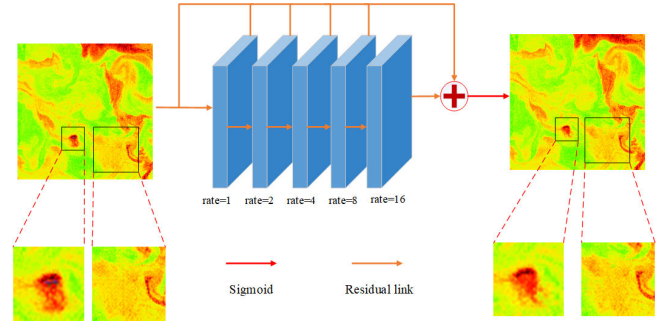
value feature. The query feature is mapped with a set of key-value feature pairs through scaled dot-product attention. In summary, the core operation of self-attention is to calculate the attention weights between query and key values, and then get the entire weights and output on the scope value.

Its calculation formula is as follows:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_n)W^o$$

(1)

$$\text{head}_i = \text{Attention}\big(QW_i^Q, KW_i^k, VW_i^v\big) \quad (2)$$

$$\text{Attention}(Q, K, V) = \text{soft max}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

where head_n represents the nth attention head, W° denotes the weight of each attention head, and $\sqrt{d_k}$ is scaling factor.

D. Union Dilated Convolution Module

Dilate convolution module (DCM) increases the perceptual range of the network through using convolutions with different dilate rates, and can retain detailed information [38]. Although pooling layer can increase the receptive field of next layer

neurons, it also results in loss of information that is difficult to recover. As shown in Fig. 5, the proposed UDCM includes five convolutional layers with different dilate rates, one BN layer, and a ReLU layer. In detail, the dilate rates of five different sizes are 1, 2, 4, 8, and 16, respectively. In addition, the kernel size, stride, and padding size of the convolutional layer are set as 4, 2, and 1, respectively. Thus, the UDCM can capture multiple larger spatial areas of restoration features with different dilate rates, benefitting from obtaining more global semantic information and enhancing the image restoration consistency. Besides, UDCM is also helpful to capture the low-frequency noise and enhance the de-noise ability of the network. To make DCMs connect to each other in the hierarchical mode, we designed a residual mechanism between two adjacent DCMs. Finally, the convolutional results were stacked, and the predictions were generated after the Sigmoid operations.

For example, when calculating the nth output feature of the dilate convolution ($F_{dn} \in \mathbb{R}^{c \times w \times h}$), the presentation can be written as follows:

$$F_{d1} = \delta_{\text{relu}}\big(\beta\big(\mathbb{C}_{3\times3}^{r=1}(F_{\text{input}})\big)\big). \tag{4}$$

$$F_{d2} = \delta_{\text{relu}}\big(\beta\big(\mathbb{C}_{3\times3}^{r=2}(F_{d1})\big)\big). \tag{5}$$

$$F_{\text{out}} = \delta_{\text{sigmoid}}\left(\sum_{j=1}^{n} F_{dn}\right). \tag{6}$$

Here, $\mathbb{C}_{3\times3}^{r=1}$ represents the convolution layer with kernel size of 3×3 and dilate rate of 1, β, δ_{relu}, and δ_{sigmoid} denote the BN layer, ReLU, and Sigmoid activation functions, respectively. F_{input} and F_{out} mean the input feature of the UDCM and the final output of the UDCM, respectively. n denotes accumulate up to the nth layer.

Through the comparison of black boxes in Fig. 5, we found that the image texture after UDCM processing is more refined and appropriate. In fact, the UDCM module can be regarded as a parallel multiresolution module to improve the model stability and reconstructed result's texture consistency.

E. Discriminator of BRGAN

In our approach, we adopted Patch GAN [39] as our GAN backbone. By distinguishing the texture structure of the real image from that of the generated image, the authenticity of the repaired image is judged, and the generator is encouraged to generate the texture structure consistent with the real image in the process of the restoration of the GOCI chlorophyll color image. The discriminator is shown in Fig. 3. Texture discriminator mainly consists of four layers of convolutions, three LeakyReLU layers, and a Sigmoid activation layer. The first three layers of convolutions are with kernel size of 4, strides of 2, and padding size of 1. The LeakyReLU activation function is used to activate neurons, and the activation function is set as negative slope of 0.2 and inplace is set as true. In the last convolution layer, the kernel size, stride, and padding size are set as 4, 1, and 1, respectively. Besides, the neurons are activated by Sigmoid function. Finally, the local discrimination of the inpainting result map is utilized to ensure the effective recovery of the low-frequency local texture structure.

F. Loss Function

The loss function of our network has four components, i.e., the pixel level loss, the perception loss, style loss, and the adversarial loss. The above union loss makes our model to have superior performances in both the visual demonstration and the quantitative analysis.

The pixel level loss can be represented as follows:

$$L_{\text{pix}} = \big\|G(I_{\text{input}}) - I_{\text{gt}}\big\|_1 \tag{7}$$

where G denotes the generator network, I_{input} and I_{gtl} mean the incomplete input image and the ground truth, respectively.

By using only the pixel level loss will lead to the mosaic phenomenon in the restored images, we added the perception loss to reinforce the learning ability for the high-level semantic information. In the implementation, we use the first three max-pooling layers outputs of the VGG16 (which was pretrained on the ImageNet) as the evaluation standards to estimate the total high-level semantic differences between the reconstructed and ground truth images. Thus, the perception loss can be written as

$$L_{\text{perceptual}} = \sum_{i=1}^{3}\big\|\varphi_i\big(G(I_{\text{input}})\big) - \varphi_i(I_{\text{gt}})\big\|_1 \tag{8}$$

where φ_i represents the output features of the ith pooling layer on the pretrained VGG16.

To make sure the restored images have same styles with the original incomplete images, we further utilize the style loss to constraint the network loss, which can be represented as

$$L_{\text{style}} = \sum_{i}\big\|\phi_i\big(G(I_{\text{input}})\big) - \phi_i\big(I_{\text{gt}}\big)\big\|_1 \tag{9}$$

where ϕ_i represents the gram matrix of the ith pooling layer on the pretrained VGG16, therefore $\phi_i(\cdot) = \varphi_i(\cdot)^T \varphi_i(\cdot)$.

Finally, to further ensure the visualization and texture consistency, we add the discriminate loss which can be written as

$$\begin{aligned} L_{\text{adversarial}} &= \min_G \max_D V(G, D) = \min_G \max_D \mathbb{E}_{I_{\text{gt}}, E_{\text{gt}}}\big[\log D\big(I_{\text{gt}}, E_{\text{gt}}\big)\big] \\ &\quad + \mathbb{E}_{I_{\text{out}}, E_{\text{out}}}\big[\log\big(1 - D\big(I_{\text{out}}, E_{\text{out}}\big)\big)\big] \end{aligned} \tag{10}$$

where D represents the discriminator of the network, I_{gt} represents the ground true, and I_{out} means the corresponds to the output of I_{input}.

Thus, our final union loss of network is designed as follows:

$$L_{\text{all}} = \lambda_1 L_{\text{pix}} + \lambda_2 L_{\text{perceptual}} + \lambda_3 L_{\text{style}} + \lambda_4 L_{\text{adversarial}} \tag{11}$$

where λ_1, λ_2, λ_3, and λ_4 are the weights of the above losses, respectively.

G. Evaluation Standards

In our experiments, we used peak signal-to-noise ratio (PSNR) and Learned Perceptual Image Patch Similarity (LPIPS) to evaluate the denoising and inpainting performance. Denote $I(i, j)$ and $K(i, j)$ as the ith row and jth column pixel values in the original and restored images, respectively. The

mean-square error (MSE) of the original and restored images can be represented as

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left[I(i, j) - K(i, j) \right]^2. \tag{12}$$

Then the PSNR can be represented as

$$\text{PSNR} = 10 \times \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) = 20 \times \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right) \tag{13}$$

where MAX_I represents the max value in the image.

And the LPIPS can be written as follows:

$$\text{Lpips}(I, K) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot \left(\hat{y}_{h,w}^l - \hat{y}_{0h,w}^l \right) \right\|_2^2. \tag{14}$$

Here, $\text{Lpips}(I, K)$ means the distance between the original true image I and the restored image K, feature stack is extracted from the ith layer, and unit regularization is carried out in channel dimension. The vector $w_l \in \mathbb{R}^{c_l}$ is used to scale the channel. Finally, $L2$ distance is calculated.

IV. EXPERIMENTS

In this section, we first introduced the dataset we made and specially designed for GOCI chlorophyll color image repair and inpainting. Then, the experimental settings and evaluation standards were illustrated. Finally, we give a detail analysis about the experimental results.

A. GOCI Denoising and Inpainting Datasets

As far as we know, there are no relevant available datasets for GOCI image denoising and inpainting application. Therefore, we collected five random images daily from NASA's official website spanning the years 2014–2019. This allowed us to create a comprehensive dataset named GOCI chlorophyll color image denoising and inpainting dataset (GCCIDID) that is well-suited for our purposes. Each original GOCI chlorophyll color image has a size of 5600×5600 pixels with a spatial resolution of 500 m. The initial data acquired by our team is in.NC format, encompassing Metadata, Flag Codings, Vector Data, Bands, Masks, and more. We employ the SeaDAS satellite remote sensing data processing software to derive oceanic chlorophyll image information, subsequently transforming it into a visual format. This serves as a foundation for subsequent stages of image refinement, filtering, and cropping processes. Fig. 6 shows a sample GOCI chlorophyll color image. It is obvious that large areas are polluted with much noise or missing lots of pixels due to the occlusion of clouds.

To guarantee that we obtain the clean data to build our dataset for training, testing, and validation, we only select areas which have little missing pixels. Besides, to make the images easier for GPU processing, we cut the original images into 224×224 patches. Thus, in detail, we only select 224×224 patches which has less than five missing pixels to build our dataset. For each original 5600×5600 image, we slide with a stride of 50 pixels to generate 224×224 patches, leading to each two adjacent patches will have much overlapping. After the selection of patches with little missing
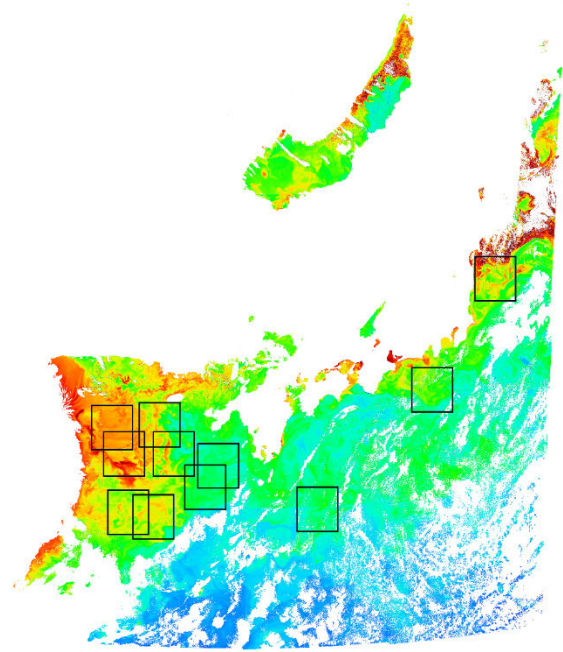

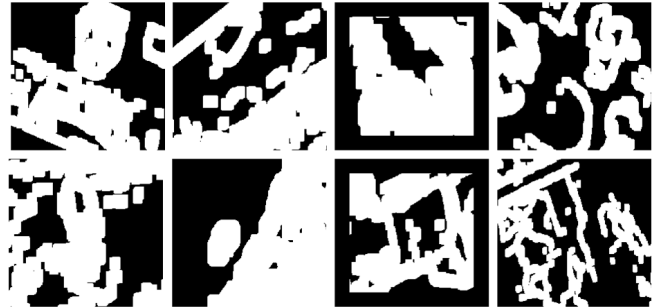Fig. 6. Sample of original GOCI chlorophyll color image (black box: the selected patches).

Fig. 7. NVIDIA irregular mask dataset.

pixels, we further check and remove the patches with much noise.

The first and second rows represent the images with noise and their corresponding original clean label images, respectively. The third and fourth rows represent the images with noise and missing pixels and their corresponding original clean label images, respectively. The fifth and sixth rows represent the images with missing pixels and their corresponding original clean label images, respectively.

We chose the original image with a little noise and a few missing pixels as the label. For the generation of training images with missing pixels, we applied mask operations as same as that in [10], [12], [40], and [41]. Fig. 7 illustrates the mask utilized in generating the dataset, sourced from the NVIDIA irregular mask dataset [10]. For the noise addition, we simulated the true noise in GOCI chlorophyll color images, majorly appearance with white, black, red, green, and blue colors. In our dataset, we further separated the images into three parts: only for denoising, only for inpainting, and for both denoising and inpainting. Finally, we generated 120 000 images for our dataset, including 60 000 training, testing, and validation images and 60 000 corresponding labels. And for our above three subdivisions, each kind has 20 000 images
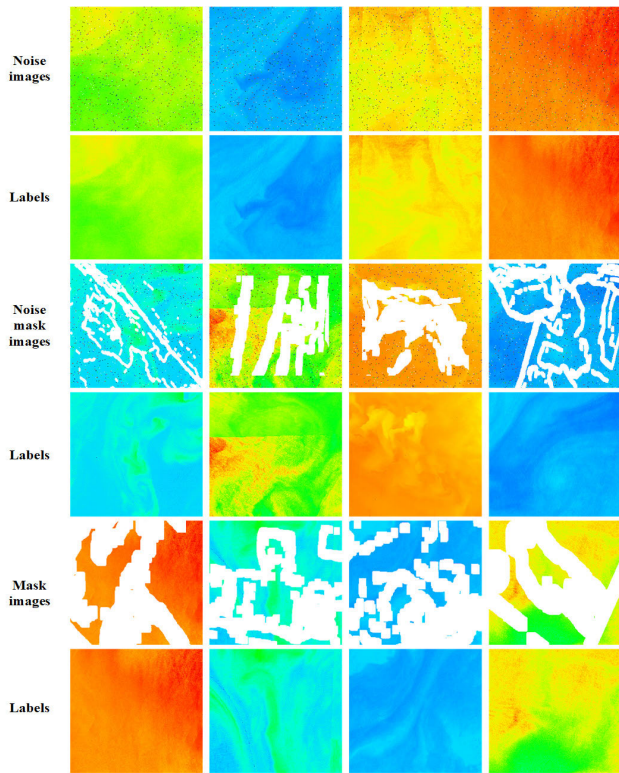
Fig. 8. Several sample images in our GCCIDID dataset.

and 20 000 corresponding labels as shown in Fig. 8. Labels in Fig. 8 represent complete ground true images, Noise images represent images with Noise, Noise Mask images represent images with both noise and missing, and Mask images represent images with missing.

In our experiments, we used about 1300 images for testing. The GCCIDIC we made for training and testing is the images in.bmp format. The memory usage of each image is about 148k, about 16.82 GB for the training dataset, and about 370 MB for the test dataset. Note that, our GCCIDID dataset is publicly available and can be obtained through the following website: http://39.98.109.195:1234/share/jUuEr7oT.

B. Experimental Settings

All of our experiments ran on a computer with Intel[1] Core[2] i9-10900K CPU @ 3.70 GHz, 128 GB memory, and an NVIDIA GeForce RTX 3090. During training, we used Adam optimizer and the learning rate decreased half when the loss did not reduce for five epochs. The initial learning rate and the batch size were set as 0.001 and 16, respectively. For our whole GCCIDID training images, we trained 100 epochs and saved the model with the best validation performance. Besides, to enhance the generalization ability, the data augmentation operations are utilized, including rotation, resize, etc., during training. It costs about 50 h for training. During the testing phase, we opted to elevate the batch size to 64, resulting in an approximately 18-s processing time for the evaluation of 1295 test images derived from the GOCI denoising and inpainting datasets. Notably, the data loading procedure

[1]Registered trademark.
[2]Trademarked.

exhibited an efficient speed of 1.05 images per second, further contributing to the overall efficiency of our testing process.

To illustrate the superior performance of our model, comparisons with other SOTA denoising and inpainting models tested on GCCIDID dataset are exhibited. The compared methods included P-Conv [10], EdgeConnect-net [40], RFR-net [12], Hypergraphs [35], Bias U-Net [42], conditional texture and structure dual generation (CTSDG) [13], and multi-level interactive siamese filtering (MISF) [41]. It needs to note that several above compared methods need to provide inpainting masks during training and testing, which was satisfied during our experiments. For justice, each model used the same experimental settings as our approach.

C. Quantitative Analysis

In this section, we quantitatively analyzed the PSNR and LPIPS performances among our method and other seven SOTA approaches verified on GCCIDID dataset. We tested for three kinds of data in GCCIDID, respectively. Besides, to authoritatively evaluate the performance scores, we used mean PSNR and mean LPIPS scores for the final comparison. The mean PSNR and mean LPIPS can be represented as follows:

$$\text{PSNR}_{\text{mean}} = \frac{1}{n} \sum_{i}^{n-1} \text{PSNR}_i. \tag{15}$$

$$\text{LPIPS}_{\text{mean}} = \frac{1}{n} \sum_{i}^{n-1} \text{LPIPS}_i. \tag{16}$$

Here, n means the number of tested images, and i means the ith image.

Table I shows the comparison results among our BrGAN and other seven SOTA approaches verified on GCCCIDID dataset. For PSNR value, the larger the better, and for LPIPS value, the smaller the better. It is obvious that our BrGAN achieved the highest PSNR values on all the three different kinds of data, obtaining the highest mean PSNR value as high as 37.06. Our mean PSNR value is higher than other approaches about 31%, 21.2%, 7.5%, 6.5%, 4.4%, 3.1%, and 6.9%. This result strongly proves the superior performance of our BrGAN model. On the other hand, our model achieved the lowest LPIPS value, as low as 0.0485. The low LPIPS value proves that our results are more consistent with human visual standards. Our mean LPIPS value is lower than other approaches about 78.1%, 41.4%, 45.3%, 32.9%, 45.8%, 18.8%, and 17.9%. This result shows that our model can generate much better results in visual, which can be clearly seen in our following visual comparisons.

D. Visualization Analysis

In this section, we analyzed the visual results of our method and other seven SOTA approaches. Fig. 9 shows several classical visual comparison results, as well as that the P-Conv obtained the worst visual results, remained holes and noise after the restoration. This visual result is consistent with the quantitative result that P-conv obtained worst PSNR and LPIPS values. The same problem also existed in the Edge-connect method. Besides, the Edge-connect model could not

TABLE I
COMPARISONS OF THE PSNR AND LPIPS VALUES

Metrics	PSNR↑				LPIPS↓			
Image types	Noise	Mask and Noise	Mask	All Test Images	Noise	Mask and Noise	Mask	All Test Images
P-Conv	27.38	27.29	29.76	28.23	0.1959	0.2295	0.191	0.2047
Edge Connect	32.07	27.9	31.59	30.58	0.0548	0.1234	0.0716	0.0827
RFR	38.07	30.67	34.69	34.48	0.0455	0.1368	0.0844	0.0886
Hyper-graphs	37.8	30.66	35.78	34.81	0.0221	0.1244	0.0706	0.0723
Bias U-Net	37.84	**32.06**	36.58	35.49	0.0308	0.1486	0.0892	0.0895
CTSDG	40.99	30.91	35.92	35.94	0.0165	0.1095	0.0542	0.0597
MISF	38.02	30.22	35.59	34.67	0.0213	0.1069	0.0505	0.0591
BRGAN	**42.27**	**32.06**	**36.86**	**37.06**	**0.0121**	**0.0918**	**0.0434**	**0.0485**

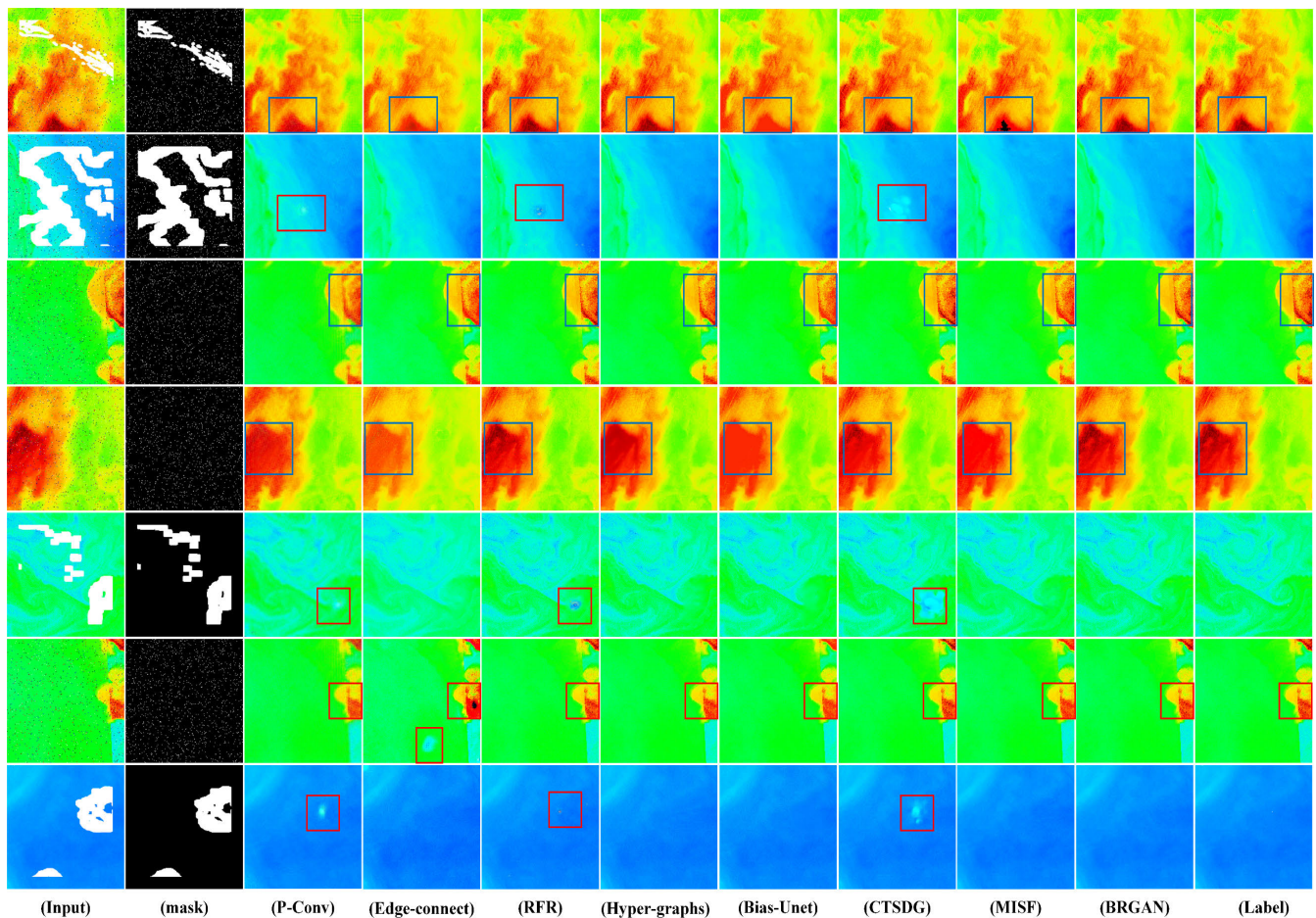| (Input) | (mask) | (P-Conv) | (Edge-connect) | (RFR) | (Hyper-graphs) | (Bias-Unet) | (CTSDG) | (MISF) | (BRGAN) | (Label) |

Fig. 9. Visual recovered comparisons among our BrGAN and other seven outstanding approaches evaluated on GOCI chlorophyll color images. The figure above denotes the input images, the mask of input images, results of P-Conv, Edge-connect, RFR, Hypergraphs, Bias U-Net, CTSDG, MISF, and our BrGAN, labels, respectively.

recover the red colors well. For RFR-net, the generated results showed that the pixel was confusing in some areas, leading to uncleared visualizations. The major weakness of hypergraphs and Bias U-Net usually generated more pale red colors than other approaches, especially for the original deep red areas. For CTSDG, some color confusion phenomenon occurs after the reconstruction. The MISF recovered well for textures of missing areas, but it got undesirable results for pixel values
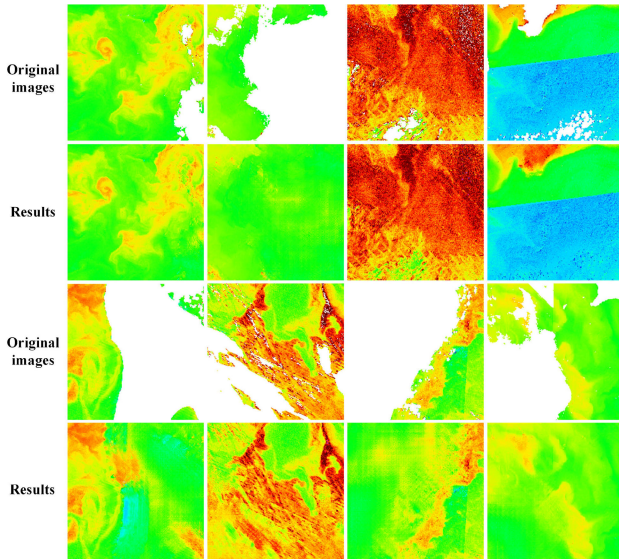
Fig. 10. Visual recovering results of our BrGAN verified on GOCI chlorophyll color images (odd and even rows denote the original incomplete targets and the recovered results, respectively).

TABLE II
INFLUENCE OF HYPERPARAMETERS IN OUR EXPERIMENT

Metrics	PSNR↑	LPIPS↓	hyper-parameter			
Image types	All Test Images	All Test Images	λ_1	λ_2	λ_3	λ_4
Original ResUnet	36.93	0.0763	10	\	\	\
BrGAN (1)	37.76	0.0727	10	\	\	\
BrGAN (2)	**37.80**	0.0547	10	0.1	\	\
BrGAN (3)	37.00	0.0495	10	0.1	250	\
BrGAN (4)	37.05	**0.0485**	10	0.1	250	0.001

TABLE III
ABLATION ANALYSIS

Metrics	PSNR↑	LPIPS↓
Image types	All Test Images	All Test Images
ResUnet-GAN128(our)	36.60	0.0510
BrGAN (w/o UDCM)	36.88	0.0505
BrGAN (w/o SA)	36.91	0.0499
BrGAN	**37.05**	**0.0485**

within missing areas. Thereafter, MISF demonstrated a worse performance for noises during the comparisons. In contrast, the recovered results by our approach demonstrated better effects in the aspects of noise removal, texture reconstruction, and color balance.

In addition, we verified our BrGAN on the actual GOCI chlorophyll color images with missing parts. The recovering results are as demonstrated in Fig. 10, showing satisfactory visual effects for dealing with targets having a large proportion of losing areas. The good performance convincingly demonstrated the superior recovering ability of our BrGAN for GOCI chlorophyll color images.

The illustrated outcomes in Fig. 10 undeniably accentuate the remarkable proficiency of our BrGAN in tackling the intricacies of images fraught with substantial data gaps. This remarkable display not only reaffirms the resilience of our methodology but also furnishes compelling substantiation of the unparalleled data recuperation potential our BrGAN introduces to GOCI chlorophyll color images. By adeptly bridging information voids with coherent and visually pleasing content, our model showcases its ability to elevate the overall quality of such images, even amidst the challenge of considerable data fragmentation.

E. Ablations and Hyperparameters Analysis

In this section, we analyzed the influence of hyperparameters in BrGAN. In addition, we verified the effectiveness of attention and dilated convolution modules.

The traditional U-Net model has shown the superiority of this structure in various studies [13], [38], [42], and the structure has been improved. We have added the residual mechanism in the encoder and decoder on the U-Net structure, which has the same application in some medical image segmentation studies [43].

As shown in Table II, ↑ denotes the bigger the better, ↓ denotes the smaller the better. BrGAN (1, 2, 3, and 4) represent the use of different hyperparameters, respectively, the repair effect of Original ResUnet in the GOCI image

reached 36.93 (PSNR). The use of a single pixel-level loss has a strong ability in image noise removal, as shown in Fig. 11(a). However, the performance in the restoration of missing images is not satisfactory, as shown in Fig. 11(b) and (c), which usually results in the fuzzy phenomenon of missing images and relatively large LPIPS, up to 0.0763 (LPIPS). Therefore, it is not possible to use a single PSNR to evaluate the repair effect of the model, but also needs to judge the repair ability of the model through LPIPS. Inspired by CTSDG, MISF, and other facial and architectural restorations, we enhanced the constraint ability of the model by adding some losses, such as perceptual loss, style loss, and adversarial loss.

In order to design the most appropriate hyperparameters in the GOCI chlorophyll color images, we conducted several restoration experiments, and finally designed a group of optimal hyperparameters. As shown in Fig. 11, the experimental restoration effect has also been significantly improved. We also analyzed the influence of adding the self-attention module and UDCM module, respectively, on the repair effect of the model, as shown in Fig. 12. After adding the self-attention module and UDCM module, the missing image could be effectively repaired. Experiments show the effectiveness and stability of these two modules. Table III shows the ablation analysis details and the hyperparameters of $\lambda1$, $\lambda2$, $\lambda3$, and $\lambda4$ used in the experiment are 10, 0.1, 250, and 0.001, respectively. The (w/o SA) and (w/o UDCM) mean without self-attention model and without DCM. As shown in Table III, the PSNR and LPIPS of our results after the addition of two proposed modules have also been significantly improved. After adding two modules at the same time, the PSNR and LPIPS of our model reached 37.05 and 0.0485, respectively. It indicates that the complete model proposed by us has superior performance in noise removal and restoration of GOCI chlorophyll color images.
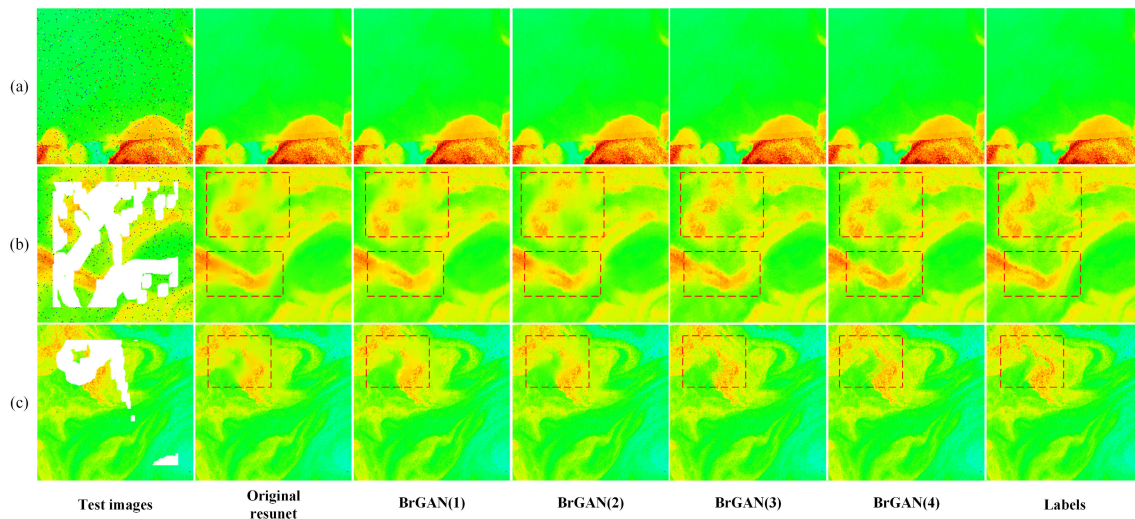
Fig. 11. Results of our model are shown on the test image after using different hyperparameters (red box: the area in the image with considerable difference).
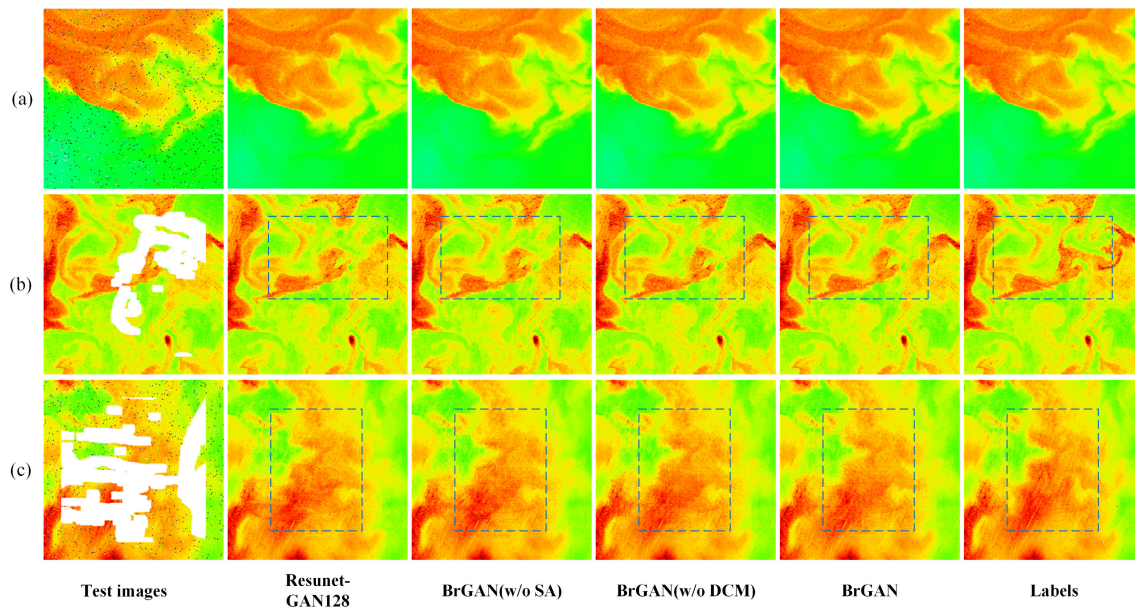

Fig. 12. Results of the ablation experiment (blue box: the area in the image with considerable difference).

V. CONCLUSION

In this article, we first build a publicly open dataset named GCCIDID that is the first GOCI chlorophyll color image restoration dataset as far as we know. Then, we proposed a GOCI chlorophyll color image restoration network named BRGAN that innovations majorly lie on a novel attention module and a superior UDCM. In BrGAN, no more assistant mask information is necessary. The attention module guides the network which focuses on areas that providing high-level semantic information and guaranteeing the texture consistency of recovered results. To further reinforce the high-level semantic information learning ability, we proposed a union dilated convolution to improve the noise distinguish ability and strengthen the digging ability to discover the relationships among incomplete and existed areas. In addition, we adopted a GAN module to further improve the color and texture consistency of recovered images. Finally, BrGAN designed a loss including the pixel level loss, the perception loss, style loss, and the discrimination loss. The effectiveness of the

designed union loss was demonstrated through our ablation experiments. Also, we compared BrGAN with other SOTA approaches evaluated on GCCIDID dataset, illustrating the superiority of our BrGAN that achieved PSNR and LPIPS as good as 37.05 and 0.0485. Our mean PSNR value is higher than other approaches about 31%, 21.2%, 7.5%, 6.5%, 4.4%, 3.1%, and 6.9%. Our mean LPIPS value is lower than other approaches about 78.1%, 41.4%, 45.3%, 32.9%, 45.8%, 18.8%, and 17.9%. We also visualized restoration results for analysis. Both the quantitative and visualization comparisons convincingly illustrate the superiority of the proposal.

REFERENCES

[1] X. Liu and M. Wang, "Analysis of ocean diurnal variations from the Korean geostationary ocean color imager measurements using the DINEOF method," *Estuarine, Coastal Shelf Sci.*, vol. 180, pp. 230–241, Oct. 2016, doi: 10.1016/j.ecss.2016.07.006.

[2] J.-M. Yeom, J.-L. Roujean, K.-S. Han, K.-S. Lee, and H.-W. Kim, "Thin cloud detection over land using background surface reflectance based on the BRDF model applied to geostationary ocean color imager (GOCI) satellite data sets," *Remote Sens. Environ.*, vol. 239, Mar. 2020, Art. no. 111610, doi: 10.1016/j.rse.2019.111610.

[3] Y. Zhou et al., "Monitoring multi-temporal and spatial variations of water transparency in the Jiaozhou bay using GOCI data," *Mar. Pollut. Bull.*, vol. 180, Jul. 2022, Art. no. 113815, doi: 10.1016/j.marpolbul.2022.113815.

[4] Y. Kang, M. Kim, E. Kang, D. Cho, and J. Im, "Improved retrievals of aerosol optical depth and fine mode fraction from GOCI geostationary satellite data using machine learning over east Asia," *ISPRS J. Photogramm. Remote Sens.*, vol. 183, pp. 253–268, Jan. 2022, doi: 10.1016/j.isprsjprs.2021.11.016.

[5] J. Wu et al., "Estimates of diurnal and daily net primary productivity using the geostationary ocean color imager (GOCI) data," *Remote Sens. Environ.*, vol. 280, Oct. 2022, Art. no. 113183, doi: 10.1016/j.rse.2022.113183.

[6] C. Feng, S. Wang, and Z. Li, "Long-term spatial variation of algal blooms extracted using the U-Net model from 10 years of GOCI imagery in the East China Sea," *J. Environ. Manage.*, vol. 321, Nov. 2022, Art. no. 115966, doi: 10.1016/j.jenvman.2022.115966.

[7] S. Wang, X. Zhang, N. Chen, and W. Wang, "Classifying diurnal changes of cyanobacterial blooms in Lake Taihu to identify hot patterns, seasons and hotspots based on hourly GOCI observations," *J. Environ. Manage.*, vol. 310, May 2022, Art. no. 114782, doi: 10.1016/j.jenvman.2022.114782.

[8] X. Ding et al., "Using geostationary satellite ocean color data and superpixel to map the diurnal dynamics of water transparency in the eastern China seas," *Ecolog. Indicators*, vol. 142, Sep. 2022, Art. no. 109219, doi: 10.1016/j.ecolind.2022.109219.

[9] Z. Hu et al., "Characterizing surface circulation in the Taiwan strait during NE monsoon from geostationary ocean color imager," *Remote Sens. Environ.*, vol. 221, pp. 687–694, Feb. 2019, doi: 10.1016/j.rse.2018.12.003.

[10] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2018, pp. 89–105, doi: 10.1007/978-3-030-01252-6_6.

[11] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4470–4479, doi: 10.1109/ICCV.2019.00457.

[12] J. Li, N. Wang, L. Zhang, B. Du, and D. Tao, "Recurrent feature reasoning for image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7757–7765, doi: 10.1109/CVPR42600.2020.00778.

[13] X. Guo, H. Yang, and D. Huang, "Image inpainting via conditional texture and structure dual generation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 14114–14123, doi: 10.1109/ICCV48922.2021.01387.

[14] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "RePaint: Inpainting using denoising diffusion probabilistic models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11451–11461, doi: 10.1109/CVPR52688.2022.01117.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural. Inf. Process. Syst.*, vol. 25, May 2012, pp. 1–9, doi: 10.1145/3065386.

[16] C. Hsu, F. Chen, and G. Wang, "High-resolution image inpainting through multiple deep networks," in *Proc. Int. Conf. Vis., Image Signal Process. (ICVISP)*, Sep. 2017, pp. 76–81, doi: 10.1109/ICVISP.2017.27.

[17] W. Wang and Y. Jia, "Damaged region filling and evaluation by symmetrical exemplar-based image inpainting for Thangka," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, pp. 1–13, Jun. 2017, doi: 10.1186/s13640-017-0186-1.

[18] J. Peng, D. Liu, S. Xu, and H. Li, "Generating diverse structure for image inpainting with hierarchical VQ-VAE," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10770–10779, doi: 10.1109/CVPR46437.2021.01063.

[19] Q. Liu et al., "Reduce information loss in transformers for pluralistic image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11337–11347, doi: 10.1109/CVPR52688.2022.01106.

[20] L. Liao, J. Xiao, Z. Wang, C.-W. Lin, and S. Satoh, "Image inpainting guided by coherence priors of semantics and textures," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 6535–6544, doi: 10.1109/CVPR46437.2021.00647.

[21] W. Li, Z. Lin, K. Zhou, L. Qi, Y. Wang, and J. Jia, "MAT: Mask-aware transformer for large hole image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10748–10758, doi: 10.1109/CVPR52688.2022.01049.

[22] Z. Qin, Q. Zeng, Y. Zong, and F. Xu, "Image inpainting based on deep learning: A review," *Displays*, vol. 69, Sep. 2021, Art. no. 102028, doi: 10.1016/j.displa.2021.102028.

[23] Y. Wei and S. Liu, "Domain-based structure-aware image inpainting," *Signal, Image Video Process.*, vol. 10, no. 5, pp. 911–919, Jan. 2016, doi: 10.1007/s11760-015-0840-y.

[24] D. Zhang, Z. Liang, G. Yang, Q. Li, L. Li, and X. Sun, "A robust forgery detection algorithm for object removal by exemplar-based image inpainting," *Multimedia Tools Appl.*, vol. 77, no. 10, pp. 11823–11842, May 2017, doi: 10.1007/s11042-017-4829-0.

[25] S. M. Muddala, R. Olsson, and M. Sjöström, "Spatio-temporal consistent depth-image-based rendering using layered depth image and inpainting," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, pp. 1–19, Feb. 2016, doi: 10.1186/s13640-016-0109-6.

[26] N. Kawai, T. Sato, and N. Yokoya, "Diminished reality based on image inpainting considering background geometry," *IEEE Trans. Vis. Comput. Graphics*, vol. 22, no. 3, pp. 1236–1247, Mar. 2016, doi: 10.1109/TVCG.2015.2462368.

[27] H. Lu, Q. Liu, M. Zhang, Y. Wang, and X. Deng, "Gradient-based low rank method and its application in image inpainting," *Multimedia Tools Appl.*, vol. 77, no. 5, pp. 5969–5993, Mar. 2018, doi: 10.1007/s11042-017-4509-0.

[28] K. H. Jin and J. C. Ye, "Annihilating filter-based low-rank Hankel matrix approach for image inpainting," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3498–3511, Nov. 2015, doi: 10.1109/TIP.2015.2446943.

[29] T. Ružic and A. Pižurica, "Context-aware patch-based image inpainting using Markov random field modeling," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 444–456, Jan. 2015, doi: 10.1109/TIP.2014.2372479.

[30] C. Zheng, T.-J. Cham, and J. Cai, "Pluralistic image completion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1438–1447, doi: 10.1109/CVPR.2019.00153.

[31] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, "Shift-Net: Image inpainting via deep feature rearrangement," in *Proc. Eur. Conf. Comput. Vis*, Sep. 2018, pp. 1–17, doi: 10.1007/978-3-030-01264-9_1.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[33] H. Liu, B. Jiang, Y. Song, W. Huang, and C. Yang, "Rethinking image inpainting via a mutual encoder–decoder with feature equalizations," in *Proc. 16th Eur. Conf. Comput. Vis.* U.K.: Springer, Aug. 2020, pp. 725–741, doi: 10.1007/978-3-030-58536-5_43.

[34] H. Liu, B. Jiang, Y. Xiao, and C. Yang, "Coherent semantic attention for image inpainting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 4169–4178, doi: 10.1109/ICCV.2019.00427.

[35] G. Wadhwa, A. Dhall, S. Murala, and U. Tariq, "Hyperrealistic image inpainting with hypergraphs," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3911–3920, doi: 10.1109/WACV48630.2021.00396.

[36] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456, doi: 10.5555/3045118.3045167.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[38] L. Zhou, C. Zhang, and M. Wu, "D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 192–1924, doi: 10.1109/CVPRW.2018.00034.

[39] 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. (CVPR)*, Jul. 2017, pp. 5967–5976, doi: 10.1109/CVPR.2017.632.

[40] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "Edge-Connect: Structure guided image inpainting using edge prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3265–3274, doi: 10.1109/ICCVW.2019.00408.

[41] X. Li, Q. Guo, D. Lin, P. Li, W. Feng, and S. Wnag, "MISF: Multi-level interactive Siamese filtering for high-fidelity image inpainting," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 1869–1878, doi: 10.1109/CVPR52688.2022.00191.

[42] Z. Chen, C. Wang, J. Li, N. Xie, Y. Han, and J. Du, "Reconstruction bias U-Net for road extraction from optical remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2284–2294, 2021, doi: 10.1109/JSTARS.2021.3053603.

[43] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted Res-UNet for high-quality retina vessel segmentation," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Oct. 2018, pp. 327–331, doi: 10.1109/ITME.2018.00080.

Ziyi Chen (Member, IEEE) received the Ph.D. degree in signal and information processing from Xiamen University, Xiamen, China, in 2016.

He is currently an Associate Professor with the Department of Computer Science and Technology, Huaqiao University, Xiamen. His research interests include computer vision, machine learning, and remote sensing image processing.

Yuhua Luo is currently pursuing the M.Sc. degree with the School of Computer Science and Technology, Huaqiao University, Xiamen, China.

His research interests include deep learning and remote sensing image processing.

Yiping Chen (Senior Member, IEEE) received the Ph.D. degree in information and communications engineering from the National University of Defense Technology, Changsha, China, in 2011.

From 2007 to 2011, she was an Assistant Researcher with The Chinese University of Hong Kong, Hong Kong. She is currently an Associate Professor with the School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai, China. She has coauthored more than 80 publications, including in top remote sensing journal such as IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, International Society for Photogrammetry and Remote Sensing (ISPRS), *Journal of Photogrammetry and Remote Sensing* (JPRS), and *International Journal of Applied Earth Observation and Geoinformation* (JAG), and in flagship Computer Vision (CV), Artificial Intelligence (AI) conferences such as IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Association for the Advancement of Artificial Intelligence (AAAI), and International Joint Conference on Artificial Intelligence (IJCAI). Her main research interests include image processing, mobile LiDAR data analysis, and GeoAI.

Dr. Chen is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the Co-Chair of the ISPRS WG I/4 on LiDAR, Laser Altimetry and Sensor Integration (2022–2026).

Jing Wang received the Ph.D. degree from the Department of Mathematics, Zhejiang University, Hangzhou, China, in 2006.

He is currently a Professor with the School of Computer Science and Technology, Huaqiao University, Xiamen, China. His research interests include machine learning, recommendation system, and matrix computation.

Dilong Li (Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2020.

He is currently an Assistant Professor with the College of Computer Science and Technology, Huaqiao University, Xiamen, China. He has authored or coauthored over 30 research articles published in refereed journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, International Society for Photogrammetry and Remote Sensing (ISPRS) *Journal of Photogrammetry and Remote Sensing*, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, and *Remote Sensing*. His research interests include 3-D computer vision, deep learning, and intelligent interpretation of point clouds.

Kyle Gao (Graduate Student Member, IEEE) received the B.Sc. degree in mathematics from the University of Waterloo, Waterloo, ON, Canada, in 2016, and the M.Sc. degree in physics from the University of Victoria, Victoria, BC, Canada, in 2020. He is currently pursuing the Ph.D. degree in systems design engineering with the Geospatial Intelligence and Mapping Group, University of Waterloo.

He has published articles in the *International Journal of Applied Earth Observation and Geoinformation*, IEEE ACCESS, and IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS. His research interests include computer vision and deep learning.

Cheng Wang (Senior Member, IEEE) received the Ph.D. degree in signal and information processing from the National University of Defense Technology, Changsha, China, in 2002.

He is currently a Professor with the School of Informatics, Xiamen University, Xiamen, China, where he is the Executive Director with the Fujian Key Laboratory of Sensing and Computing for Smart Cities. He has coauthored more than 150 articles in refereed journals including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Pattern Recognition*, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and International Society for Photogrammetry and Remote Sensing (ISPRS) *Journal of Photogrammetry and Remote Sensing*, and top conferences such as IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Association for the Advancement of Artificial Intelligence (AAAI), International Joint Conference on Artificial Intelligence (IJCAI), IEEE International Geoscience and Remote Sensing Society (IGARSS), and ISPRS. His research interests include point cloud analysis, multisensor fusion, mobile mapping, and geospatial big data.

Jonathan Li (Fellow, IEEE) received the Ph.D. degree in geomatics engineering from the University of Cape Town, Cape Town, South Africa, in 2000.

He is currently a Professor of geomatics and systems design engineering with the University of Waterloo, Waterloo, ON, Canada. He has coauthored over 600 publications, more than 150 of which were published in top remote sensing journals, including *Remote Sensing of Environment*, International Society for Photogrammetry and Remote Sensing (ISPRS) *Journal of Photogrammetry and Remote Sensing*, *International Journal of Applied Earth Observation and Geoinformation* (JAG), and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He has also published papers in flagship conferences in computer vision and AI, including IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Association for the Advancement of Artificial Intelligence (AAAI), and International Joint Conference on Artificial Intelligence (IJCAI). He has supervised nearly 200 master's/Ph.D. students as well as post-doctoral fellows/visiting scholars to completion. His main research interests include AI-based information extraction from earth observation images and LiDAR point clouds, photogrammetry, 3-D vision and GeoAI for digital twin cities, and autonomous vehicles.

Dr. Li is a fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada. He is the Editor-in-Chief of JAG and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.