ISPRS Journal of Photogrammetry and Remote Sensing 171 (2021) 266–277

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Missing data reconstruction in VHR images based on progressive structure prediction and texture generation

Hanwen Xu [a], Xinming Tang [b], Bo Ai [a,*], Xiaoming Gao [b], Fanlin Yang [a], Zhen Wen [a]

[a] Shandong University of Science and Technology, Qingdao, China
[b] Land Satellite Remote Sensing Application Center, Beijing, China

ARTICLE INFO

Keywords:
Missing data reconstruction
VHR images
Progressive structure prediction
Texture generation
Deep learning

ABSTRACT

Very high resolution (VHR) satellite and aerial images often suffer from scene occlusion caused by redundant objects. The task of removing these redundant objects can be solved by missing data reconstruction technology. However, when dealing with VHR images with large-scale missing regions, existing spatial-based methods often destroy the structural information of ground objects. To alleviate this problem, this paper proposes a novel missing data reconstruction method based on deep learning. The reconstruction process is divided into two parts: structure prediction and texture generation. First, a progressive edge generation network (PEGN) is designed to predict the edges of objects in missing regions in a progressive manner. Then, the edge map predicted by PEGN is input to a texture generation network (TGN) as structural information to produce the reconstruction results. This is a spatial-based method that can produce realistic and reasonable results without any need for auxiliary spectral or temporal data. Experiments demonstrate that our model can better restore the structure of ground objects in VHR images than other spatial-based methods and outperform them in SSIM and PSNR indices. In addition, our model also has a strong generalization capability by introducing Poisson blending and histogram matching.

1. Introduction

Abundant very high resolution (VHR) satellite and aerial images are generated from different sensors and platforms. Benefiting from their high resolution, VHR images have been used for various applications like scene recognition, object detection, and land use classification. However, these images often suffer from scene occlusions caused by redundant objects as shown in Fig. 1, which will affect their subsequent applications. Therefore, the question of how to remove these redundant objects quickly and thoroughly to restore the original appearances of VHR images has become an important topic. This is a task that can be solved by missing data reconstruction technology.

Shen et al. (2015a) divided methods for reconstructing the missing data in remote sensing images into four categories: spatial-based methods, spectral-based methods, temporal-based methods, and hybrid methods. This paper focuses on the spatial-based methods which, in the field of computer vision, are known as image inpainting. Spatial-based methods are guided by the assumption that the data in missing and remaining regions share the same statistical or geometrical structures (Guillemot and Le Meur, 2013). These methods utilize the data in remaining regions to fill the missing regions without any need for complementary spectral or temporal data. However, as uncertainty and error accumulate along with propagation, the spatial-based methods can hardly deal with large-scale missing regions. One of the goals of this paper is to implement a method that can handle both small and large missing regions at the same time.

In general, the results of spatial-based reconstruction methods should meet two requirements: reasonable object structure and realistic textural details. However, it is difficult to reach these two requirements simultaneously. Additionally, the structure information of VHR images is more obvious and complex than that of low-resolution images, which makes the reconstruction more difficult. At present, well-performed spatial-based large-scale missing data reconstruction methods include exemplar (patch)-based methods and deep learning-based methods. Exemplar-based methods (Barnes et al., 2009; Li et al., 2016; Criminisi et al., 2004; Lorenzi et al., 2011; Chen et al., 2005), such as the classic image inpainting method PatchMatch (Barnes et al., 2009), were first proposed in the field of digital image processing. Recently, many excellent exemplar-based methods (Li et al., 2016; Lorenzi et al., 2011; Chen et al., 2005) have been proposed for remote sensing image reconstruction. These methods can recover the texture of missing regions effectively, but they have no concept of visual semantics

* Corresponding author at: College of Geodesy and Geomatics, Shandong University of Science and Technology, 266590 Qingdao, Shandong, China.
E-mail addresses: hanwen5xu@gmail.com (H. Xu), aibo@sdust.edu.cn (B. Ai).

https://doi.org/10.1016/j.isprsjprs.2020.11.020
Received 11 July 2020; Received in revised form 18 November 2020; Accepted 21 November 2020
Available online 9 December 2020

(a) (b) (c)

Fig. 1. VHR images with redundant objects. The definition of redundant objects in this paper is unrestricted. Any objects that are uninteresting or cause scene occlusion can be regarded as redundant objects. (a) These planes can be regarded as extraneous objects when they are not public. (b) These cars are not needed if the goal is a clean scene. (c) Image with map annotations.

information of ground objects and are too dependent on the available image statistics. If there is no data in the remaining regions that matches the missing regions, the reconstruction results will collapse.

In the last few years, deep learning has attracted the attention of researchers because of the increase of available data and computing power. The basic framework by which deep learning solves missing data reconstruction problems is taking the images with missing data as inputs and the complete images as labels, then training a neural network to perform the reconstruction. Numerous deep learning-based methods (Pathak et al., 2016; Iizuka et al., 2017; Liu et al., 2018; Yu et al., 2019; Yu et al., 2018; Nazeri et al., 2019; Xiong et al., 2019; Zhang et al., 2018; Xu et al., 2019) have been proposed for missing data reconstruction. In early papers (Pathak et al., 2016; Iizuka et al., 2017; Liu et al., 2018; Yu et al., 2019; Yu et al., 2018), the reconstruction process is usually performed in only one convolutional neural network (CNN). Researchers are committed to designing modified CNN architectures, such as partial convolution (Liu et al., 2018) and gated convolution (Yu et al., 2019), to improve reconstruction performance. However, the reconstruction task may be too burdensome to complete in a single network, hence many methods of combining auxiliary information have been proposed. This auxiliary information includes edges (Nazeri et al., 2019; Xiong et al., 2019), spectral (Zhang et al., 2018), flow (Xu et al., 2019), etc. In contrast to exemplar-based methods, deep learning-based methods can predict the structural of objects in missing regions and directly generate the reconstruction results via an end-to-end form.

In this paper, a novel deep learning-based missing data reconstruction is proposed for VHR images with large-scale missing regions. Like Nazeri et al. (2019) and Xiong et al. (2019), we use edges as structural constraints to improve the authenticity of reconstruction results. To adapt to large-scale edge prediction in VHR images, a progressive edge generation network (PEGN) is proposed by analyzing various edge prediction methods. Meanwhile, a texture generation network (TGN) is designed to finish the texture generation task. In summary, the proposed method has two parts: structure prediction and texture generation, as shown in Fig. 2. Structure prediction is used to predict the structure of objects in missing regions, while texture generation is used to generate textures according to the predicted structure. Our main contributions can be summarized as follows:

1) The reconstruction process is divided into two sub-tasks: structure prediction and texture generation, which reduces the burden of each network and realizes the decoupling of structure and texture.
2) We analyze the characteristics of four edge prediction methods and propose the PEGN based on these characteristics. Compared with other methods, PEGN is more suitable for structure prediction in VHR images with large-scale missing regions.
3) Qualitative and quantitative experiments on the ISPRS Vaihingen dataset shows that our method outperforms other frequently-used spatial-based methods in SSIM (Wang et al., 2004) and PSNR indices. In addition, our model also has a strong generalization capability by introducing Poisson blending (Pérez et al., 2003) and histogram matching.

2. Related work

2.1. Traditional spatial-based methods

Different from spectral-based (Wang et al., 2006; Shen et al., 2010; Shen et al., 2013) and temporal-based (Li et al., 2014; Zeng et al., 2013; Li et al., 2019) methods requiring other clear and complete band of data, spatial-based methods can fill the missing regions directly without any complementary data. Scholars proposed numerous spatial-based methods before the rapid development of deep learning technology. For example, many variation-based methods (Shen and Zhang, 2008; Cheng et al., 2013), propagated diffusion methods (Maalouf et al., 2009; Mendez-Rial et al., 2011) and interpolation-based methods (Zhang et al., 2007) were proposed for reconstructing small missing regions. If the missing regions get larger, however, the results of these methods tend to get blurry. Therefore, exemplar (patch)-based methods were developed for reconstructing large-scale missing regions. Based on the low-level

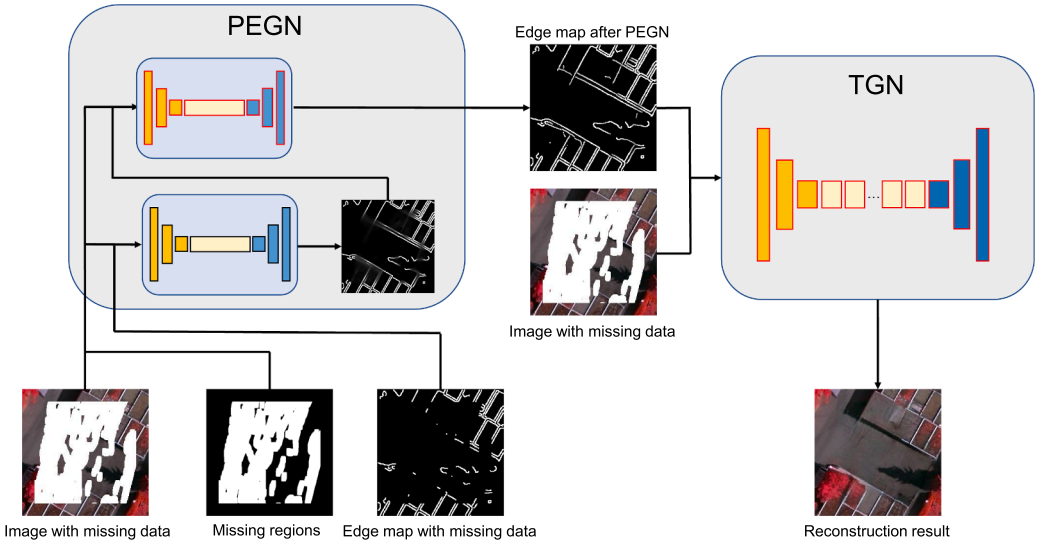


Fig. 2. Overview of our method. First, PEGN is used to predict the structure of objects in missing regions by a progressive manner. Then, TGN is used to generate texture based on the predicted structure to complete the reconstruction.

features of images, exemplar-based methods find patches that best match the missing regions in the remaining regions and copy them to the missing regions. Chen et al. (2005) presented an improved fast fragment-based image completion method to remove clouds and shadows in high-resolution remote sensing images. Barnes et al. (2009) used a randomized algorithm to quickly find the best fitting patches to fill the missing regions. Lorenzi et al. (2011) presented three different solutions to reconstruct VHR images by propagating the spectro-geometrical information retrieved from the remaining parts. Li et al. (2016) proposed a patch matching-based multitemporal group sparse representation method for reconstructing missing data in remote sensing images. The exemplar-based methods are commonly used in reconstruction tasks with large-scale missing regions. However, exemplar-based methods often produce disorganized results because of the no concept of objects semantic information.

2.2. Deep learning-based methods

The emergence of deep learning has inspired recent works on designing various neural network architectures for missing data reconstruction task. Among them, generative adversarial network (GAN) (Goodfellow et al., 2014) and its derivatives have shown remarkable advancements in the field of image generation in recent years. It has ability to simulate the semantic characteristics of objects and generate realistic images, which can make up for the shortcomings of exemplar-based methods. Therefore, GAN was naturally introduced into the field of missing data reconstruction. Pathak et al. (2016) proposed Context Encoder, who introduced deep learning and GAN into the field of image inpainting for the first time. Their results, however, were blurry and the size of input images were limited. To alleviate these problems, Iizuka et al. (2017) divided the loss function into global and local loss, enabling the model to handle higher resolution images. Subsequently, Liu et al. (2018) and Yu et al. (2019) presented partial convolution and gated convolution, respectively, which made the convolution operation more suitable for image inpainting tasks.

Later, some methods further improved the reconstruction accuracy by introducing auxiliary information. Nazeri et al. (2019) and Xiong et al. (2019) used edges as structural information to make the reconstruction results more reasonable. Zhang et al. (2018) developed a unified deep CNN model combined with spatial–temporal-spectral as supplementary information. Furthermore, Xu et al. (2019) proposed a flow-guided video inpainting approach where video inpainting was considered as a pixel propagation problem. Experiment results in these papers show that adding auxiliary information does improve the accuracy of reconstruction. We also use edges as auxiliary information to enhance the authenticity of the reconstruction results. But unlike Nazeri et al. (2019) and Xiong et al. (2019), our model generates edges in a progressive manner to accommodate large-scale edge prediction in VHR images.

2.3. Generative adversarial network

As an implicit generative model (Mohamed and Lakshminarayanan, 2016), GAN can directly capture the data distribution via adversarial learning without designing a specific likelihood function, which gives GAN the ability to generate realistic images. GAN consists of a generator G and a discriminator D. Intuitively, the G is trained to generate images to confuse the D, while the D is trained to distinguish the images generated by G from the real ones. Mathematically, the adversarial learning can be formulated as a two-player minimax game with the following value function $V(D,G)$:

$$\begin{aligned}&\min_G \max_D V(D,G)\\&= E_{x\sim q_{data}}[\log D(x)] + E_{x'\sim q_g}[\log(1-D(x'))]\end{aligned} \tag{1}$$

where q_{data} is the true data distribution, and q_g is the data distribution generated by G. $V(D,G)$ is to evaluate the distance between q_{data} and q_g, which contains many forms. Formulation (1) is the original form Goodfellow et al. (2014) that is equivalent to reducing the Jensen-Shannon divergence between q_{data} and q_g.

Original GAN has ability to generate realistic objects, but its training process is unstable due to the mechanism of adversarial learning which makes it difficult to produce expected results. Arjovsky et al. (2017) proposed a method called Wasserstein GAN (WGAN) to stabilize the adversarial learning. It replaced Jensen–Shannon divergence with Wasserstein distance and limited the gradient of the discriminator to satisfy the 1-Lipschitz condition by weight clipping. Gulrajani et al. (2017) abandoned the weight clipping of WGAN, and instead used a linear interpolation between the generated data and the target data to construct the gradient penalty. The above two methods optimize the adversarial training from the perspective of $V(D,G)$. Another example is the spectrally normalized GAN (SN-GAN) proposed by Miyato et al. (2018). It directly limited the gradient of parameters inside the discriminator network through spectral normalization and achieve better performance. In this paper, spectral normalization is used to stabilize the adversarial training in our model.

3. Progress edge generation network

As mentioned above, the missing data reconstruction process is divided into two sub-processes: structure prediction and texture generation. In this section, we will describe the structural prediction model in detail.

3.1. Motivation

Structure prediction is to predict the edges in missing regions. We use Canny algorithm (Canny, 1986) to extract edges, because it is a general algorithm that can be easily used in any images. Nazeri et al. (2019) and Xiong et al. (2019) provided two reference models for edge prediction. The network architectures of these two methods are both encoder-decoder architectures, but their loss functions are different. After experiments and summary, L1 loss, binary cross-entropy (BCE) loss, feature-matching loss (Nazeri et al., 2019) and adversarial loss can be used to predict edges in missing regions. We verify the characteristics of these loss functions on the ISPRS Vaihingen dataset through an encoder-decoder network. The network architecture will be described in Section 3.2. The experimental results are shown in Fig. 3. These loss functions can be divided into two categories. The first type are pixel-level methods that directly perform calculation on the image pixel, including L1 and BCE loss functions, which are good at generating regular shapes such as lines and rectangles. The second type are feature-level methods that perform calculation after converting images to a feature space through CNN, including feature-matching and adversarial loss functions, which can produce some irregular edges. Fig. 3 shows the edge prediction results of these two type loss functions in small-scale and large-scale missing regions respectively.

From Fig. 3, we find that: 1) All four methods can achieve edge prediction in small-scale missing regions, but none of them perform well in large-scale missing regions. 2) The results of L1 and BCE (pixel-level) losses are clean and tidy, but feature-matching and adversarial (feature-level) losses will produce lots of trivial edges. 3) L1 and BCE losses are more suitable for generating straight edges, but not for curved edges; both straight and curved edges can be generated by feature-matching and adversarial losses, but they often misjudge the structure of objects. Vividly, pixel-level methods are “cautious” and feature-level methods are “reckless”. Based on the above viewpoints, the PEGN is designed for VHR images with large-scale missing regions. Since large-scale edge prediction task cannot be solved using only one neural network, we propose a step-by-step approach to alleviate this problem. At the same time, for making better use of the characteristics of pixel-

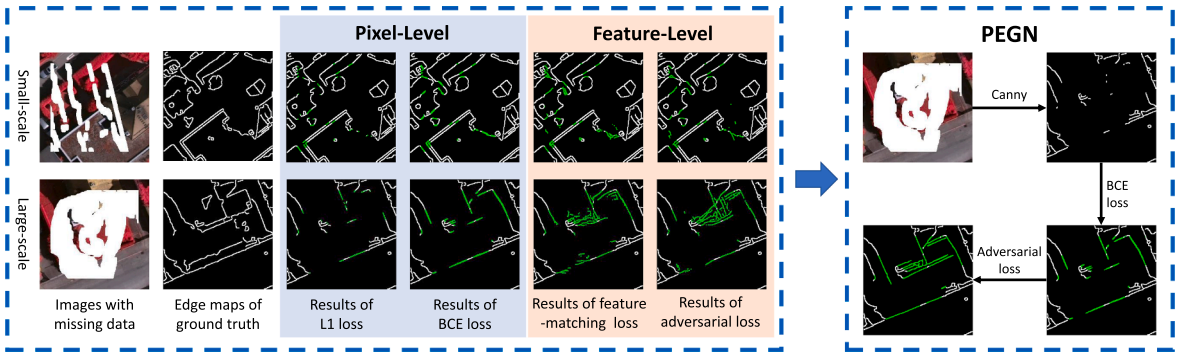


Fig. 3. Influence of different loss functions on edge prediction and the motivation of PEGN.

level and feature-level methods, we first use BCE loss to generate straight edges with reasonable structure, then use adversarial loss to generate more complex structures based on the results of BCE loss. Fig. 3 shows the basic framework of PEGN. PEGN can predict the structure of objects more reasonably and realize effective edge prediction in large-scale missing regions.

3.2. Network architecture

According to the basic framework in Fig. 3, PEGN is designed by combining two CNNs, each of which has its own loss function to achieve progressive edge generation. As shown in Fig. 4, PEGN can be defined as

$$\begin{aligned} S^{mid} &= G_1(I^m, M, S^m) \\ S^{pred} &= G_2(I^m, M, S^{mid}) \end{aligned} \tag{2}$$

where G_1 and G_2 represent the first and second generators used to predict edges. M is a binary image used to simulate missing regions. I^m and S^m are the VHR image and the edge map with missing data. S^{mid} and S^{pred} are the results of these two generators respectively, where S^{pred} is the final result of PEGN.

G_1 and G_2 are both a CNN with encoder-decoder architecture similar to the model proposed Johnson et al. (2016). The encoder consists of several strided convolution layers used to downsample the input image, and followed by 8 residual blocks (He et al., 2016). The decoder uses several transposed convolution layers to upsample the activation maps and produces the results of edge prediction. Meanwhile, the Markovian discriminator (D) (Isola et al., 2017) is designed to train G_2 in an adversarial manner based GAN. It outputs an N × N matrix where each element in the matrix shows a value of either "real" or "fake" to represent the quality of each patch in the image. This is equivalent to dividing the image into multiple blocks, then evaluating the quality of each block. The Markovian discriminator realizes the extraction of local image features, which is beneficial when generating high resolution images.

To stabilize the adversarial training of G_2 and D, spectral normalization (Miyato et al., 2018) is introduced to limit the gradient of the discriminator. After a rigorous mathematical derivation, spectral normalization can be simply summarized as

$$\begin{aligned} \sigma(W) &= \max_{h:h\neq 0} \frac{||Wh||_2}{||h||_2} = \max_{||h||_2\leq 1} ||Wh||_2 \\ W_{SN}(W) &= \frac{W}{\sigma(W)} \end{aligned} \tag{3}$$

where h and W are the input activation maps and parameter matrix respectively of each convolution layer in the discriminator. $\sigma(W)$ is the spectral norm of the matrix W equivalenting to calculate the largest singular value of W, and $W_{SN}(W)$ is the result of spectral normalization. According to the description in the original paper (Miyato et al., 2018),

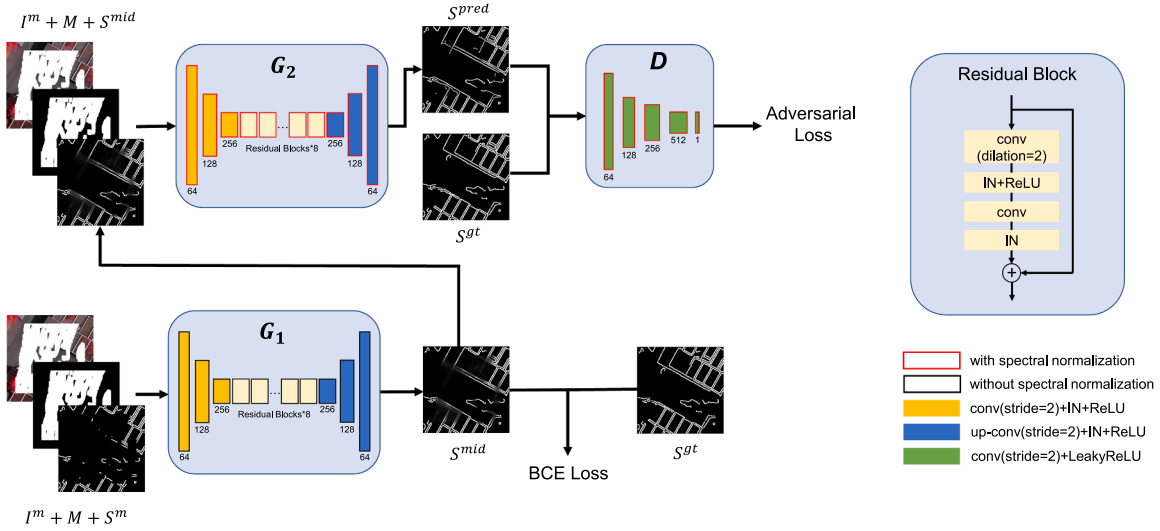


Fig. 4. Structure of PEGN. G_1 uses BCE loss to generate straight edges with reasonable structure, and G_2 uses adversarial loss to generate more complex structures to refine the results of G_1. G_1 and G_2 are all CNNs based on model proposed by Johnson et al. (2016). D is designed based on SN-GAN (Miyato et al., 2018) and Markovian discriminator (Isola et al., 2017). In detail, IN is instance normalization (Ulyanov et al., 2016), the kernel size of each convolution layers is 3 × 3, up-conv represents the transposed convolution, the slope of LeakyReLU is 0.2 and the dilation factor of the first convolution layer in each residual block is set to 2 to expand the receptive field. Additionally, the sigmoid function is added to the end of G_1 and G_2 for more convenient representation of edges.

if each W is normalized using formulation (3), the Lipschitz constant of discriminator will be bounded from above by 1, thereby enhancing the stability of adversarial training. Recently, Zhang et al. (2019) demonstrates that spectral normalization not only makes the discriminator more stable, but also improves the performance of the generator. Therefore, both G_2 and D in PEGN employ the spectral normalization to stabilize the adversarial training.

3.3. Loss functions

As mentioned in Section 3.2, PEGN has two generators with different loss functions. First, the BCE loss function is used to train G_1. To deal with the sparsity of the edges, the weight of missing regions is adjusted dynamically based on the idea of focal loss (Lin et al., 2017). The adjusted BCE loss is

$$\begin{aligned} L_1 = {} & \lambda M \odot \left(S^{mid} - S^{gt}\right)^{\gamma} L_{BCE}\left(S^{mid}, S^{gt}\right) \\ & + (1-M) \odot \left(S^{mid} - S^{gt}\right)^{\gamma} L_{BCE}\left(S^{mid}, S^{gt}\right) \end{aligned} \tag{4}$$

where M represents the missing regions, $\odot$ denotes the Hadamard product, S^{gt} is the edge map of ground truth; λ and γ are set to 5 and 2 respectively to increase the importance of missing regions.

Second, the adversarial loss function shown in formula (5) is employed to train G_2. Benefiting from spectral normalization, the adversarial training process can be more steadily.

$$\begin{aligned} L_{adv} = {} & \min_{G_2} \max_{D} E[\log D(I^{gt}, S^{gt}) + \\ & \log(1 - D(I^{gt}, S^{pred}))] \end{aligned} \tag{5}$$

These two loss functions have different effects. BCE loss predicts the probability that each pixel belongs to an edge and visualizes the probability by grayscale. It is used to guide network G_1 to generate simple and straight edges with reasonable structures, such as the edges of roads and buildings. However, BCE loss usually cannot generate complex edges. Hence, adversarial loss is added to accommodate the shortcomings of BCE loss. It will guide G_2 to generate complex and irregular edges that D cannot distinguish between true and false. Utilizing these two loss functions allows our model achieve progressive edge generation in large-scale missing regions.

4. Texture generation network

After PEGN predicts the structure of the missing regions, the next step is to use this structure information for image reconstruction. Similar to EdgeConnect (Nazeri et al., 2019), a texture generation network (TGN) is designed to generate the textures of missing regions based on the predicted structural information. The process of texture generation can be defined as

$$\begin{aligned} I^{pred} &= TGN(I^{m}, S^{pred}) \\ I^{re} &= (1-M) \odot I^{m} + M \odot I^{pred} \end{aligned} \tag{6}$$

where I^{m} refers to the image with missing data, I^{pred} is the output of TGN, and I^{re} represents the final reconstruction result of our model.

4.1. Network architecture

To generate realistic textures, we introduce GAN technology that divides the TGN into two parts: generator and discriminator, as shown in Fig. 5. The generator (G) accepts a combination of images with missing data and corresponding edge maps predicted by PEGN as input, and outputs the reconstruction results. And the discriminator (D) is trained to distinguish the reconstruction results from the real images. Through adversarial training, we can get a G with the ability to generate realistic texture. In TGN, G is the core wherein entire texture reconstruction process is executed, while D is only an auxiliary network.

The network architectures of G and D in TGN are same as G_2 and D in PEGN. We do not use specialized convolution operations such as partial convolution (Liu et al., 2018) or gated convolution (Yu et al., 2019). The reasons are (1) partial convolution is incompatible with the input edge maps, because it classifies all spatial locations in the missing regions to be invalid which will neglect the roles of edges as structural constraints. (2) the addition of extra modules in partial and gated convolution layers doubles the amount of the network parameters. (3) with the constraints of edges, good reconstruction results can be obtained simply and efficiently using standard convolution layers as described in Section 6.2. The illustration of the three convolution layers is shown in Fig. 6.

4.2. Loss functions

The loss function of TGN is a combination of L1, adversarial, and style loss functions. L1 loss calculates the L1 distance between ground-truth I^{gt} and I^{pred}, as

$$\begin{aligned} L_{L1} = {} & E[\lambda || M \odot (I^{pred} - I^{gt}) ||_1 + \\ & || (1-M) \odot (I^{pred} - I^{gt}) ||_1] \end{aligned} \tag{7}$$

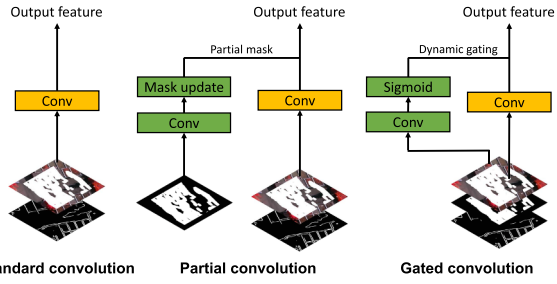


Fig. 6. Illustration of standard, partial and gated convolution layers. The green parts represent the extra modules compared to the standard convolution. Partial convolution masks and renormalizes the features conditionally on only valid regions. Gated convolution generalizes partial convolution by introducing a learnable dynamic feature selection mechanism at each spatial location, similar to attention mechanism (Vaswani et al., 2017).

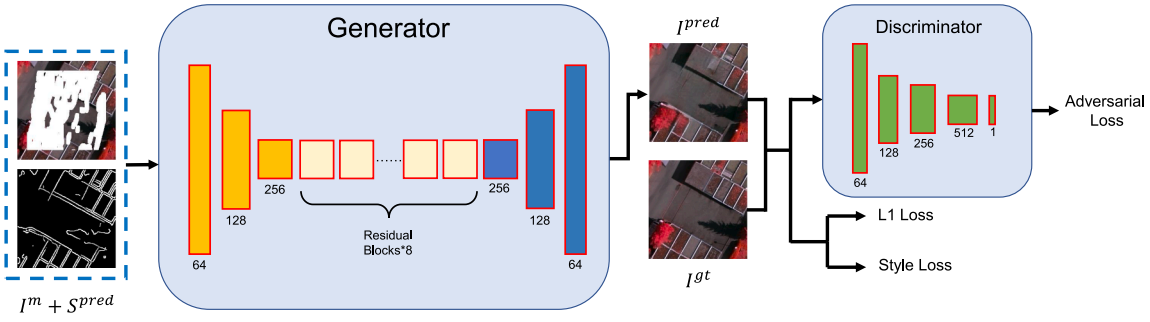


Fig. 5. Structure of TGN, where each shape has the same meaning as Fig. 4. Tanh function is added to the end of G for the prediction of pixel intensities.

in which λ is set to 5 to increase the importance of missing regions. The purpose of L1 loss is to make the reconstruction results as similar as possible to the ground truth. However, it will reduce the sharpness and flexibility of the reconstruction results. Adversarial loss is the core of GAN, which can guide the generator to generate realistic and various images. In addition, the adversarial training process is more stable with the introduction of spectral normalization. The formula for adversarial loss is as follows

$$L_{adv} = \min_G \max_D E[\log D(I^{gt}) + \log(1 - D(G(I^m, S^{pred})))] \tag{8}$$

Furthermore, style loss (Gatys et al., 2016; Johnson et al., 2016) commonly used in image style transfer fields is introduced to our model to pursue better texture details. Its formula is

$$L_{style} = E[||GM^{\phi^n}(I^{pred}) - GM^{\phi^n}(I^{gt})||_1] \tag{9}$$

where ϕ^n represents the activation maps from relu1_1, relu2_1, relu3_1, relu4_1, and relu5_1 layers of the VGG-19 (Simonyan and Zisserman, 2014) to extract features of images; *GM* is the Gram matrix. Measuring the differences between Gram matrices from two images is equivalent to measuring their differences in style information.

Finally, the total loss function is formed by combining these three loss functions, as

$$L_{total} = \lambda_{L1} L_{L1} + \lambda_{adv} L_{adv} + \lambda_{style} L_{style} \tag{10}$$

In practice, we set $\lambda_{L1} = 1$, $\lambda_{adv} = 0.4$ and $\lambda_{style} = 250$. Experiments on the effects of these loss functions is described in Section 6.2.

5. Experiments

5.1. Implementation details

Experiments are conducted on the ISPRS Vaihingen dataset that contains 33 orthorectified image patch mosaics with 3 spectral bands (near-infrared, red and green) to validate the performance of our model. This dataset involves 5 object classes and 1 background class, and the spatial resolution is 9 cm. We choose 16 images for training and 17 for testing. Because the average size of each image is 2494 × 2064, each image is randomly cropped to 256 × 256 or 512 × 512 during training. We also conduct experiments using the ISPRS Potsdam dataset and DOTA dataset (Xia et al., 2018) to test the generalization capability of our model. The ISPRS Potsdam dataset conntains 38 images tiles and each tile has 6000 × 6000 pixels and 4 channels (near-infrared, red, green and blue). The DOTA dataset has 2806 aerial images from different sensors and platforms and each size is about 4000 × 4000 with 3 cahnnels (red, green and blue).

In order to create images with missing data, we use binary mask images provided by Liu et al. (2018) as shown in Fig. 7 to simulate the missing regions. This dataset contains 6 categories of mask images with different ratios and each category has 2000 images.

Our model is trained using the Adam (Kingma and Ba, 2014) algorithm as the gradient descent optimization method with a learning rate of 0.0001. We also apply random scaling (0.5 to 2.0), horizontal flipping, and random crop over the training dataset for data augmentation. In the training process, our model is iterated for 80,000 iterations on an NVIDIA RTX 2070 GPU with batch size of 4.

5.2. Simulated experiments

We artificially create images with missing data in ISPRS Vaihingen dataset using binary mask images, and compare our model with Inpaint_Telea (Telea, 2004) (the inpainting function in OpenCV), PatchMatch (Barnes et al., 2009), Partial Conv (Liu et al., 2018), and EdgeConnect (Nazeri et al., 2019). Among them, Inpaint_Telea is a propagation-based method, PatchMatch is an exemplar-based method, and the others are deep learning-based methods. Fig. 8 shows the comparison results. For reconstruction tasks in small-scale missing regions, the capabilities of these five methods is similar; however, the results are very different in large-scale missing regions. Inpaint_Telea produces some fuzzy results, in which the texture and structure of ground objects are severely destroyed. Since the reconstruction results of PatchMatch are merged from the remaining regions, its texture details are realistic. But PatchMatch cannot capture the semantic information of ground objects, which leads to the appearance of structural disorder. Partial Conv destroys the boundary of the objects because it fails to correctly predict the structure. EdgeConnect and our model are both edge-based methods and produce acceptable texture effects, but the predicted structure of our model is more reasonable as shown in Fig. 8.

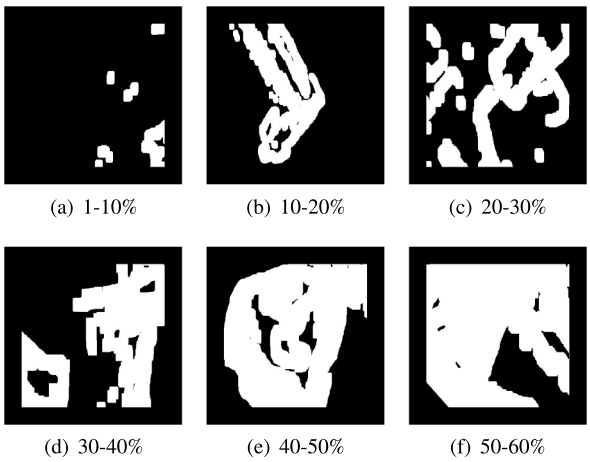


Fig. 7. Some binary mask images to simulate the missing regions. The numerical range below represents the proportion of missing regions.

To evaluate the above methods numerically, we use peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) (Wang et al., 2004) indices for quantitative evaluation. The quantitative evaluation results are divided into six groups according to the ratio of masked regions for comparing clearly. The results shown in Table 1 indicate that our method becomes more advantageous as the ratio of missing regions grows.

5.3. Classification experiments

More comprehensively, we evaluate our model qualitatively and quantitatively from the perspective of classification. First, a DeepLab V3+ (Chen et al., 2018) classification model is trained on the ISPRS Vaihingen training dataset with image size of 512 × 512 and batch size of 4. Then, we use this trained model to classify the reconstruction results on the test dataset and calculate the classification accuracy. The classification results are shown in Fig. 9. It should be noted that these classification results have some fixed errors due to the limitations of DeepLab V3+ itself. Fig. 9(d)-(f) indicate that classification results of our model are closer to the labels.

Furthermore, we use overall accuracy (OA), F1 scores, and mean intersection over union (mIoU), which are commonly used in remote sensing classification tasks as accuracy indicators, to evaluate the classification results of reconstruction results from different methods. As shown in Table 2, there is a significant gap between the Partial Conv and the other two methods. And the classification accuracy of our model is slightly better than EdgeConnect. This experiment indicates that our model still has an advantage from classification perspective.

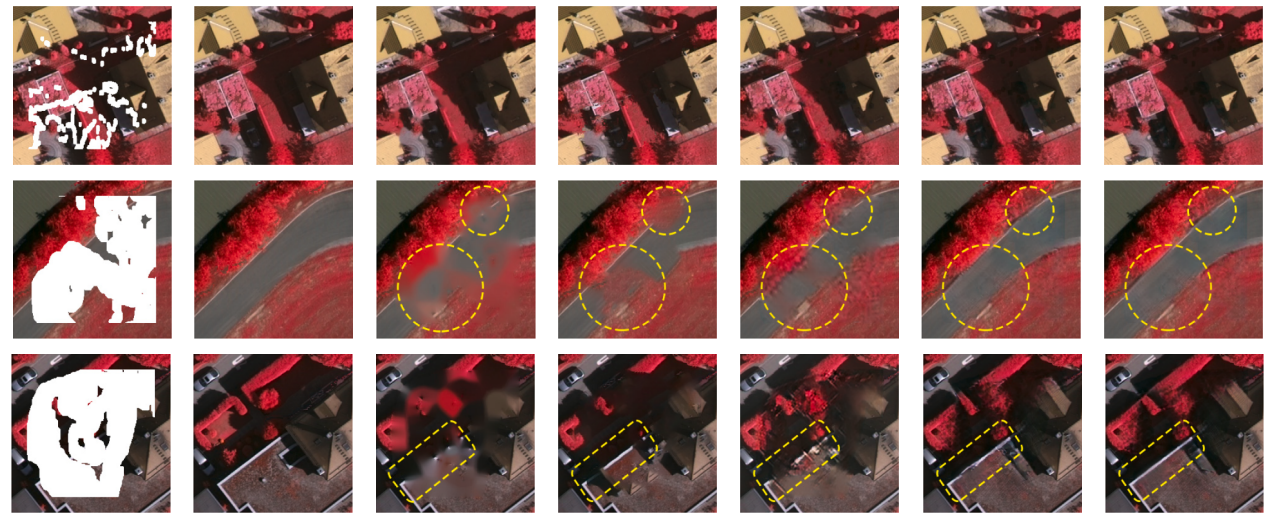


Fig. 8. Comparison of reconstruction results from the perspective of classification with DeepLab V3+. (a) Images with missing data. (b) Class label images. (c) Classification results of original images. (d)-(f) Classification results of reconstructed images produced by Partial Conv, EdgeConnect, and our model.

Table 1
Quantitative evaluation with different ratio of missing regions.

Mask Ratio	Methods	PSNR	SSIM
1%-10%	Inpaint_Telea	36.773	0.978
	PatchMatch	**37.774**	0.986
	Partial Conv	36.283	0.984
	EdgeConnect	37.228	0.986
	Ours	37.723	**0.987**
10%-20%	Inpaint_Telea	31.303	0.943
	PatchMatch	30.426	0.947
	Partial Conv	31.021	0.961
	EdgeConnect	31.851	0.966
	Ours	**32.112**	**0.967**
20%-30%	Inpaint_Telea	28.056	0.896
	PatchMatch	26.117	0.913
	Partial Conv	27.744	0.928
	EdgeConnect	**28.911**	0.938
	Ours	28.860	**0.940**
30%-40%	Inpaint_Telea	25.790	0.849
	PatchMatch	22.181	0.881
	Partial Conv	25.557	0.895
	EdgeConnect	26.410	0.908
	Ours	**26.770**	**0.912**
40%-50%	Inpaint_Telea	23.904	0.800
	PatchMatch	20.528	0.852
	Partial Conv	23.779	0.858
	EdgeConnect	24.603	0.876
	Ours	**24.708**	**0.879**
50%-60%	Inpaint_Telea	21.581	0.740
	PatchMatch	20.842	0.826
	Partial Conv	21.179	0.803
	EdgeConnect	21.986	0.829
	Ours	**22.038**	**0.833**

5.4. Application experiments

This section shows the redundant objects removal effect of our model. Suppose we want to obtain a clean scene from the ISPRS Vaihingen dataset, such that the cars in these images can be considered as redundant objects that we want to remove with our model. Since the original image is too large to be input into our model at once, we first crop the original image into many sub-blocks where the size of each sub-block is 512×512, then perform reconstruction operations on each sub-block, and finally match them together to complete the cars removal task. Fig. 10(a) is the original image with lots of cars. Fig. 10(b) is the binary mask image representing the position of cars extracted from the classification labels. However, due to the original mask usually cannot completely wrap the cars, we use dilate operation to expand the range of missing regions to obtain better reconstruction effect. Fig. 10(c) is the reconstruction result where cars have been removed. Because some shadows of cars are not wrapped by mask image, the reconstruction result contains a few fuzzy black shadows. In the second scenario, the planes in the image from DOTA dataset are regarded as redundant objects. A benefit of this dataset is that the positions of the bounding boxes of these planes are provided, so the mask image can be obtained directly through those bounding boxes as shown in Fig. 11(b). Similarly, we also divide the image into multiple sub-blocks and perform reconstruction separately. Fig. 11(c) shows the planes removal results.

6. Discussion

6.1. Discussion of PEGN

One of the highlights of our model is that the reconstruction process is divided into two parts: structural prediction and texture generation. Structural prediction refers to predicting the edges of ground objects in missing regions. Edges play an important role in our model, where different edges will make a significant impact on the final reconstruction results. To visually explain the effect of edges, Fig. 12 shows the influence of different edge maps on reconstruction results under fixed TGN.

In Fig. 12(b), the image is reconstructed with an incomplete edge map. We find that the reconstruction result is blurry at the regions where the edge information is missing. Fig. 12(c) shows the results of using a complete edge map to reconstruct the image: the blur phenomenon is gone and clear boundaries between objects remain. By comparing Fig. 12(b) and (c), the importance of edges for our model can be represented well, that is, having complete and reasonable edges will help TGN to produce better reconstruction results. This is a “double-edged sword”, in that unreasonable edges will mislead the reconstruction result.

The above experiment indicates that the results of PEGN can directly affect the quality of the final reconstruction results. So, a powerful edge prediction method is important for our model. Fig. 3 has shown some results of different edge prediction methods, but none of them perform

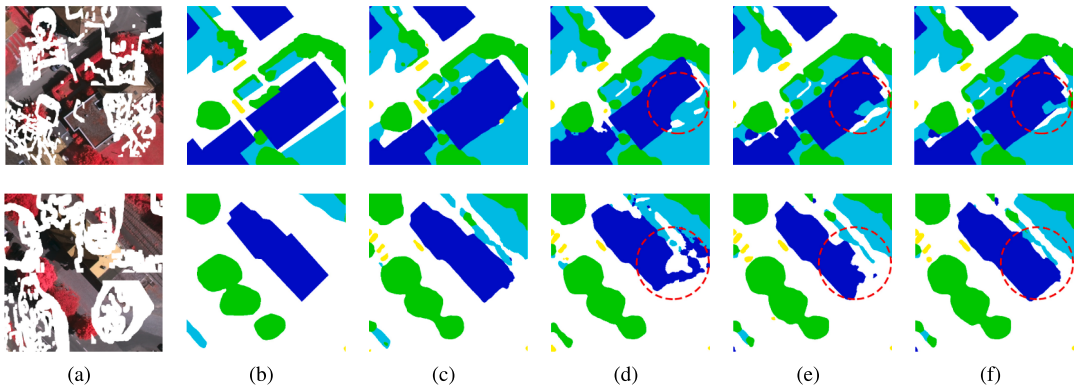


Fig. 9. Comparison results of simulating experiments.

Table 2
Classification accuracy of reconstruction results under different ratios of missing regions.

Mask Ratio	Methods	OA	F1	mIoU
1%-10%	Partial Conv	87.830	85.714	75.247
	EdgeConnect	87.869	85.919	75.582
	Ours	**87.982**	**85.984**	**75.678**
10%-20%	Partial Conv	87.077	84.572	73.438
	EdgeConnect	87.580	85.345	74.667
	Ours	**87.688**	**85.379**	**74.732**
20%-30%	Partial Conv	85.953	82.991	70.955
	EdgeConnect	**87.132**	84.350	73.188
	Ours	87.124	**84.455**	**73.307**
30%-40%	Partial Conv	83.891	80.335	66.994
	EdgeConnect	86.140	82.857	70.868
	Ours	**86.148**	**83.000**	**71.110**
40%-50%	Partial Conv	81.410	77.494	62.928
	EdgeConnect	84.746	80.888	67.926
	Ours	**85.056**	**81.170**	**68.321**
50%-60%	Partial Conv	74.802	70.953	54.275
	EdgeConnect	80.635	75.999	60.917
	Ours	**81.368**	**77.047**	**62.420**

well in VHR images with large-scale missing regions and complex structures. Fig. 13 shows the comparison result between the state-of-the-art inpainting model EdgeConnect (feature-matching loss) and our PEGN. It can be seen that EdgeConnect produces some trivial edges with disordered structure leading to the destruction of the structure of the ground objects that does not appear in PEGN.

To further verify the rationality of PEGN, we conduct an ablation study by removing and modifying some modules in PEGN and observing how that affects performance. First, we remove G_2 in PEGN and only use BCE loss and adversarial loss respectively to train G_1. Fig. 14(c) and (d) show that only use G_1 cannot complete a large-scale edge prediction task. Then, we change the loss of G_2 to BCE loss, that is, the loss of G_1 and G_2 is the same. Because BCE loss is a "cautious" method that is only good at generating straight and regular edges as mentioned in Section 3.1, some complex object structure cannot be predicted by BCE loss as shown in Fig. 14(e). Relatively, the edge prediction result of our PEGN shown in Fig. 14(f) has the most reasonable object structure.

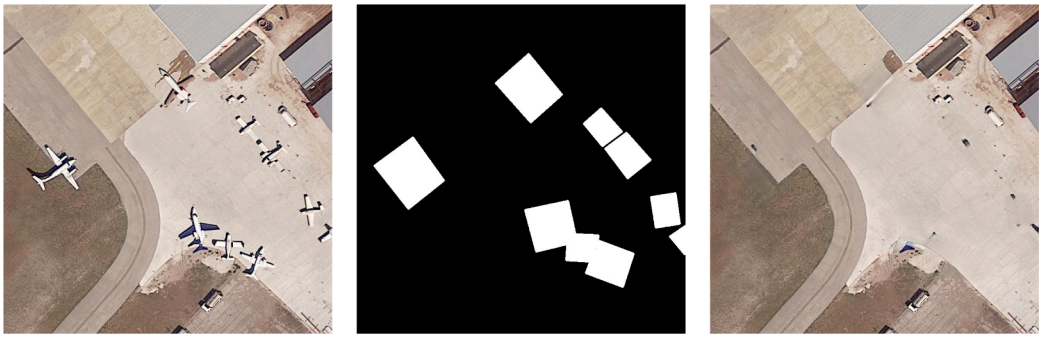
6.2. Discussion of TGN

As mentioned in Section 4.2, TGN has a total loss function comprised of three loss functions. We design an ablation experiment as shown in Fig. 15 to verify the effect of each loss function. When only L1 loss is used, the result is very blurry as shown in Fig. 15(c). This is because using L1 loss is equivalent to estimating the median pixel value of the missing regions (Bishop, 2006). Fig. 15(d) shows the result of joining L1 and adversarial losses, where the blur phenomenon has been improved but "checkerboard" artifacts occurred. When adding style loss, the result becomes more obviously realistic as shown in Fig. 15(e). Furthermore, style loss is also an effective tool for combatting "checkerboard" artifacts.

Partial convolution and gated convolution are two specialized convolution layers for image inpainting. However, their advantages cannot be reflected under the constraint of the edges as mentioned in Section 4.1. We first conduct an experiment to compare the performance of partial, gated and standard convolution layers in ISPRS Vaihingen

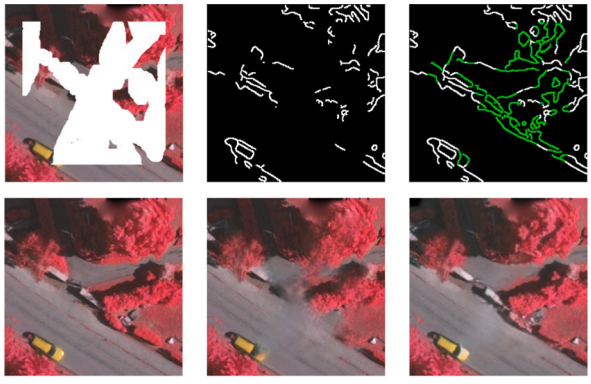
(a) Original Image (b) Mask image of cars (c) Cars removal result

Fig. 10. Cars removal experiment.



(a) Original Image (b) Mask image of planes from bounding boxes (c) Planes removal result

Fig. 11. Planes removal experiment.



(a) (b) (c)

Fig. 12. Influence of different edge maps on reconstruction results. (a) Image with missing data and its ground truth. (b) Incomplete edge map and its result. (c) Edge map extracted from ground truth and its result.

dataset. For fairness, the basic network architecture is same as Fig. 5, and the only difference is in the convolution layers. The results are shown in Fig. 16 and Fig. 17. Partial convolution cannot make good use of the structural information provided in the edges, resulting in a fuzzy boundary of objects. In contrast, gated and standard convolution can take advantage of the edges and produce similar results.

We further compare these three convolution layers in term of efficiency in an NVIDIA RTX 2070 GPU, including parameters, GPU memory when training and FPS when testing. As shown in Table 3, standard convolution has 2 × less parameters and 2 × improvement in FPS than the others, 1.55 × less GPU memory than gated convolution. To summarize, partial convolution is not suitable for reconstruction under edges constraints. Gated convolution has the best accuracy, but with poor efficiency. Comprehensively, standard convolution is a better choice.

6.3. Analysis of generalization ability

Improving the generalization ability of deep learning models has always been a challenge for researchers. An effective solution is using abundant and comprehensive data for model training, but most people have limited data and computing power. Therefore, how to use limited data to train the model to apply to more scenes is what we need to solve. To test the generalization ability, we trained our model on the ISPRS Vaihingen dataset with NIR-R-G band sequence and tried to apply this trained model to the ISPRS Potsdam dataset with NIR-R-G band sequence and DOTA dataset with R-G-B band sequence. It's worth noting that the ISPRS Potsdam dataset has a similar spectral distribution to ISPRS Vaihingen dataset, while the spectral distribution of DOTA dataset is completely different. Fig. 18(b) shows that PEGN has a strong generalization ability, it can produce reasonable edges even if the spectral distribution of the testing data was very different from that of the training data. TGN works well on Potsdam dataset, but its results on DOTA dataset suffer from significant chromatic aberrations as shown in Fig. 18(c) and (d). This indicates that TGN cannot transfer to the images that are too far apart from the training images. The good news is that TGN shows no serious structural collapse under the constraint of PEGN. Therefore, we can improve the generalization ability of our model by alleviating the chromatic aberrations.

Our model uses the fusion operation $(1-M)\odot I^{m}+M\odot I^{pred}$ to get the reconstruction image I^{re}, where I^{pred} is the output of TGN. The reason for the chromatic aberrations in I^{re} is that the spectral distributions of I^{m}

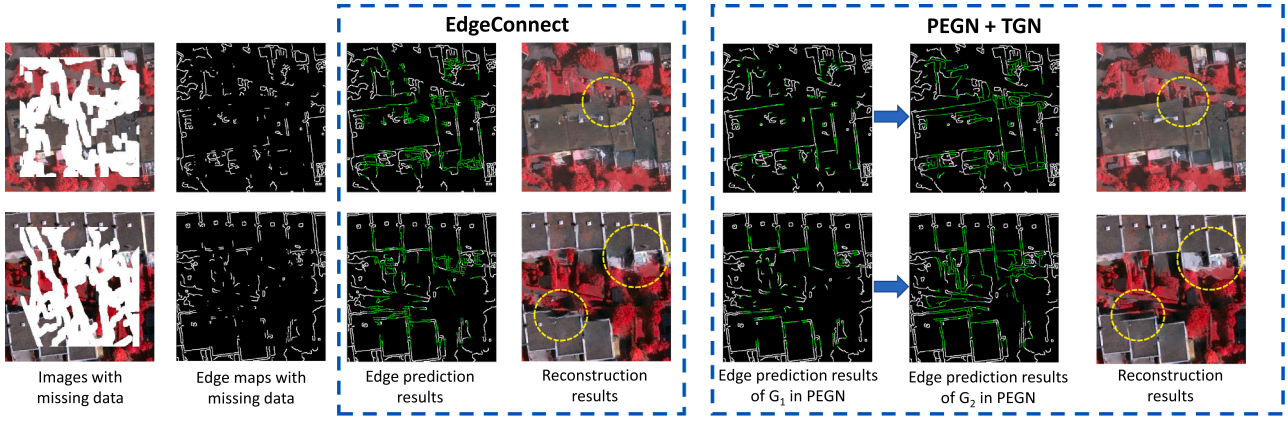
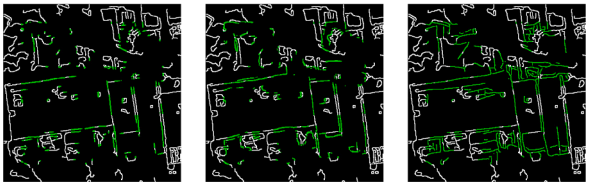


Fig. 13. Edge prediction results in large-scale missing regions of EdgeConnect and PEGN.

(a) Image with missing data (b) Edge map with missing data (c) Only Adv



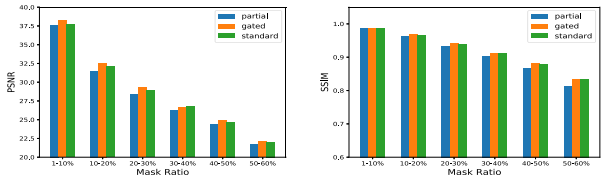
(d) Only BCE (e) BCE + BCE (f) BCE + Adv

Fig. 14. Ablation study of PEGN. A + B means G_1 with A loss and G_2 with B loss. BCE means binary cross-entropy loss and Adv means adversarial loss.

and I^{pred} are different, which causes the color of I^m and I^{pred} to be inconsistent. The breakthrough point is that I^{pred} itself has reasonable structure and no chromatic aberrations inside as shown in Fig. 18(c). Hence, the chromatic aberrations can be alleviated by adjusting the spectral distribution of I^{pred} as close to I^m as possible. Based on this idea, we propose a histogram matching-based method to eliminate the chromatic aberrations. This method uses the histogram matching to make the histogram of each channel of I^{pred} as consistent as possible with the histogram of I^m, and apply this pixel mapping rule to the missing data regions to eliminate chromatic aberrations phenomena as shown in Fig. 19. Another way is to use Poisson blending technology (Pérez et al., 2003), as described in Iizuka et al. (2017), which first employs the fast-marching method (Telea, 2004) followed by Poisson image blending. Fig. 18 shows the effects of these two chromatic aberration elimination methods. Restricted by the fast-marching method, Poisson blending produces some blurred boundaries that are not presented in histogram matching. Table 4 shows the quantitative evaluation results in the DOTA dataset where histogram matching is slightly better than Poisson blending.

7. Conclusions

In this paper, a novel deep learning-based missing data reconstruction method is presented for VHR satellite and aerial images. It is a spatial-based method that can generate realistic results without any auxiliary spectral or temporal data. The reconstruction process in our



(a) PSNR (b) SSIM

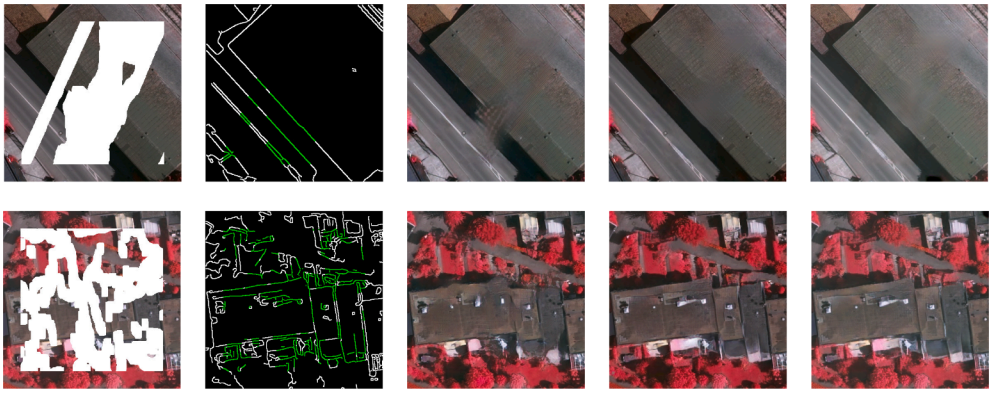
Fig. 17. Quantitative comparison results of partial, gated and standard convolution layers.

Table 3
Efficiency comparison of partial, gated and standard convolution layers when processing a image of size 512 × 512. The network architecture for experiment is described in Fig. 5

Convolution	Params (MB)	Memory (MB)	FPS
Partial	77.8	3852.0	7.1
Gated	80.2	5976.0	6.3
Standard	**40.0**	**3849.5**	**15.2**

(a) (b) (c) (d) (e)

Fig. 15. Ablation study of TGN. (a) Image with missing data. (b) Edge map. (c) Only L1 loss. (d) Joining L1 loss and adversarial loss. (e) Joining L1 loss, adversarial loss and style loss.



(a) (b) (c) (d) (e)

Fig. 16. Qualitative comparison results of partial, gated and standard convolution layers. (a) Images with missing data. (b) Edge maps. (c)-(d) Reconstruction results of partial, gated and standard convolution layers respectively under the constraint of edge maps (b).

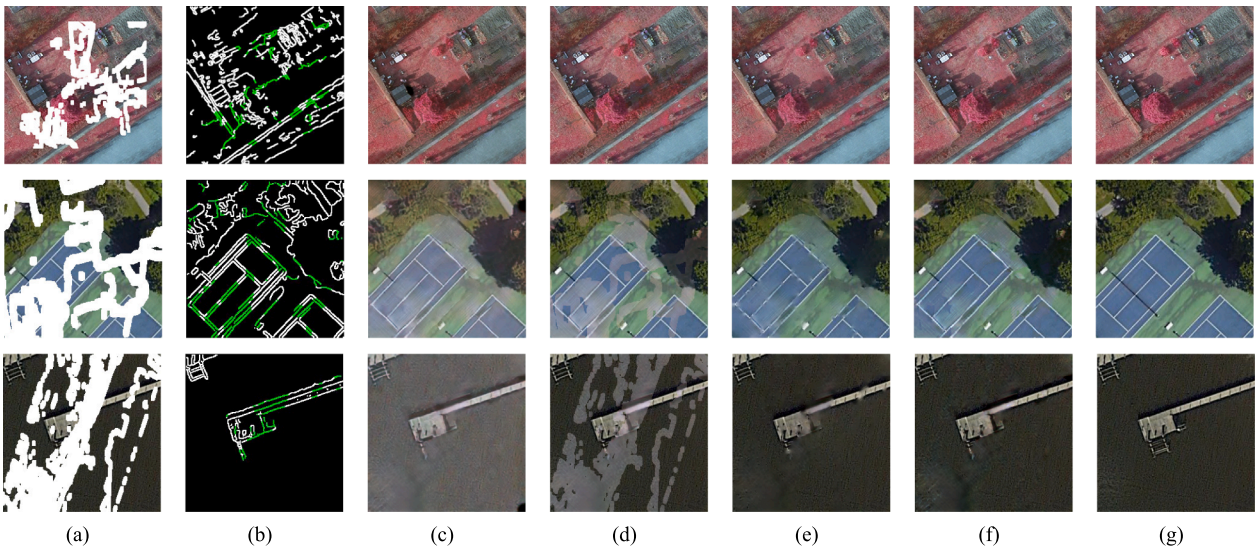


Fig. 18. Generalization ability experiment. The model is trained on the ISPRS Vaihingen dataset and tested on the ISPRS Potsdam dataset (the first row) and DOTA dataset (the second and third row). (a) I^m, images with missing data. (b) Edge maps produced by PEGN. (c) I^{pred}, the outputs of TGN. (d) I^{re} ($I^{re}=(1-M)\odot I^m + M\odot I^{pred}$), the final results of our model. (e) Results of Poisson blending. (f) Results of histogram matching. (g) Ground Truth.

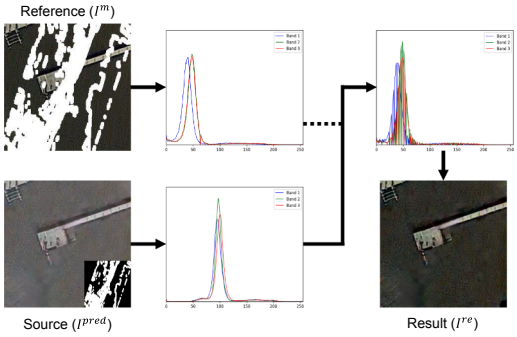


Fig. 19. Process of histogram matching to eliminate chromatic aberrations.

Table 4
Quantitative evaluation of chromatic aberrations elimination algorithms, where model is trained on the ISPRS Vaihingen dataset and tested on the DOTA dataset.

Mask Ratio	Methods	PSNR	SSIM
1%-10%	Ours	33.393	0.973
	+Poisson blending	36.987	0.982
	+Histogram Matching	**37.699**	**0.985**
10%-20%	Ours	28.704	0.939
	+Poisson blending	32.318	0.958
	+Histogram Matching	**33.185**	**0.964**
20%-30%	Ours	26.282	0.906
	+Poisson blending	29.842	0.930
	+Histogram Matching	**30.487**	**0.937**
30%-40%	Ours	24.502	0.874
	+Poisson blending	28.080	0.904
	+Histogram Matching	**28.472**	**0.909**
40%-50%	Ours	23.231	0.844
	+Poisson blending	26.744	**0.878**
	+Histogram Matching	**26.926**	**0.878**
50%-60%	Ours	21.699	0.812
	+Poisson blending	24.766	**0.842**
	+Histogram Matching	**24.780**	0.837

model is divided into two parts: structure prediction and texture generation, which reduces the burden on each individual network. In addition, we propose a two-stage structure prediction model, PEGN, which is more suitable for edge prediction in VHR images with large-scale missing regions. Experiments show that our model can better predict the structure of the missing regions and achieve good scores in SSIM and PSNR indices when compared with other spatial-based methods. And through Poisson blending and histogram matching, our model can acquire a strong generalization ability.

The proposed model performs well in experiments, but also has some shortcomings in practical applications: 1) requiring sufficient training data; 2) relying too much on the structural prediction model; 3) influenced by edge detection methods. Moreover, compared with spectral-based methods (Wang et al., 2006; Shen et al., 2010; Shen et al., 2013), temporal-based methods (Li et al., 2014; Zeng et al., 2013; Li et al., 2019), and hybrid methods (Zhang et al., 2018; Cheng et al., 2014), the authenticity of our model cannot be guaranteed, which is a common problem for spatial-based methods. Some possible solutions to alleviate these problems might include: using a more powerful edge extraction algorithm (Shen et al., 2015b), combining semantic segmentation techniques to conduct the reconstruction process (Park et al., 2019), utilizing few-show learning to reduce training data requirements, exploring other useful auxiliary information, etc.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Key Program of the National Natural Science Foundation of China [Grant No. 41930535]; Open Fund of Key Laboratory of Urban Natural Resources Monitoring and Simulation, Ministry of Natural Resources [Grant Nos. KF-2019-04-070], and the SDUST Research Fund [Grant No. 2019TDJH103].

References

Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein gan. arXiv preprint arXiv: 1701.07875.

Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B., 2009. Patchmatch: A randomized correspondence algorithm for structural image editing. ACM Trans. Graph. 28, 24.
Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer.
Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Machine Intell. 679–698.
Chen, F., Zhao, Z., Peng, L., Yan, D., 2005. Clouds and cloud shadows removal from high-resolution remote sensing images. In: Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05, pp. 4256–4259. Ieee volume 6.
Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818.
Cheng, Q., Shen, H., Zhang, L., Li, P., 2013. Inpainting for remotely sensed images with a multichannel nonlocal total variation model. IEEE Trans. Geosci. Remote Sens. 52, 175–187.
Cheng, Q., Shen, H., Zhang, L., Yuan, Q., Zeng, C., 2014. Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal mrf model. ISPRS J. Photogramm. Remote Sens. 92, 54–68.
Criminisi, A., Pérez, P., Toyama, K., 2004. Region filling and object removal by exemplar-based image inpainting. IEEE Trans. Image Process. 13, 1200–1212.
Gatys, L.A., Ecker, A.S., Bethge, M., 2016. Image style transfer using convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680.
Guillemot, C., Le Meur, O., 2013. Image inpainting: Overview and recent advances. IEEE Signal Process. Mag. 31, 127–144.
Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C., 2017. Improved training of wasserstein gans. In: Advances in Neural Information Processing Systems, pp. 5767–5777.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
Iizuka, S., Simo-Serra, E., Ishikawa, H., 2017. Globally and locally consistent image completion. ACM Trans. Graphics (ToG) 36, 1–14.
Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134.
Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision. Springer, pp. 694–711.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Li, X., Shen, H., Li, H., Zhang, L., 2016. Patch matching-based multitemporal group sparse representation for the missing information reconstruction of remote-sensing images. IEEE J. Sel. Top. Appl. Earth Obser. Remote Sens. 9, 3629–3641.
Li, X., Shen, H., Zhang, L., Zhang, H., Yuan, Q., Yang, G., 2014. Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning. IEEE Trans. Geosci. Remote Sens. 52, 7086–7098.
Li, X., Wang, L., Cheng, Q., Wu, P., Gan, W., Fang, L., 2019. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. ISPRS J. Photogramm. Remote Sens. 148, 103–113.
Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988.
Liu, G., Reda, F.A., Shih, K.J., Wang, T.-C., Tao, A., Catanzaro, B., 2018. Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 85–100.
Lorenzi, L., Melgani, F., Mercier, G., 2011. Inpainting strategies for reconstruction of missing data in vhr images. IEEE Geosci. Remote Sens. Lett. 8, 914–918.
Maalouf, A., Carré, P., Augereau, B., Fernandez-Maloigne, C., 2009. A bandelet-based inpainting technique for clouds removal from remotely sensed images. IEEE Trans. Geosci. Remote Sens. 47, 2363–2371.
Mendez-Rial, R., Calvino-Cancela, M., Martin-Herrero, J., 2011. Anisotropic inpainting of the hypercube. IEEE Geosci. Remote Sens. Lett. 9, 214–218.
Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y., 2018. Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957.
Mohamed, S., Lakshminarayanan, B., 2016. Learning in implicit generative models. arXiv preprint arXiv:1610.03483.
Nazeri, K., Ng, E., Joseph, T., Qureshi, F.Z., Ebrahimi, M., 2019. Edgeconnect: Generative image inpainting with adversarial edge learning. arXiv preprint arXiv:1901.00212.
Park, T., Liu, M.-Y., Wang, T.-C., Zhu, J.-Y., 2019. Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2337–2346.
Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A., 2016. Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2536–2544.
Pérez, P., Gangnet, M., Blake, A., 2003. Poisson image editing. In: ACM SIGGRAPH 2003 Papers, pp. 313–318.
Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., Zhang, L., 2015a. Missing information reconstruction of remote sensing data: A technical review. IEEE Geosci. Remote Sens. Mag. 3, 61–85.
Shen, H., Li, X., Zhang, L., Tao, D., Zeng, C., 2013. Compressed sensing-based inpainting of aqua moderate resolution imaging spectroradiometer band 6 using adaptive spectrum-weighted sparse bayesian dictionary learning. IEEE Trans. Geosci. Remote Sens. 52, 894–906.
Shen, H., Zeng, C., Zhang, L., 2010. Recovering reflectance of aqua modis band 6 based on within-class local fitting. IEEE J. Sel. Top. Appl. Earth Obser. Remote Sens. 4, 185–192.
Shen, H., Zhang, L., 2008. A map-based algorithm for destriping and inpainting of remotely sensed images. IEEE Trans. Geosci. Remote Sens. 47, 1492–1502.
Shen, W., Wang, X., Wang, Y., Bai, X., Zhang, Z., 2015b. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3982–3991.
Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
Telea, A., 2004. An image inpainting technique based on the fast marching method. J. Graphics Tools 9, 23–34.
Ulyanov, D., Vedaldi, A., Lempitsky, V., 2016. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: Advances in Neural Information Processing Systems, pp. 5998–6008.
Wang, L., Qu, J.J., Xiong, X., Hao, X., Xie, Y., Che, N., 2006. A new method for retrieving band 6 of aqua modis. IEEE Geosci. Remote Sens. Lett. 3, 267–270.
Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13, 600–612.
Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. Dota: A large-scale dataset for object detection in aerial images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3974–3983.
Xiong, W., Yu, J., Lin, Z., Yang, J., Lu, X., Barnes, C., Luo, J., 2019. Foreground-aware image inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5840–5848.
Xu, R., Li, X., Zhou, B., Loy, C.C., 2019. Deep flow-guided video inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3723–3732.
Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S., 2018. Generative image inpainting with contextual attention. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5505–5514.
Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S., 2019. Free-form image inpainting with gated convolution. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4471–4480.
Zeng, C., Shen, H., Zhang, L., 2013. Recovering missing pixels for landsat etm+ slc-off imagery using multi-temporal regression analysis and a regularization method. Remote Sens. Environ. 131, 182–194.
Zhang, C., Li, W., Travis, D., 2007. Gaps-fill of slc-off landsat etm+ satellite image using a geostatistical approach. Int. J. Remote Sens. 28, 5103–5122.
Zhang, H., Goodfellow, I., Metaxas, D., Odena, A., 2019. Self-attention generative adversarial networks. In: International Conference on Machine Learning, pp. 7354–7363.
Zhang, Q., Yuan, Q., Zeng, C., Li, X., Wei, Y., 2018. Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network. IEEE Trans. Geosci. Remote Sens. 56, 4274–4288.