ISPRS Journal of Photogrammetry and Remote Sensing 199 (2023) 226–241

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Cross-sensor remote sensing imagery super-resolution via an edge-guided attention-based network

Zhonghang Qiu [a], Huanfeng Shen [a,c], Linwei Yue [b,*], Guizhou Zheng [b]

[a] School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
[b] School of Geography and Information Engineering, China University of Geosciences, Wuhan, China
[c] Hubei Luojia Laboratory, Wuhan, China

ARTICLE INFO

Keywords:
Super-resolution
Remote sensing imagery
Degradation modeling
Edge prior

ABSTRACT

The deep learning based super-resolution (SR) methods have recently achieved remarkable progress in the reconstruction of ideally simulated high-quality remote sensing image datasets. However, due to the large variation in image quality caused by the complex degradation factors, their performance decreases dramatically on real-world images acquired by different satellite sensors. To this end, we propose a cross-sensor SR framework that consists of a cross-sensor degradation modeling strategy for bridging the gap between the images obtained by the source and target sensors, and an edge-guided attention-based SR (EGASR) network to promote the learning of high-frequency feature representation. Specifically, we build a degradation pool on the low-resolution (LR) target sensor to produce a degraded training dataset simulated from the high-resolution (HR) images obtained by the source sensor. Furthermore, the EGASR network, which employs the edge-guided residual attention block (EGRAB) to introduce implicit edge prior to enhance edge-related information, is embedded in the cross-sensor SR framework for reconstructing HR results with sharp details. The proposed method is applied on images from the Chinese Gaofen (GF) satellite sensors and compared to several representative SR methods. An ideally simulated GF-2 LR/HR image set with only downsampling considered is first used to evaluate the effectiveness of the proposed EGASR network. Moreover, GF-2/GF-1 and GF-2/GF-6 cross-sensor SR datasets are constructed by synthesizing GF-2 degraded image pairs with the degradation pools estimated from the GF-1 and GF-6 images, respectively. The results show that: 1) the proposed EGASR model shows superiority in reconstructing textural details and edge features, and achieves the best results among the state-of-art SR methods involved in comparison; 2) the cross-sensor SR framework significantly promotes the model's ability to super-resolve real-world LR images acquired by the target satellite sensors, e.g., the NIQE values are improved by at least 30% and 34% on average with respect to other comparative methods for GF-2/GF-1 and GF-2/GF-6 datasets in the real experiments, respectively. Code is available at https://github.com/zhonghangqiu/EGASR.

1. Introduction

High-resolution (HR) remote sensing images contain finer textural details than low-resolution (LR) images. With the rapid development of remote sensing imaging techniques, the resolution of images has been boosted in the last decades. However, the extensive demand for fine-scale image parsing applications (e.g., change detection (Liu et al., 2022b), fine-grained classification (Zhu et al., 2021), and semantic segmentation (Zheng et al., 2020)) still has a high requirement for the spatial resolution of the images. Moreover, the spatial details of remote sensing images are often degraded by multiple factors (e.g., optical diffraction, blur, and noise) through the acquisition process. Considering the high cost and limitations in improving the imaging equipment, algorithmic-based image super-resolution (SR) technology has been a popular research topic.

Image SR is an ill-posed inverse problem that involves recovering an HR image from one or multiple LR images, which has now been developed for nearly-four decades. Overall, the SR algorithms can be roughly summarized into three main categories, i.e., interpolation-based (Wang et al., 2015; Zhang and Wu, 2006), reconstruction-based (Zhang et al., 2012), and learning-based methods (Dong et al., 2011; Gao et al., 2012; Yang et al., 2010). Before the explosion of learning-based SR methods,

* Corresponding author.
E-mail address: yuelw@cug.edu.cn (L. Yue).

https://doi.org/10.1016/j.isprsjprs.2023.04.016
Received 13 July 2022; Received in revised form 3 April 2023; Accepted 14 April 2023
Available online 24 April 2023

the effectiveness of single-image SR (SISR) methods was limited without the involvement of external information (Molini et al., 2019). In recent years, the deep learning based methods (Anwar and Barnes, 2022; Dong et al., 2015; Haris et al., 2021) have obtained remarkable performances on various public benchmark datasets, and have shown their superiority over the traditional SISR methods. State-of-the-art results have been obtained with the frequent improvements in network design, such as residual learning (He et al., 2016), recursive learning (Tai et al., 2017), attention mechanisms (Hu et al., 2020), dense connections (Zhang et al., 2021), and the recent popular transformer model (Liang et al., 2021; Lu et al., 2022). Compared with multi-frame SR methods that require sequential input images, SISR is more practical and convenient, especially for remote sensing data. Therefore, we mainly refer to SISR in this work.

Overall, the progress of deep learning techniques has resulted in significant improvements in image SR (Lepcha et al., 2023). However, compared with natural images, satellite images have some distinctive characteristics.

(1) The images acquired with spaceborne imaging systems are generally contaminated with complex degradation factors, such as atmospheric scattering and absorption, sensor distortion, and system noise. The degradation also differs with the different satellite sensors, as well as the acquisition conditions.
(2) The ground objects with various scales and texture details within large-scale image scenes can make it difficult to super-resolve the high-frequency information from the LR imagery. Moreover, the blurring and noise in satellite images pose an extra challenge for the models distinguishing the crucial features with anomalies.

A variety of deep learning based methods have been proposed to address the SISR issues for remote sensing images, and have achieved some success (Sdraka et al., 2022; Wang et al., 2022a). Nevertheless, these methods rarely pay attention to the variation in image quality associated with complex degradation factors (Kang et al., 2022; Lanaras et al., 2018; Yin et al., 2022). For example, most methods train the SR model on datasets composed of HR images and simulated LR image pairs collected by the source sensor, and then apply the trained model to real LR images acquired from the target sensor (Pouliot et al., 2018; Xiong et al., 2020). However, the LR images in the training set are often generated by applying a simple downsampling operator to the corresponding HR images. Without full consideration of the varied degradation conditions for different satellite sensors, the effectiveness of the trained model can be reduced in practical applications, due to the poor generalization capability (Chen et al., 2022; Liu et al., 2022a). In recent years, some researchers have begun to explore methods considering complicated degradations to solve the real-world remote sensing image SR task. Zhang et al. (2022a) proposed an unsupervised SR framework aided by multi-degradation, which can handle complex degradation schemes existing in remote sensing images. However, the unsupervised framework is usually difficult to train and may produce artifacts in some cases which are harmful to real-world applications. Dong et al. (2022) proposed a degradation model to simulate blur and noise degradation for remote sensing images, which considered the domain gap between the simulated training data and real-world images. However, they only considered simple Gaussian noise and JPEG compression with regard to the noise factor in the degradation assumption, which still needs extensions to be more comprehensive for real-world settings. In addition to the degradation modeling-based methods, another popular pipeline is to collect real-world LR/HR image pairs of the same scene for network training. For example, to train the SR model, Galar et al. (2020) constructed a dataset by collecting Sentinel-2 images and reference Planet images with similar spectral characteristics at the same locations. More recently, Wang et al. (2021a) introduced a real-world multi-sensor dataset consisting of Landsat 8 Operational Land Imager (OLI) and Sentinel-2 Multispectral Instrument (MSI) image pairs to train the SR model, which can more effectively super-resolve Landsat 8 data than a model trained on simulated data. Nevertheless, collecting LR/HR image pairs from multiple sensors is laborious, due to the difference between cross-sensor images, i.e., cloud contamination, environmental changes, and atmospheric conditions. Therefore, it is of great significance to develop a generic SR approach for the task of cross-sensor image SR.

Another key issue is that the complicated edges in remote sensing images increase the difficulty of recovering accurate and detailed information by the means of SR. The convolutional neural network (CNN)-based SR methods have recently been the subject of much attention, with special attention being paid to the network architectures (Dong et al., 2021; Lei et al., 2022; Lei et al., 2017; Pan et al., 2019). These CNN-based models optimized by pixel-wise loss keep obtaining state-of-the-art results, but often leave blurred impressions (Wang et al., 2022b). The generative adversarial networks (GAN) are popularly employed in many SR works (Jia et al., 2022; Li et al., 2022; Tu et al., 2022) to promoted the perceptual quality of the recovered remote sensing images. However, the GAN-based methods usually tend to produce unnatural textures in the cases with complex data distributions, which are difficult to accurately model. Image priors, and especially edge priors, have been shown to be effective in recovering fine details (Fang et al., 2020; Zhang et al., 2022b). For example, Jiang et al. (2019) proposed a robust GAN-based method that integrates an edge extraction operation into the network to mitigate the structural distortion caused by adversarial learning in the SR process. Li et al. (2021) developed a progressive split-merge SR framework with gradient guidance and achieved a significant improvement in both the numerical index results and visual quality. Most of these prior-guided methods explicitly extract edge or gradient information from the input LR images and incorporate it into the network (Shen et al., 2022). However, the extracted edge information can be sensitive to the quality of the images, resulting in possible artifacts and ambiguous textures in the reconstruction images (Huan et al., 2022).

In general, the SR reconstruction of remote sensing images still exhibits a crucial challenge from two aspects, i.e., investigating robust real-world SR methods to handle complicated degradation for practical applications and reconstructing remote sensing images with sharp details in the context of imaging degradation. To address the above issues, we proposed a cross-sensor SR framework by simultaneously considering the two important perspectives, which consists of a data degradation modeling strategy to bridge the gaps between images obtained by the source and target sensors, and an edge-guided attention-based SR network, namely EGASR, for reconstructing sharp spatial details. Firstly, a degradation pool was built to construct to perform image degradation on HR images of the source sensor to generate LR images, which contain similar degradation characteristics to that of the target sensor images. In this way, the SR model trained on the degraded LR-HR paired dataset can be applied to super-resolve real-world images acquired by the target sensor. Secondly, EGASR was specially designed to capture spatially precise structural representations with the guidance of the edge prior to recover sharp boundary details in the images. In contrast to the existing methods that explicitly extract edge maps from the input LR images, edge information is implicitly extracted from the feature maps through the edge-guided residual attention block (EGRAB), thus suppressing the noise and artifacts in the images. The major contributions of this study are:

(1) We propose a cross-sensor SR framework to deal with the real-world images acquired by the target satellite sensor, considering the variations in image quality caused by the complex degradation factors.
(2) A degradation pool is built with blind noise and blur-kernel estimation in the image set of the target sensor and incorporated into the training phase. As a result, the model trained with synthesized LR/HR image pairs from the source sensor can be

adapted to the target sensor, which significantly improves the model's performance in the task of cross-sensor SR.

(3) We propose the EGASR network for remote sensing SISR, where the core component, the EGRAB, implicitly extracts edge features and guides the network to focus on structural feature representation, resulting in robust recovery of sharp and clear contours of the ground objects.

The rest of this paper is organized as follows. In Section 2, we introduce the details of the proposed method, including the cross-sensor SR framework and the structure of the EGASR network. The experimental results are given in Section 3, and a further detailed analysis and discussion of the effectiveness of the proposed method is provided in Section 4. Section 5 presents our conclusions.

2. Methodology

In this section, the idea and procedures of the cross-sensor SR framework are first introduced. The architectures of the proposed EGASR method and the specially designed internal modules are then described in detail.

2.1. Cross-sensor image super-resolution framework

Deep learning based SR methods performed in a supervised manner are restricted to training data. Therefore, the models trained on simulated datasets tend to perform poorly on real-world data. To this end, we developed a cross-sensor SR strategy which can apply the SR model trained on the HR image set from the source sensor to real LR data of the target sensor. The overall framework is shown in Fig. 1, which includes two main stages. Firstly, an image degradation model is built and a degradation pool is constructed by estimating the blur kernels and extracting noise patches from the real LR data of satellite sensor B. Based on the degradation pool, it is possible to generate synthetic degraded LR data from the HR images of sensor A, which contain degradation characteristics similar to that of the realistic data of target sensor B. Note that this strategy can avoid the requirement for an extensive collection of paired data with the same location and similar observation conditions from the two sensors, and the subsequent geometric registration. The SR model used in the proposed framework is the EGASR network, which is trained on the synthetically degraded SR data and finally applied to super-resolve the real LR data of sensor B.

2.1.1. Degradation pool building

Limited by the acquisition systems and environmental conditions in the imaging process, remote sensing image quality is often affected by varying degrees of degradation, including blurring, downsampling, and noise (Yue et al., 2016). Except for the resolution sampling, blurring and noise are mainly considered in this paper. Generally speaking, the degradation process to generate LR images is modeled as:

$$I_{LR} = (I_{HR} \otimes k)\downarrow_s + n \tag{1}$$

where I_{HR} is the HR image, I_{LR} denotes the obtained degraded LR image, $\downarrow_s$ represents the downsampling operation with a scale factor of s, and $\otimes$ denotes the linear (2D) convolution operation. In addition, k and n denote the blur kernel and additive noise, respectively.

To solve the cross-sensor remote sensing image SR task, a degradation pool is built by the means of blind noise and blur-kernel estimation. Firstly, the KernelGAN process proposed by Bell-Kligler et al. (2019) is introduced to estimate the blur kernels from the realistic images of satellite sensor B. KernelGAN is an unsupervised image-specific internal-GAN, which learns the internal distribution of the input imagery and generates a downsampled version of it. KernelGAN needs to meet the following optimization objective during the training phase:

$$arg\min_G \min_D \{ \mathbb{E}_{x\sim patches(I_{In})}[|D(x) - 1| + |D(G(x))|] + \mathscr{R} \} \tag{2}$$

where $\times$ denotes the patch extracted from the input image I_{In}, and G and D denote the generator and the discriminator, respectively. $\mathscr{R}$ is the regularization term on the kernel. Since the generator of KernelGAN is a deep linear network consisting of several linear layers with no activations, the kernel that is an array can be obtained by convolving all the filters of the generator. Mathematically, estimating the kernels by the use of KernelGAN can be formulated as follows:

$$k = F_{KernelGAN}(I_{LR\_T}) \tag{3}$$

where I_{LR_T} is the input realistic LR image of target sensor B, and k represents the estimated kernel. By estimating the kernels on a large amount of real LR images from target sensor B, it is possible to construct a blur-kernel pool $\{k_1, k_2 \cdots, k_m\}$.

After building a kernel pool, the noise is also taken into consideration to generate more realistic degraded images. Noise patches are directly collected from the noise-dominant LR images from target sensor B, which can have a more similar noise distribution to the realistic images. It is assumed that the expectation of the noise distribution is zero, and a low-pass filter is applied to a noisy LR image to generate a smooth image. Subtracting the smooth image from the noisy LR image then yields an approximate noise patch. To reduce the influence of complex background information, a set of image blocks with heterogeneous textures (e.g., lakes and bare soil) are first extracted from the LR images. Furthermore, a standard is applied to collect noise patches from these eligible image blocks:

$$|Mean(|p|) - Mean(|q|)| \le Mean(|q|) and |Var(p) - Var(q)| \le Var(q) \tag{4}$$

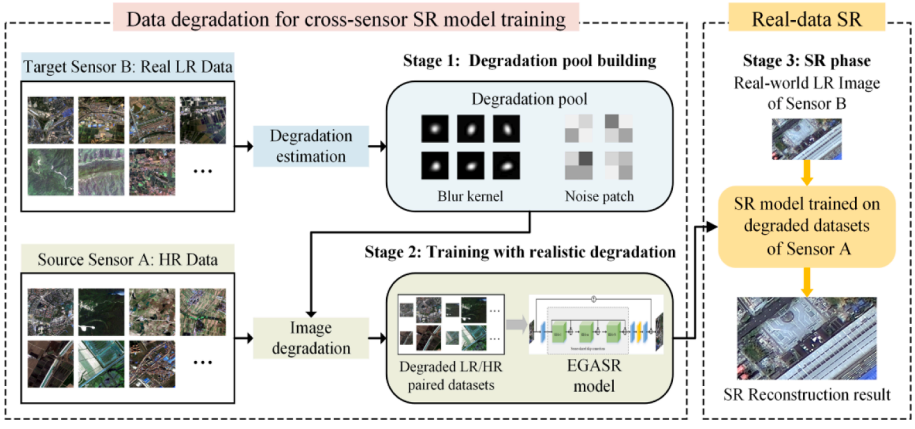

Fig. 1. The framework of the proposed cross-sensor SR method.

where p is the local noise patch (i.e., size of 40 × 40 in our tests), q is the global image block (i.e., size of 2000 × 2000), and $Mean(\cdot)$ and $Var(\cdot)$ denote the mean and variance calculations, respectively. It is then possible to utilize the collected series of noise patches $\{n_1, n_2 \cdots, n_l\}$, together with the blur-kernel pool, to build a joint degradation pool.

2.1.2. Cross-sensor realistic image super-resolution

With the degradation pool constructed, the simulation process for cross-sensor training can be implemented as follows. The images from source sensor A are used as the HR images, and the degraded LR images can be synthesized as:

$$I_{LR\_S} = (I_{HR\_S} \otimes k_i)\downarrow_s + n_j, i \in \{1, 2 \cdots, m\} and j \in \{1, 2 \cdots, l\} \tag{5}$$

where I_{HR_S} and I_{LR_S} represent the HR image and its degraded LR version, respectively; k_i and n_j are the i-th kernel and j-th noise patch randomly selected from the degradation pool, respectively; and $\downarrow_s$ denotes the strided downsampling operation.

Finally, a degraded LR-HR paired dataset simulated from the images of source sensor A can be obtained and used to train the EGASR model. The trained model can then be applied to super-resolve the realistic LR image I_{LR_T} of target sensor B.

2.2. EGASR network architecture

2.2.1. Overall pipeline

The architecture of the proposed EGASR network is shown in Fig. 2. The degraded LR-HR paired dataset simulated from the images of source sensor is used to train the EGASR model. Given an LR image $I_{LR\_S} \in \mathbb{R}^{H \times W \times N}$ with N bands as input, a convolutional layer with kernel size of 3 × 3 is first applied to obtain the initial shallow feature map:

$$F_0 = f_{Conv3}(I_{LR\_S}) \tag{6}$$

where $f_{Conv3}(\cdot)$ represents the 3 × 3 convolution operation, and $F_0 \in \mathbb{R}^{H \times W \times C}$ is the extracted feature map. To make full use of the shallow layer information and capture long-range dependencies to improve the model performance, a source-shared skip connection (SSKC) structure is added in the EGASR network. The initial feature map F_0 is fed into each of a stack of residual edge-enhanced groups (REGs), where the deep feature map of the n-th REG can be expressed as follows:

$$F_{Gn} = f_{REGn}(F_{Gn-1}) + W_S F_0 \tag{7}$$

where $f_{REGn}(\cdot)$ denotes the function of the n-th REG; F_{Gn-1} and F_{Gn} represent the input and output feature maps of the n-th REG, respectively; and W_S is a learnable parameter. The final output deep feature of the REGs is obtained as:

$$F_D = F_{GN} + W_S F_0 \tag{8}$$

where F_{GN} is the output feature of the last REG, which is then passed into a convolutional layer followed by an upsampled module:

$$F_{UP} = f_{UP}(f_{Conv3}(F_{DF})) \tag{9}$$

where F_{UP} and $f_{UP}(\cdot)$ represent the upsampled features and upsampled module, respectively. In the proposed approach, the commonly used sub-pixel convolutional layer (Shi et al., 2016) is first used as the upsampling layer. F_{UP} is then passed into a convolutional layer to obtain the reconstructed feature:

$$F_R = f_{Conv3}(F_{UP}) \tag{10}$$

At the tail of the EGASR network, the final SR result can be obtained as follows:

$$I_{SR} = F_R + f_{BL}(I_{LR\_S}) = F_{EGASR}(I_{LR\_S}) \tag{11}$$

where $f_{BL}(\cdot)$ and $F_{EGASR}(\cdot)$ are the upsampled operation of bilinear interpolation and the function of EGASR, respectively; and "+" is the element-wise addition operation.

Finally, the EGASR network is optimized with a certain loss function. We chose to use the robust Charbonnier loss function (Lai et al., 2017), which can handle outliers and improve the performance. The difference between the ground-truth HR image and the SR reconstruction result is minimized with the:

$$L_{SR} = \sqrt{\|I_{HR\_S} - I_{SR}\|^2 + \varepsilon^2} \tag{12}$$

where ε is an empirical constant, which was set to 1×10^{-3} in the experiments conducted in this study; and I_{SR} and I_{HR_S} are the SR result and the corresponding HR image, respectively.

2.2.2. Edge-guided residual attention block

As shown in Fig. 2, within each REG, several EGRABs and a 3 × 3 convolutional layer are stacked with short skip connections. In other words, the EGRAB is the core component and basic unit of the proposed network. The structure of the EGRAB is displayed in Fig. 2, which is made up of two residual blocks (RBs), an edge-enhanced spatial attention module (ESAM), and a multi-feature adaptive fusion module (MAFM).

(A) Residual block

In the EGRAB, given an initial feature $F_G \in \mathbb{R}^{H \times W \times C}$ as input, it is first processed by an RB that consists of two 3 × 3 convolutional layers and a parametric rectified linear unit (PReLU) activation function (He et al., 2015). The obtained immediate feature is then fed into the two branches, respectively. The RB in one of the branches is aimed at facilitating the feature representation and preserving the base performance of the network. The output feature of the RB can be written as:

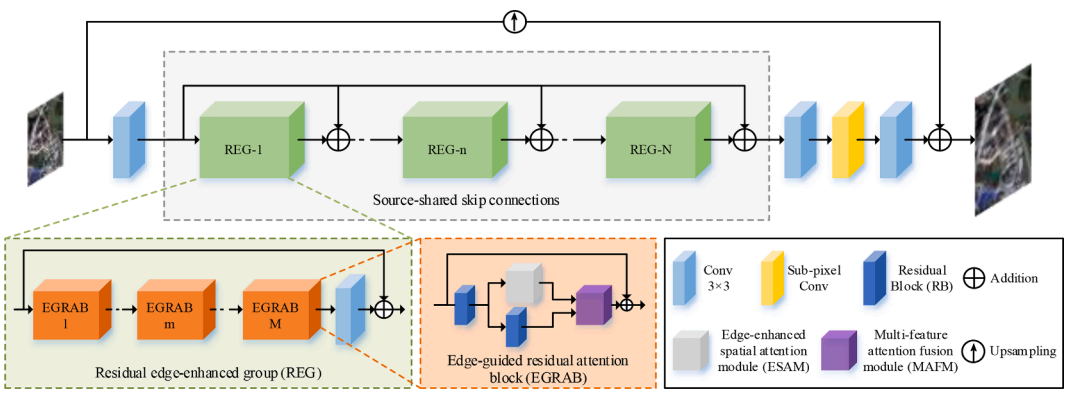

Fig. 2. Overview structure of the proposed EGASR network.

$$F_X = f_{Conv3}(\rho(f_{Conv3}(F_G))) \tag{13}$$

$$F_C = f_{Conv3}(\rho(f_{Conv3}(F_X))) \tag{14}$$

where $\rho(\cdot)$ denotes the PReLU activation function; and F_X and F_C represent the immediate feature and the obtained convolutional feature, respectively.

(B) Edge-enhanced spatial attention module

Spatial information determines the edges and textures of an image. In the proposed approach, an ESAM is constructed to capture and enhance the useful spatial information, to reconstruct images with more accurate and sharp details. Differing from some methods that build an additional branch of the network to explicitly extract edge maps using an edge operator, the edge extraction operation is implicitly incorporated into the design of the core block. As shown in Fig. 3, a three-branch structure is adopted to perform edge feature extraction in the head of the ESAM. The first-order Sobel operator and the second-order Laplace operator that are easily implemented and convenient to compute are separately employed in the three branches. The filter templates of the Sobel and Laplacian operators can be formulated as follows:

$$T_{Sx} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, T_{Sy} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ -1 & 2 & 1 \end{bmatrix} \tag{15}$$

$$T_{Lp} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{16}$$

where T_{Sx} and T_{Sy} denote the Sobel edge filter in the horizontal and vertical directions, respectively; and T_{Lp} represents the Laplacian edge filter. These edge operator templates are set as fixed kernels in the convolutional layers. In this way, the edge extraction operations are implicitly embedded in the ESAM. Specifically, in each branch, the input immediate feature F_X is first passed through a 1×1 convolutional layer to obtain the processed feature. The edge feature is then extracted from the processed feature by utilizing the edge filters, followed with a channel-wise scaling operation. The edge information extraction process can be expressed as:

$$\begin{aligned} F_{Sx} &= f_{Conv1}(F_X) \otimes (T_{Sx} * S_{Sx}) + B_{Sx} \\ F_{Sy} &= f_{Conv1}(F_X) \otimes (T_{Sy} * S_{Sy}) + B_{Sy} \\ F_{Lp} &= f_{Conv1}(F_X) \otimes (T_{Lp} * S_{Lp}) + B_{Lp} \end{aligned} \tag{17}$$

where $f_{Conv1}(\cdot)$ denotes the 1×1 convolutional layer; $\otimes$ denotes the depth-wise convolution; * represents the channel-wise broadcasting multiplication; S_{Sx}, S_{Sy}, S_{Lp}, B_{Sx}, B_{Sy}, and B_{Lp} are the scaling parameters and bias in the edge convolutional layer, respectively; and F_{Sx}, F_{Sy}, and F_{Lp} are the extracted edge features of the Sobel operator in the horizontal and vertical directions and the Laplacian operator, respectively. The edge features are concatenated and then fused by applying a 1×1 convolutional layer:

$$F_{edge} = f_{Conv1}([F_{Sx}, F_{Sy}, F_{Lp}]) \tag{18}$$

where [.,.] represents the concatenation operation. A spatial attention (SA) module is then utilized to enhance the edge information. In the SA module, the feature is first compressed in the channel dimension to obtain the maximum and average feature maps, which are concatenated and passed into a convolutional layer followed by a sigmoid activation function to generate an SA map. The enhanced features can be obtained by multiplying the edge features F_{edge} with the spatially refined map, which can be written as:

$$M_{SA} = \sigma(f_{Conv7}([P_{avg}(F_{edge}), P_{max}(F_{edge})])) \tag{19}$$

$$F_E = F_{edge} \otimes M_{SA} \tag{20}$$

where F_E represents the enhanced edge feature; $M_{SA} \in \mathbb{R}^{H \times W}$ represents the spatially refined map; $P_{avg}(\cdot)$ and $P_{max}(\cdot)$ denote the average pooling operation and the maximum pooling operation, respectively; $f_{Conv7}(\cdot)$ is the function of a 7×7 convolutional layer; and $\sigma(\cdot)$ represents the sigmoid activation function.

(C) Multi-feature adaptive fusion module

The ESAM in one branch captures and enhances the edge information, which is important for SR, while the RB in the other branch facilitates the base feature representation. The commonly used approaches for fusing different features include concatenation followed by a convolutional layer or direct summation. However, these simple feature aggregation approaches limit the expressive power of the features. Inspired by the excellent work of SKNet (Li et al., 2019), in the MAFM, soft attention is introduced to adaptively fuse the features coming from the two branches. The architecture of the MAFM is displayed in Fig. 4.

The features of the two branches are first aggregated by element-wise summation. A global average pooling operation is then performed for the added feature map in the spatial dimension to generate a channel-wise vector. The vector is squeezed and expanded by a fully connected layer and then activated by a sigmoid function to obtain the selective weights:

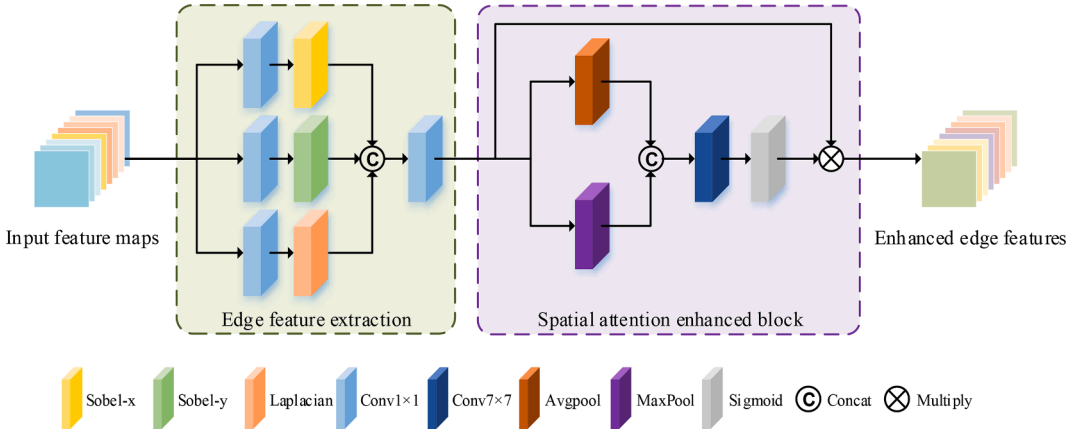

Fig. 3. The detailed structure of the edge-enhanced spatial attention module (ESAM), which consists of an edge feature extraction part and a spatial attention enhanced block.

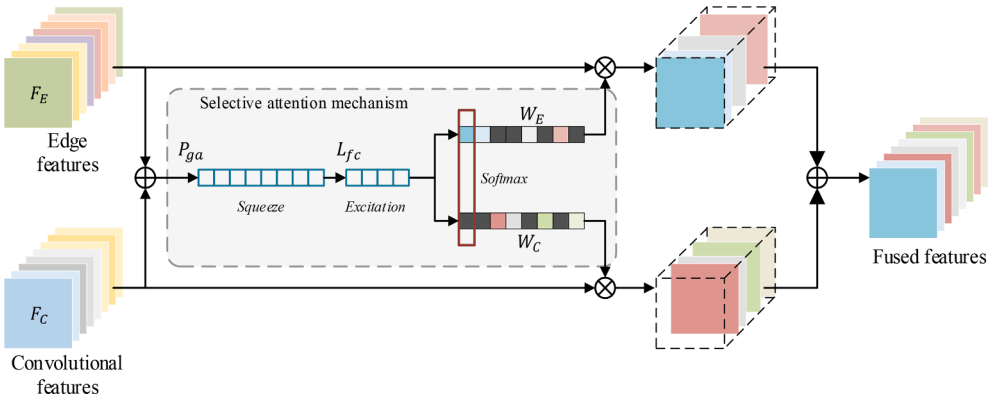

Fig. 4. The architecture of the multi-feature adaptive fusion module (MAFM).

$$z = L_{fc}\left(P_{ga}(F_E + F_C)\right) \tag{21}$$

$$d = max(C/r, L) \tag{22}$$

where $z \in \mathbb{R}^{d\times 1}$ is a compact feature for adaptive selection; r is the ratio of dimension reduction for computing d, which was set to 2 in the experiments; L is the minimum length of d, which was set to 32 in the experiments; and $P_{ga}(\cdot)$ and $L_{fc}(\cdot)$ denote the global average pooling function and fully connected layer, respectively.

Two attention weights are obtained by applying the softmax function on the channel-wise digits, which are utilized to adaptively recalibrate the two input features of the two branches. The process of the adaptive selection operation to obtain the final fused feature can be expressed as:

$$W_{Ck} = \frac{e^{C_k z}}{e^{C_k z} + e^{E_k z}}, W_{Ek} = \frac{e^{E_k z}}{e^{C_k z} + e^{E_k z}} \tag{23}$$

$$F_f = W_C * F_C + W_E * F_E \tag{24}$$

where W_C and W_E are the weights for the features of the two branches, respectively; $C, E \in \mathbb{R}^{C\times d}$ are the learnable parameters; $C_k \in \mathbb{R}^{1\times d}$ is the k-th row of C and W_{Ck} is the k-th element of W_C, likewise E_k and W_E; and $F_f \in \mathbb{R}^{H\times W\times C}$ is the final fused feature.

3. Experiments

3.1. Data preparation

In this study, we conducted experiments on datasets from three Gaofen (GF) series satellite sensors: Gaofen-1 (GF-1), Gaofen-2 (GF-2), and Gaofen-6 (GF-6), for which the spatial resolution and swath width properties are complementary (Table 1). To fully utilize the spatial and spectral information among the data, the images were processed with rational polynomial coefficient (RPC) correction, panchromatic/multispectral (PAN/MS) image registration, and pansharpening (Meng et al., 2019). Generally, the high-quality images can serve as the reference of the SR output in the training process for the SR tasks, as in (Wang et al., 2021a), and we prescribed a basic assumption to the proposed cross-sensor SR framework, i.e., the images acquired by the source sensor have higher quality than those of the target sensor. Therefore, among the different images, the GF-2 images with the highest spatial resolution of 1 m were used to construct the training dataset. The performance of the proposed SR method was verified from two aspects. Firstly, the ideally simulated GF-2 LR/HR image set with only downsampling considered was used to evaluate the superiority of the proposed EGASR network. To demonstrate the effectiveness of the proposed cross-sensor SR framework, we further adopted the model trained on the LR/HR GF-2 image set simulated with the estimated degradation pool to improve the spatial resolution of the GF-1 and GF-6 images, as shown in Fig. 1. To this end, two cross-sensor datasets were constructed for GF-2/GF-1 and GF-2/GF-6 in the experiments. The detailed information about the multi-sensor images and datasets is given in Table 1.

3.1.1. GF-2 simulated dataset

This dataset included the training and test sets, which were both obtained from GF-2. Taking the Wuhan urban agglomeration as the study region, we collected images of various scenes and different seasons in 2020. Specially, we conducted strict screening to ensure that the images were not perceptually contaminated with clouds and obvious artifacts in the process of collecting images. Finally, 950 images with the size of 1000 × 1000 were obtained. The HR images were downsampled to obtain LR images with the size of 500 × 500 × 4 (resp., 250 × 250 × 4) when the scale factor was 2 (resp., 4), which were the input for the network. These images are split into the training set and test set, while there are no overlaps existing between them. Note that part of the training set constructed the validation set, which is used to calculate the

Table 1
The technical specifications and image set information for the GF series of sensors.

	Satellite Sensors		GF-2	GF-1	GF-6
Technical specifications	Spectral range (μm)	Panchromatic	0.45 to 0.90	0.45 to 0.90	0.45 to 0.90
		Multispectral	0.45 to 0.52	0.45 to 0.52	0.45 to 0.52
			0.52 to 0.59	0.52 to 0.59	0.52 to 0.59
			0.63 to 0.69	0.63 to 0.69	0.63 to 0.69
			0.77 to 0.89	0.77 to 0.89	0.77 to 0.90
	Spatial resolution (m)	Panchromatic	1	2	2
		Multispectral	4	8	8
	Revisit interval (day)		5	2 (GF-1/GF-6 joint revisit period)	
	Scale range (km)		45	60	90
Image set information	Data domain		Source	Target	Target
	Number of collected images		850 (train), 100 (test)	100 (test)	100 (test)
	Size of collected images		1000 × 1000	500 × 500	500 × 500
	Resolution (pansharpening fusion) (m)		1	2	2

PSNR and loss value per epoch during the training phase. In this study, the GF-2 training dataset contained 850 LR/HR image pairs, 800 of which are used for training while another 50 for validation. The remaining 100 images formed the test set for assessing the performance of the trained model, covering four representative scenes of urban, rural, field, and mountain areas, each with 25 images. The test LR images were downsampled from the HR images with the same spatial size as the training data, where the spatial resolution was 2 m and 4 m with a scale factor of 2 and 4, respectively. The original HR images were used as the reference to calculate the quantitative metrics in the simulated tests.

3.1.2. GF-2/GF-1 cross-sensor dataset

The cross-sensor dataset used the images obtained from the different satellite sensors to make up the training and test data. In the experiments, the images in the cross-sensor training set were the same as those in the GF-2 simulated dataset. However, the LR images input to the network were degraded from the corresponding HR images, considering blur and noise (Sdraka et al., 2022; Yue et al., 2022). As mentioned in Section 2.1.1, we collected a random set of GF-1 images which also covered various scenes and estimated the blur kernels and noise features from these images to build a degradation pool for the target sensor. The cross-sensor training data were then constructed with the LR images simulated as described in Equation (1). The independent test set was composed of 100 real GF-1 images, each with a spatial resolution of 2 m and a size of 500 × 500 × 4. These images were collected from urban, rural, field, and mountain areas, each with 25 images.

3.1.3. GF-2/GF-6 cross-sensor dataset

The other cross-sensor dataset was constructed in the same way as the GF-2/GF-1 dataset, while the target sensor was GF-6. The results included the training set composed of 850 HR GF-2 images along with corresponding LR image pairs simulated with the degradation pool for GF-6, and the test set composed of 100 GF-6 images covering urban, rural, field, and mountain areas, each with 25 images. The details can be found in Table 1.

3.2. Implementation details

In the training phase, we augmented the training data by randomly employing 90°, 180°, and 270° rotation and horizontal and vertical flipping (Wang et al., 2021b). In each mini-batch, 16 LR images with a patch size of 48 × 48 were provided as inputs for the model, and the corresponding HR image served as the ground truth for calculating the loss. The LR-HR paired data and the output of the network were all four-band images.

The models were optimized using the ADAM optimizer (Kingma and Ba, 2014) with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$. The initial learning rate was set to 10^{-4} and then decreased by half every 250 epochs. A total of 1000 epochs were used for training the models since more epochs did not bring further improvements. The proposed EGASR network was implemented using Python 3.8 and the PyTorch 1.8 framework on an Nvidia GeForce RTX 3090 GPU.

3.3. Evaluation metrics

In this paper, we use the commonly used full-reference image evaluation metrics to evaluate the quality of the SR reconstruction results, i.e., the peak signal-to-noise ratio (PSNR), the structural similarity index (SSIM) (Wang et al., 2004), the spectral angle mapper (SAM) (Vivone et al., 2015), the relative dimensionless global error in synthesis (ERGAS) (Wald, 2002), and the spatial correlation coefficient (SCC) (Zhou et al., 1998). Among them, PSNR and ERGAS are two metrics in terms of comprehensive spatial-spectral image quality, SSIM and SCC are used to quantify the spatial quality, and SAM is utilized to quantify the spectral distortion. We calculated the mean values of these metrics across all the recovered spectral bands in the SR images. Specially, larger values of PSNR, SSIM, and SCC, and smaller values of SAM and ERGAS indicate better SR results.

In addition, we introduce no-reference image evaluation metrics, i.e., the natural image quality evaluator (NIQE) (Mittal et al., 2013), average gradient (AG) (Chen et al., 2018), the spatial frequency (SF) (Eskicioglu and Fisher, 1995), and the entropy (Li et al., 1999) to evaluate the real-world SR results, without an HR image for reference. Smaller values of NIQE and larger values of AG, SF, and entropy indicate better SR results.

3.4. Comparison with other CNN-based methods on the simulated datasets

In this section, we describe the simulated experiments conducted on the GF-2 dataset to validate the SR reconstruction performance of the proposed EGASR network. The comparison methods included several state-of-the-art CNN-based methods, i.e., the enhanced deep residual SR network (EDSR) (Lim et al., 2017), the residual channel attention network (RCAN) (Zhang et al., 2018), the second-order attention network (SAN) (Dai et al., 2019), the holistic attention network (HAN) (Niu et al., 2020), and the soft-edge assisted network (SeaNet) (Fang et al., 2020). Among them, EDSR wins the champion in NTIRE 2017 Challenge on Single Image Super-Resolution, RCAN, SAN, and HAN are representative SISR methods that use attention mechanism, and SeaNet is a well-performed edge-assisted SR method that introduces edge prior into CNN. For a fair comparison, all the models were trained on the simulated GF-2 training dataset, as introduced in Section 3.1.1. The trained models were then evaluated with the independent test set.

The quantitative results in terms of PSNR, SSIM, SCC, ERGAS, and SAM are reported in Table 2, where the bold font represents the best performance for each index. It can be seen that EGASR achieves the highest PSNR, SSIM, and SCC values and the lowest ERGAS and SAM values, on average, for the four scenes on both the × 2 and × 4 SR. Compared with the CNN-based methods, the proposed EGASR method obtains better results for all four test scenes. Especially on the urban scene with rich texture information, EGASR surpasses most of the comparison methods by a large margin in terms of PSNR. Even when compared with SAN with the second-best score, EGASR still shows a PSNR value gain of 0.17 dB and 0.15 dB, on average, for the four scenes in the × 2 and × 4 SR results, respectively.

Fig. 5 and Fig. 6 show visual comparisons for the test images in the GF-2 test set with × 2 and × 4 SR, respectively. The zoomed-in views within the red box are provided for facilitating the visual comparison. It can be seen that the EGASR method can reconstruct images with more clean details and sharper edges than the other CNN-based SR methods. For example, in "Img_07" of the urban scene in Fig. 5, the textures of the steps within the SR images of the other CNN-based methods suffer from different degrees of blurring, while EGASR gives clear details. Similarly, the CNN-based methods loss the internal texture of the field in "Img_05", and only EGASR can recover more details in the field which are more faithful to the ground truth. For the × 4 SR, as shown in "Img_03" of the urban area in Fig. 6, the other methods generate images with fuzzy artifacts, while EGASR can reconstruct the main structure of the building site. In particular, in "Img_11" of the rural scene, all the compared methods over-smooth the small white objects above the roof and on the side of the street. In contrast, the white objects in the SR results of EGASR contain clearer contours and can be distinguished well. In general, compared with the other SR methods, the proposed EGASR method shows superiority in reconstructing finer details and sharper edges of the ground objects in remote sensing images.

To further study the reconstruction accuracy of the SR results, boxplots of the root-mean-square error (RMSE) between the SR results and the HR references in each band are displayed in Fig. 7. It can be seen from the indicators (e.g., median and mean, box region from the 25th to the 75th, 1st, and the 99th percentile value) in Fig. 7 that the CNN-based methods are clearly superior to bicubic interpolation, and the proposed

Table 2
Quantitative comparison on the GF-2 test set for 2 × and 4 × SR, where the bold font indicates the best performance.

Scale	Test set	Metric	Bicubic	EDSR	RCAN	SAN	HAN	SeaNet	Proposed
×2	Urban	PSNR	40.0131	43.9863	44.2242	44.4884	44.4044	44.3710	**44.6538**
		SSIM	0.9612	0.9833	0.9840	0.9848	0.9846	0.9845	**0.9854**
		SCC	0.7113	0.8562	0.8641	0.8721	0.8693	0.8684	**0.8768**
		ERGAS	2.6663	1.6877	1.6422	1.5927	1.6086	1.6145	**1.5631**
		SAM	0.4557	0.3199	0.3196	0.3053	0.3028	0.3042	**0.2933**
	Rural	PSNR	44.3498	48.4423	48.7077	48.8414	48.8177	48.7905	**49.0562**
		SSIM	0.9791	0.9914	0.9918	0.9920	0.9920	0.9920	**0.9924**
		SCC	0.7545	0.8742	0.8815	0.8849	0.8840	0.8834	**0.8902**
		ERGAS	2.4814	1.5486	1.5040	1.4795	1.4856	1.4889	**1.4461**
		SAM	0.6760	0.4410	0.4342	0.4243	0.4245	0.4246	**0.4124**
	Field	PSNR	47.4634	49.8121	49.9032	49.9347	49.9991	49.9501	**50.0620**
		SSIM	0.9784	0.9875	0.9879	0.9880	0.9881	0.9880	**0.9883**
		SCC	0.7315	0.8070	0.8114	0.8126	0.8151	0.8134	**0.8179**
		ERGAS	1.6331	1.2469	1.2311	1.2267	1.2192	1.2251	**1.2081**
		SAM	0.6083	0.4859	0.4764	0.4695	0.4707	0.4714	**0.4602**
	Mountain	PSNR	45.7107	49.6216	49.9193	50.1594	50.0897	50.0538	**50.3344**
		SSIM	0.9837	0.9930	0.9933	0.9936	0.9935	0.9935	**0.9938**
		SCC	0.7715	0.8820	0.8898	0.8957	0.8938	0.8931	**0.8997**
		ERGAS	1.6120	1.0277	0.9939	0.9667	0.9742	0.9779	**0.9478**
		SAM	0.4159	0.2793	0.2735	0.2683	0.2668	0.2679	**0.2612**
	Average	PSNR	44.3843	47.9656	48.1886	48.3560	48.3277	48.2913	**48.5266**
		SSIM	0.9756	0.9888	0.9892	0.9896	0.9896	0.9895	**0.9899**
		SCC	0.7422	0.8548	0.8617	0.8663	0.8655	0.8646	**0.8711**
		ERGAS	2.0982	1.3777	1.3428	1.3164	1.3219	1.3266	**1.2913**
		SAM	0.5390	0.3815	0.3759	0.3669	0.3662	0.3670	**0.3568**
×4	Urban	PSNR	34.2845	35.5513	35.6013	35.7466	35.7067	35.6648	**35.9496**
		SSIM	0.8501	0.8850	0.8860	0.8895	0.8888	0.8875	**0.8942**
		SCC	0.2868	0.4080	0.4127	0.4309	0.4254	0.4219	**0.4555**
		ERGAS	2.5777	2.2281	2.2153	2.1786	2.1885	2.1992	**2.1285**
		SAM	0.8983	0.7811	0.7968	0.7760	0.7702	0.7671	**0.7466**
	Rural	PSNR	37.6319	39.1304	39.1182	39.2571	39.2347	39.2075	**39.3956**
		SSIM	0.9038	0.9294	0.9292	0.9312	0.9310	0.9305	**0.9332**
		SCC	0.3462	0.4757	0.4723	0.4889	0.4846	0.4838	**0.5029**
		ERGAS	2.6890	2.2651	2.2679	2.2320	2.2375	2.2453	**2.1964**
		SAM	1.4778	1.2160	1.2313	1.2055	1.2042	1.2039	**1.1831**
	Field	PSNR	41.6935	42.3938	42.4123	42.4623	42.4387	42.4390	**42.5337**
		SSIM	0.9197	0.9301	0.9307	0.9312	0.9307	0.9308	**0.9322**
		SCC	0.4044	0.4389	0.4387	0.4418	0.4415	0.4426	**0.4517**
		ERGAS	1.5886	1.4684	1.4639	1.4558	1.4608	1.4613	**1.4441**
		SAM	1.1865	1.1115	1.1216	1.1047	1.1084	1.1061	**1.0968**
	Mountain	PSNR	39.2030	40.8179	40.8380	40.9601	40.9616	40.9313	**41.0953**
		SSIM	0.9298	0.9521	0.9522	0.9536	0.9537	0.9532	**0.9550**
		SCC	0.4204	0.5419	0.5421	0.5553	0.5552	0.5518	**0.5683**
		ERGAS	1.7054	1.4091	1.4049	1.3840	1.3842	1.3908	**1.3634**
		SAM	0.8770	0.6609	0.6769	0.6573	0.6549	0.6513	**0.6419**
	Average	PSNR	38.2032	39.4733	39.4924	39.6065	39.5854	39.5606	**39.7436**
		SSIM	0.9009	0.9241	0.9245	0.9264	0.9261	0.9255	**0.9286**
		SCC	0.3644	0.4661	0.4665	0.4793	0.4767	0.4750	**0.4946**
		ERGAS	2.1402	1.8427	1.8380	1.8126	1.8178	1.8241	**1.7831**
		SAM	1.1099	0.9424	0.9567	0.9359	0.9344	0.9321	**0.9171**

EGASR method obtains the highest accuracy in all four bands, which indicates that the SR result of EGASR is closest to the HR reference.

3.5. Cross-sensor SR with real datasets

Cross-sensor SR with real datasets was conducted to evaluate the performance of the SR methods on real-world images. Note that the CNN-based SR models, including the proposed EGASR method, were all trained on the GF-2 training set with bicubic degradation. In the real-data experiments, the proposed cross-sensor SR framework was applied to improve the generalization of the EGASR method to real-world images and obtain a robust SR model, namely EGASR-CS. We tested all the SR models on the GF-1 and GF-6 realistic test sets with a scale factor of 2.

The quantitative results in terms of AG, SF, entropy, and NIQE for the GF-1 and GF-6 test sets are reported in Tables 3 and 4, respectively. It can be seen that the CNN-based methods trained with bicubic degradation obtain close results, among which the proposed EGASR method obtains superior results. It is worth noting that EGASR-CS far surpasses all the methods by a large margin in terms of all no-reference evaluation metrics, which indicates that the images reconstructed by EGASR-CS possess better visual fidelity and richer details.

We also qualitatively compared the performance of the different methods in super-resolving realistic images. The visual results for the GF-1 and GF-6 test sets are displayed in Fig. 8 and Fig. 9, respectively. As shown in Fig. 8, all the CNN-based methods obtain similar blurred images, which are only slightly better than the bicubic method. In contrast, EGASR-CS can obtain visually pleasing results with sharper edges and details. Fig. 9 gives similar trends. Taking “Img_09” from the urban area as an example, only EGASR-CS is capable of recovering the outlines on the roof and reconstructing the small white objects with clear contours. The results obtained on the realistic test sets confirm that the proposed cross-sensor framework is plausible and competitive in practical use, compared to the supervised methods trained with an ideal dataset.

Furthermore, to explore the applicability of the SR result of the proposed method in further applications, we conducted ground feature extraction experiments on the SR results of the bicubic, EGASR, and EGASR-CS methods in super-resolving GF-1 and GF-6 real-world images.

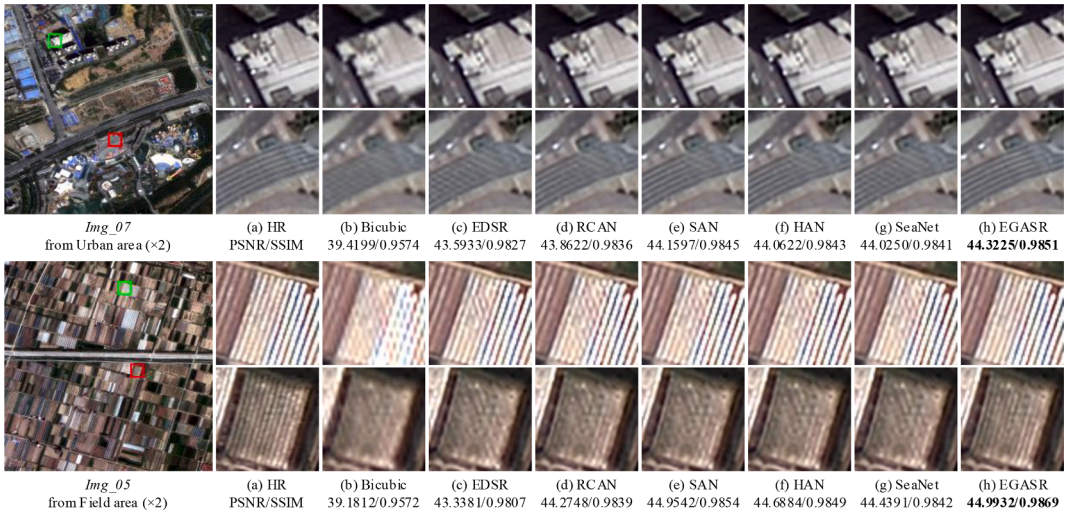

Fig. 5. Qualitative results obtained on the GF-2 simulated test sets with bicubic degradation for 2 × SR. The different SR results within the red box are zoomed in on for better visualization. The best quantitative results are marked in bold.

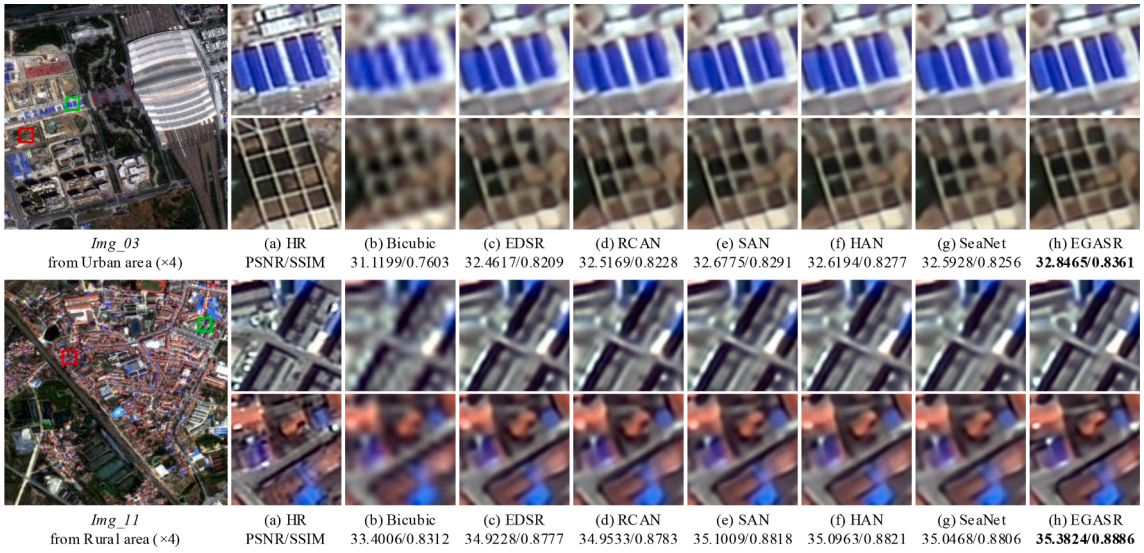

Fig. 6. Qualitative results obtained on the GF-2 simulated test sets with bicubic degradation for 4 × SR. The different SR results within the red box are zoomed in on for better visualization. The best quantitative results are marked in bold.

Specifically, the built-in edge-based segmentation algorithm in ENVI was first utilized to segment the image at multiple scales, followed by the full lambda-schedule algorithm (Robinson et al., 2002) to fuse adjacent small patches with spatial and spectral feature information. We utilized the segment-only feature extraction workflow in ENVI 5.3 to implement the complete extraction process, and set the parameters to the same values in each step for the different SR reconstruction results used in the experiments. The ground object extraction results are displayed in Fig. 10. Taking "Img_005" from the urban area as example, by comparing the extracted ground objects of the SR images generated by each method, the number of extracted ground objects in the SR results of EGASR-CS are much more than those in the SR results of the bicubic and EGASR methods. In particular, the very small white objects next to the buildings can be extracted in the SR results of EGASR-CS, while almost no extraction results can be seen in the results of the other two methods. Overall, the extraction results of EGASR-CS are better than those of EGASR in object boundaries and separability, which demonstrates the effectiveness of the proposed cross-sensor SR framework in improving the model's ability to super-resolve HR images with more high-frequency details in practical applications.

4. Discussion

4.1. Ablation studies

In Section 3.4, we demonstrated the superiority of the proposed EGASR network over the other compared SISR networks. In this subsection, we describe how we further conducted a set of ablation experiments to analyze the effects of some of the important components in EGASR, including the ESAM and MAFM in the EGRAB, and the SSKC structure. The quantitative results of the ablation studies conducted on the GF-2 simulated test set for × 4 SR are given in Table 5.

The baseline model was obtained by removing all three components in the standard EGASR network. From Table 5, it can be observed that Model1 can bring a 0.05 dB improvement in terms of PSNR over the baseline model, which is largely due to the fact that the SSKC structure can transfer rich low-level information to the deep layers, to improve the SR performance.

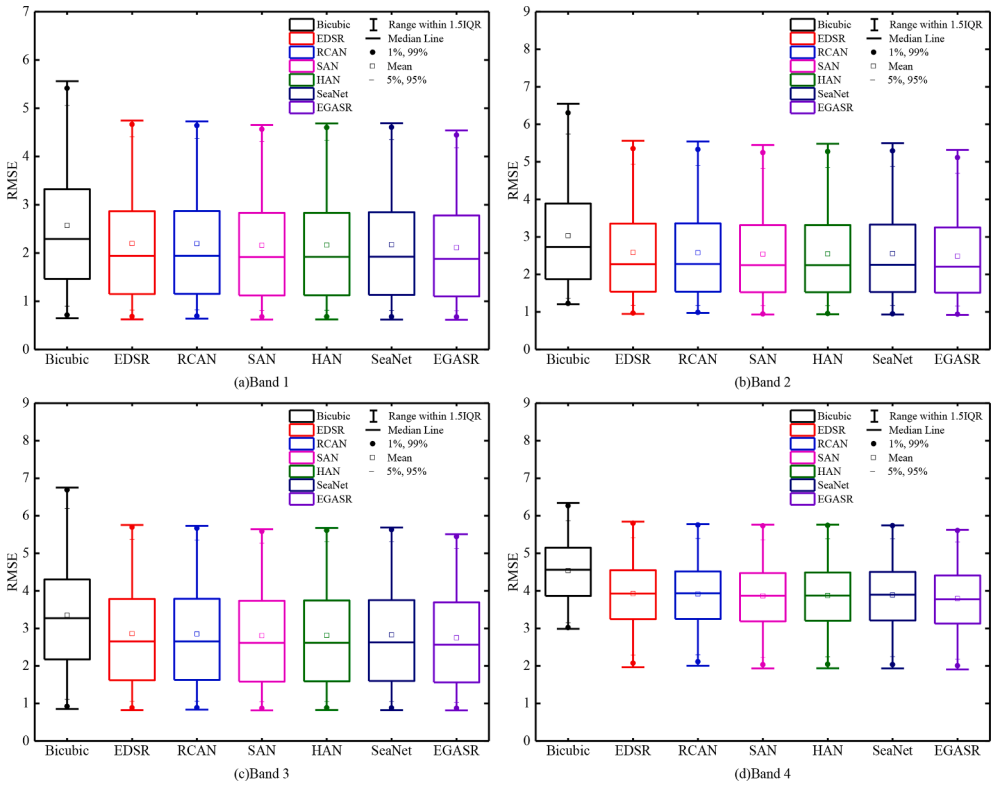

Fig. 7. Boxplots of the RMSEs between the SR results and HR images of the different methods in each band.

Table 3
Quantitative comparison with the GF-1 real-data test sets for 2 × SR, where the bold font indicates the best performance.

Test set	Metric	Bicubic	EDSR	RCAN	SAN	HAN	SeaNet	EGASR	EGASR-CS
Urban	AG	18.3541	19.1799	19.2451	19.2705	19.2408	19.2481	19.2917	**33.5140**
	SF	4.3141	4.6946	4.6974	4.7022	4.7004	4.7096	4.6834	**8.4237**
	entropy	5.7399	5.7450	5.7458	5.7458	5.7452	5.7454	5.7457	**5.9074**
	NIQE	9.4480	7.8692	7.9080	7.9116	7.8408	7.8825	7.9236	**5.1154**
Rural	AG	13.1004	13.6893	13.7449	13.7567	13.7293	13.7459	13.7675	**23.4204**
	SF	3.5134	3.7727	3.7862	3.7865	3.7850	3.7921	3.7682	**6.6228**
	entropy	5.4782	5.4827	5.4839	5.4837	5.4833	5.4835	5.4838	**5.6537**
	NIQE	9.5677	7.8232	7.8540	7.8387	7.8075	7.8462	7.8153	**5.0728**
Mountain	AG	10.6581	11.3480	11.3774	11.3973	11.3729	11.4091	11.4048	**23.0941**
	SF	2.4754	2.6539	2.6592	2.6655	2.6622	2.6663	2.6632	**5.4339**
	entropy	4.1365	4.1517	4.1553	4.1581	4.1535	4.1586	4.1573	**4.5751**
	NIQE	12.5012	10.9077	10.8204	10.8373	10.6980	10.8457	10.7458	**8.6518**
Field	AG	7.2079	7.7298	7.7650	7.7698	7.7587	7.7598	7.7799	**11.3526**
	SF	2.2527	2.4464	2.4542	2.4575	2.4571	2.4590	2.4526	**3.5956**
	entropy	4.7968	4.7991	4.8001	4.7998	4.7992	4.7995	4.8004	**4.8807**
	NIQE	10.3366	8.0894	8.0317	8.1000	8.0241	8.1096	8.1148	**5.3356**
Average	AG	12.3301	12.9868	13.0331	13.0486	13.0254	13.0407	13.0610	**22.8453**
	SF	3.1389	3.3919	3.3993	3.4030	3.4012	3.4067	3.3919	**6.0190**
	entropy	5.0379	5.0446	5.0463	5.0469	5.0453	5.0467	5.0468	**5.2542**
	NIQE	10.4634	8.6724	8.6535	8.6719	8.5926	8.6710	8.6499	**6.0439**

The quantitative comparison between the models with and without using the ESAM are reported in Table 5, where it can be found that Model2 using the ESAM outperforms Model1 without using the ESAM by a large margin in terms of PSNR. Fig. 11 displays the visualization results of Model1 and Model2. Model1 without using the ESAM generates images with distorted textures and fuzzy artifacts, while Model2 using the ESAM can recover sharper images with clearer edges and textures. The quantitative and qualitative results demonstrate that introducing the edge prior into the network using the ESAM contributes to the preservation of the structural information.

Finally, it can be found that the PSNR value increases from 39.69 to 39.74 dB when comparing the evaluation metrics of Model2 and EGASR. These results prove that the MAFM is an important component for fusing the different features of the two branches, instead of simply combining the features.

4.2. Discussion on the edge extraction operators

The Sobel operator is a combination of Gaussian smoothing and a differential operation, which has a strong anti-noise ability, while the Laplacian operator is isotropic and can extract edges in any direction. To explore the function of extracting edge features using these operators to guide the SR process, we designed several models with different edge operator combinations in the ESAM and conducted a set of experiments. The quantitative results of these models are given in Table 6. In EGASR-woE, we removed the edge feature extraction operation via the edge

Table 4
Quantitative comparison on the GF-6 real-data test sets for 2 × SR, where the bold font indicates the best performance.

Test set	Metric	Bicubic	EDSR	RCAN	SAN	HAN	SeaNet	EGASR	EGASR-CS
Urban	AG	18.9287	19.4793	19.5203	19.5987	19.5495	19.5201	19.5302	**37.5616**
	SF	4.8825	5.1851	5.2067	5.2360	5.2196	5.2168	5.1893	**11.3187**
	entropy	6.6828	6.6820	6.6831	6.6820	6.6832	6.6845	6.6838	**6.7873**
	NIQE	9.3844	7.7936	7.8316	7.8436	7.8595	7.8237	7.7448	**5.0934**
Rural	AG	7.1191	7.1728	7.1762	7.2104	7.1765	7.1743	7.1802	**13.7465**
	SF	1.9594	1.9923	2.0066	2.0235	2.0044	2.0028	1.9920	**4.1548**
	entropy	6.2389	6.2369	6.2375	6.2385	6.2360	6.2369	6.2374	**6.3662**
	NIQE	11.5513	9.5748	9.4587	9.4238	9.4384	9.5798	9.4494	**4.8235**
Mountain	AG	7.6730	8.1199	8.2095	8.2326	8.2334	8.1436	8.1427	**14.7428**
	SF	1.8650	1.9857	2.0201	2.0373	2.0271	2.0040	1.9904	**3.5222**
	entropy	5.0265	5.0274	5.0370	5.0316	5.0335	5.0338	5.0345	**5.2676**
	NIQE	11.2700	8.9784	8.9666	8.5280	8.6301	8.5936	8.7090	**6.5728**
Field	AG	8.8020	9.1609	9.1846	9.2616	9.2178	9.1798	9.1700	**15.6220**
	SF	2.5808	2.7376	2.7710	2.8027	2.7742	2.7627	2.7336	**4.8067**
	entropy	5.5218	5.5231	5.5241	5.5259	5.5255	5.5253	5.5250	**5.6408**
	NIQE	8.1308	6.4915	6.0226	6.3303	6.3574	6.3206	6.3375	**4.6054**
Average	AG	10.6307	10.9832	11.0226	11.0758	11.0443	11.0045	11.0058	**20.4182**
	SF	2.8219	2.9752	3.0011	3.0249	3.0064	2.9966	2.9763	**5.9506**
	entropy	5.8675	5.8674	5.8704	5.8695	5.8696	5.8701	5.8701	**6.0155**
	NIQE	10.0841	8.2096	8.0699	8.0314	8.0714	8.0794	8.0602	**5.2738**

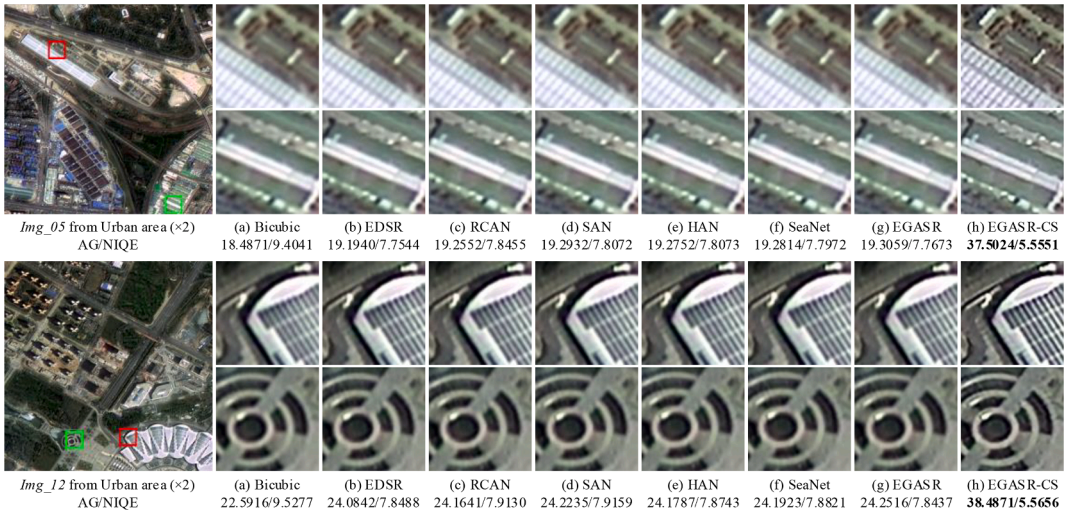

Fig. 8. SR results for the GF-1 real LR remote sensing images at scale × 2. (a) Bicubic. (b) EDSR. (c) RCAN. (d) SAN. (e) HAN. (f) SeaNet. (g) EGASR. (h) EGASR-CS.

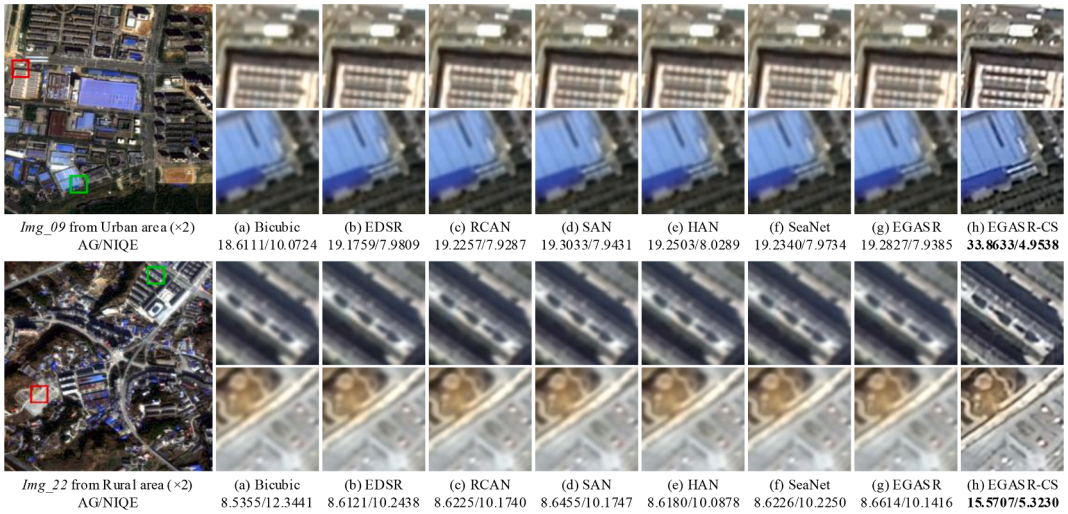

Fig. 9. SR results for the GF-6 real LR remote sensing images at scale × 2. (a) Bicubic. (b) EDSR. (c) RCAN. (d) SAN. (e) HAN. (f) SeaNet. (g) EGASR. (h) EGASR-CS.

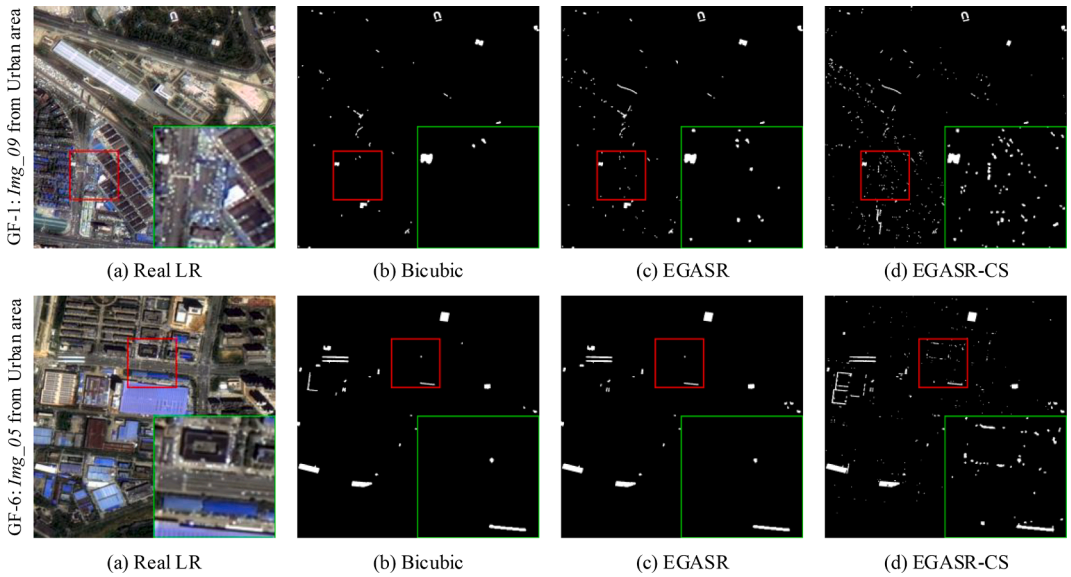

Fig. 10. Visual comparison of the ground features extracted from the SR results of the different methods: (a) real LR image, (b) bicubic, (c) EGASR, and (d) EGASR-CS.

Table 5
Ablation study on the GF-2 simulated test sets when the scale factor is 4.

Model	Baseline	Model1	Model2	EGASR
SSKC	×	✓	✓	✓
ESAM	×	×	✓	✓
MAFM	×	×	×	✓
PSNR	39.4944	39.5453	39.6913	**39.7436**
SSIM	0.9209	0.9247	0.9277	**0.9286**
SCC	0.4671	0.4727	0.4889	**0.4946**
ERGAS	1.8367	1.8271	1.7928	**1.7831**
SAM	0.9408	0.9339	0.9229	**0.9171**

operators but preserved the remaining convolutional layers and SA module in the ESAM. The first column represents the results of EGASR-woE, where it can be found that all the combinations that utilize edge operators surpass the performance of EGASR-woE, which does not employ the edge extraction operation. Moreover, EGASR-Sxy, which employs Sobel operators in both the horizontal and vertical directions, can obtain better results than EGASR-Sx and EGASR-Sy, which use only one direction. EGASR-Sxy can also surpass EGASR-Lap in terms of PSNR. The best performance is found when both the Sobel and Laplacian operators are combined.

In addition, we also display the visual results of these models in Fig. 12. The results reveal that EGASR-woE fails to recover the textures on the roof, while all the other models that utilize an edge prior can generate more details. EGASR can reconstruct clearer results than the other models. Both the quantitative and qualitative results demonstrate the effectiveness of edge guidance for recovering more high-frequency textures and details, so we use both the first-order and second-order differential edge operators in the ESAM.

4.3. Discussion on the effect of degradation factors

As mentioned in Section 2.1.1, noise and blur are considered in the image degradation process and in building the degradation pool, which

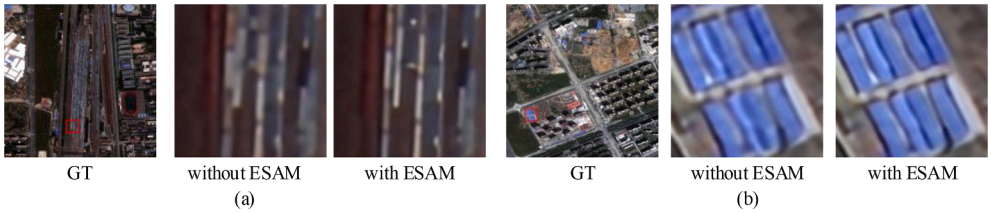

Fig. 11. Visual comparison between the networks using the ESAM and without the ESAM for 4 × SR on the GF-2 test set. (a) 13th image from the urban area. (b) 19th image from the urban area.

Table 6
Ablation study for the edge extraction operators in the ESAM when the scale factor is 4.

Model	EGASR-woE	EGASR-Sx	EGASR-Sy	EGASR-Sxy	EGASR-Lap	EGASR
Sobel-x	×	✓	×	✓	×	✓
Sobel-y	×	×	✓	✓	×	✓
Laplacian	×	×	×	×	✓	✓
PSNR	39.6711	39.6989	39.7105	39.7195	39.7037	**39.7436**
SSIM	0.9260	0.9279	0.9280	0.9285	0.9282	**0.9286**
SCC	0.4863	0.4903	0.4921	0.4929	0.4914	**0.4946**
ERGAS	1.7958	1.7916	1.7889	1.7881	1.7901	**1.7831**
SAM	0.9310	0.9209	0.9210	0.9171	0.9187	**0.9171**

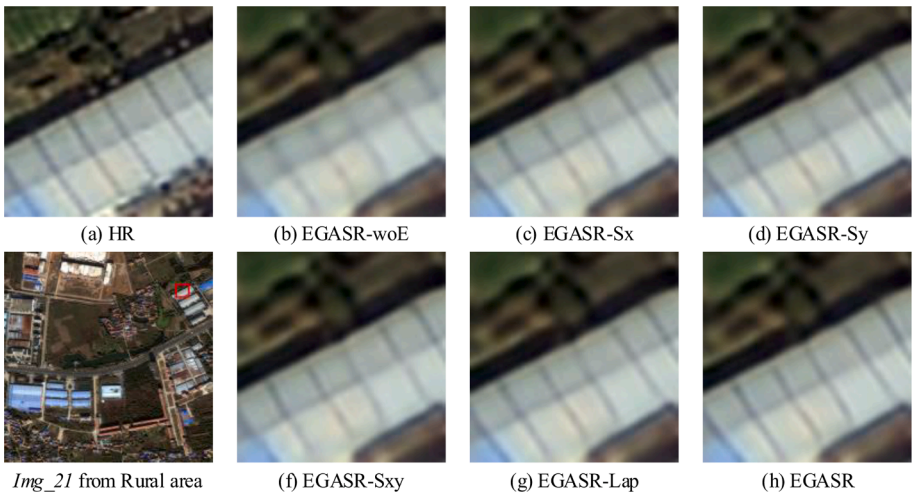

Fig. 12. Visual comparison between the networks using different combinations of edge operators for 4 × SR on the GF-2 test set.

consists of a blur-kernel pool and a noise pool. The EGASR-CS model that was trained with realistic degradation achieved an excellent performance in the real-data experiments. To investigate the effect of the degradation factors, we tested two independent cases with only noise or blur considered in the process of simulating the degraded images for training, which are denoted as EGASR-N and EGASR-B, respectively. We also took the EGASR model trained with bicubic degradation into consideration. The quantitative results are listed in Table 7. As shown in Table 7, the performance of EGASR-N is close to that of EGASR, and EGASR-CS and EGASR-B outperform them by a large margin on both the GF-1 and GF-6 real test sets in terms of all the no-reference metrics.

From the visual results in Fig. 13, it can also be found that both EGASR and EGASR-N obtain blurred results, while EGASR-B and EGASR-CS can generate sharper edges and clearer textures in the images. However, the SR image of EGASR-B is sharp but suffers from artifacts, while EGASR-CS gives a clear result. Overall, only considering noise in the degradation process cannot bring significant quantitative and qualitative improvements, and only considering blur fails to obtain visually pleasing results when encountering noise in real-world images. Only EGASR-CS that takes both noise and blur into consideration can obtain a promising performance in coping with a realistic scene.

4.4. Impacts and limitations of the cross-sensor SR framework

In this section, we further discuss the generalization performance and limitations of the proposed cross-sensor SR framework. Specifically, the EDSR and SAN methods with the relative lowest and highest accuracy among the compared deep learning based SR methods were selected and embedded in the cross-sensor SR framework to train two new models, namely EDSR-CS and SAN-CS models. The original and retrained EDSR and SAN models were tested on real-world GF-6 images, and the qualitative results are displayed in Fig. 14. It is clear that the performances of EDSR-CS and SAN-CS models are significantly improved with the proposed cross-sensor SR framework applied in the training process, reflecting on clearer edges and contours in the SR results (see Fig. 14(c) and (e)) than those of the original models (see Fig. 14(b) and (d)). The EGASR-CS method that introduced edge prior to promote the learning of edge information can still obtain sharper details in the SR results (see Fig. 14(f)), compared to the EDSR-CS and SAN-CS models trained in the same manner.

Table 7
Quantitative comparison of models trained with different degradation factors on the GF-1 and GF-6 real test sets.

Test set	Index\model	EGASR	EGASR-N	EGASR-B	EGASR-CS
GF-1	AG	13.0610	13.0573	22.7181	**22.8453**
	SF	3.3919	3.4061	6.0065	**6.0190**
	entropy	5.0468	5.0486	5.2506	**5.2542**
	NIQE	8.6499	8.6333	6.3116	**6.0439**
GF-6	AG	11.0058	11.2610	19.9765	**20.4182**
	SF	2.9763	3.0374	5.6844	**5.9506**
	entropy	5.8701	5.8861	5.9913	**6.0155**
	NIQE	8.0602	8.0331	5.3701	**5.2738**

The experimental results demonstrate that the well-designed cross-sensor SR framework has good generalization performance and can effectively improve the performance of deep learning-based SR methods adopted in real-world scenarios. In contrast to the image-pair-based SR methods (Joze et al., 2020; Wang et al., 2021a), which collected LR-HR image pairs acquired from different sensors for the same scene, we synthesized degraded training datasets via the cross-sensor degradation modeling. In this way, the trained model can bridge the gap between the images obtained by the source and target sensors, with no special need for cross-sensor accurate image registration and laborious processing operations. Another category of unsupervised domain-based methods (Maeda, 2020; Wei et al., 2021) show their advantages in directly capturing the underlying degradation process through learning with unpaired dataset. These methods implicitly modeled the complex real-world degradation and usually employed the GAN-based framework to learn the domain translation process (Chen et al., 2022). However, GAN-based frameworks can be difficult to train and often result in severe artifacts in the SR results (Liu et al., 2022a; Ma et al., 2022), which may hinder the use of reconstructed RS images in further applications. Furthermore, in contrast to such a one-stage unsupervised learning manner, we exploited the SR model in the framework to learn from dataset in a supervised manner, which results in a more stable training process and can achieve visually pleasing results.

While the proposed framework demonstrates promising results, there are potential limitations that need to be addressed. The degradation pool used in this study varies for different satellite sensors, which requires the training dataset needs to be reconstructed to achieve optimal results, and the model needs to be fine-tuned to adapt to dealing with different target sensors. Therefore, further study is needed to develop an adaptive strategy for the SR of real-world cross-sensor remote sensing images.

5. Conclusions

In this paper, we have proposed a cross-sensor SR framework for tackling the real-world remote sensing image SR task. The main contributions include a novel cross-sensor training strategy and an edge-

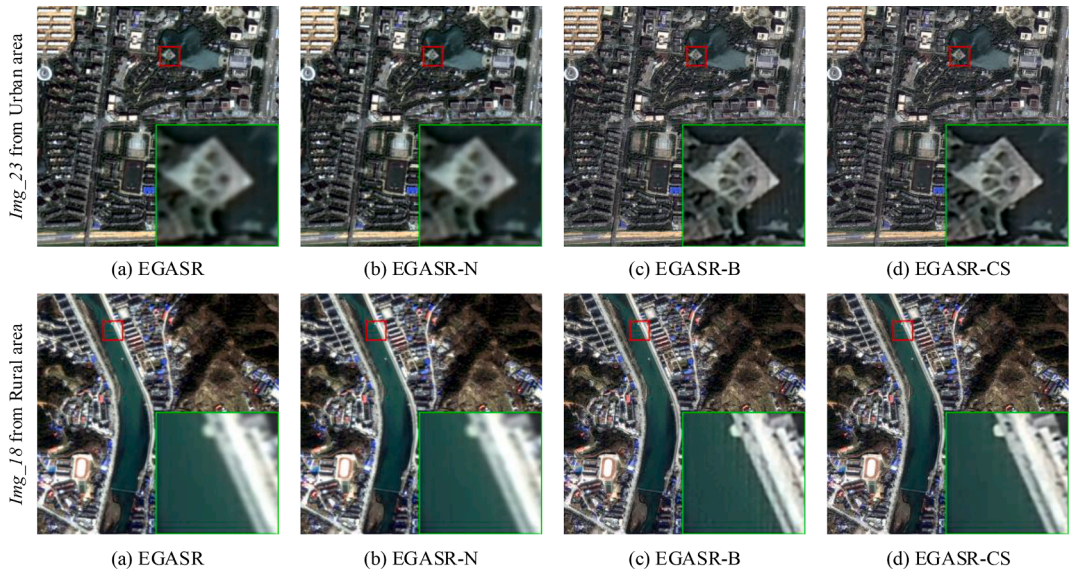

Fig. 13. Visual comparison between the models trained with different degradation factors for 2 × SR on the GF-6 realistic LR test set.

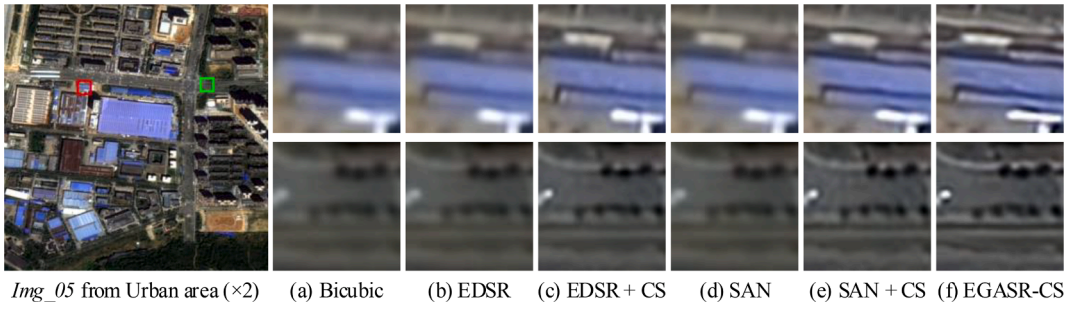

Fig. 14. Visual comparison between the models for 2 × SR on the GF-6 realistic LR test set. (a) Bicubic. (b) EDSR. (c) EDSR model trained under the cross-sensor SR framework. (d) SAN. (e) SAN model trained under the cross-sensor SR framework. (f) The proposed EGASR-CS method.

guided attention-based deep network. Firstly, to alleviate the domain shift between the images of the source sensor and target sensor, a degradation pool is first built by estimating the degradation factors, including blur and noise, from the LR images of the target sensor. The degraded LR images and the paired HR images constructed from the source domain are then used to train the deep learning model. Secondly, the EGASR network was proposed for super-resolving remote sensing images with complex edge details, which introduces an edge prior to orient the SR process in an implicit manner.

Both the simulated and real-data experiments on GF remote sensing images showed that the proposed method can achieve better evaluation metrics and visual results with higher fidelity and richer high-frequency information than the representative CNN-based SR methods used in the comparison. In the real cross-sensor experiments on GF-1/GF-2 and GF-6/GF-2 datasets, the proposed cross-sensor SR framework trained with the training datasets simulated from the source-sensor HR images significantly promoted the model's ability to super-resolve real-world LR images acquired by the target satellite sensor. The ablation analysis also validated the effectiveness of the edge prior in assisting the model to reconstruct sharper details in the SR images. In general, with the well-designed cross-sensor SR framework and open-source code package, the proposed method could potentially promote the extensive use of deep learning SR techniques in the processing of remote sensing imagery and related applications.

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.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 42130108 and Grant 42171375 and in part by the Open Fund of Hubei Luojia Laboratory under Grant 220100041. The authors would like to thank all the researchers who kindly shared the codes used in this paper.

References

Anwar, S., Barnes, N., 2022. Densely Residual Laplacian Super-Resolution. IEEE Trans. Pattern Anal. Mach. Intell. 44 (3), 1192–1204.
Bell-Kligler, S., Shocher, A., Irani, M., 2019. Blind super-resolution kernel estimation using an internal-gan. Adv. Neural Inf. Proces. Syst. 32, 284–293.
Chen, A.A., Chai, X., Chen, B., Bian, R. and Chen, Q., 2018. A novel stochastic stratified average gradient method: Convergence rate and its complexity, in: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1-8.
Chen, H., He, X., Qing, L., Wu, Y., Ren, C., Sheriff, R.E., Zhu, C., 2022. Real-world single image super-resolution: A brief review. Inform. Fusion 79, 124–145.
Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L., 2019. Second-Order Attention Network for Single Image Super-Resolution. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11057–11066.
Dong, C., Loy, C.C., He, K., Tang, X., 2015. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 38 (2), 295–307.

Dong, R., Mou, L., Zhang, L., Fu, H., Zhu, X.X., 2022. Real-world remote sensing image super-resolution via a practical degradation model and a kernel-aware network. ISPRS J. Photogramm. Remote Sens. 191, 155–170.
Dong, X., Wang, L., Sun, X., Jia, X., Gao, L., Zhang, B., 2021. Remote Sensing Image Super-Resolution Using Second-Order Multi-Scale Networks. IEEE Trans. Geosci. Remote Sens. 59 (4), 3473–3485.
Dong, W., Zhang, L., Shi, G., Wu, X., 2011. Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization. IEEE Trans. Image Process. 20 (7), 1838–1857.
Eskicioglu, A.M., Fisher, P.S., 1995. Image quality measures and their performance. IEEE Trans. Commun. 43 (12), 2959–2965.
Fang, F., Li, J., Zeng, T., 2020. Soft-Edge Assisted Network for Single Image Super-Resolution. IEEE Trans. Image Process. 29, 4656–4668.
Galar, M., Sesma, R., Ayala, C., Albizua, L. and Aranda, C., 2020. Super-Resolution of Sentinel-2 Images Using Convolutional Neural Networks and Real Ground Truth Data, Remote Sens.
Gao, X., Zhang, K., Tao, D., Li, X., 2012. Image Super-Resolution With Sparse Neighbor Embedding. IEEE Trans. Image Process. 21 (7), 3194–3205.
Haris, M., Shakhnarovich, G., Ukita, N., 2021. Deep Back-ProjectiNetworks for Single Image Super-Resolution. IEEE Trans. Pattern Anal. Mach. Intell. 43 (12), 4323–4337.
He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1026–1034.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778.
Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2020. Squeeze-and-Excitation Networks. IEEE Trans. Pattern Anal. Mach. Intell. 42 (8), 2011–2023.
Huan, L., Xue, N., Zheng, X., He, W., Gong, J., Xia, G.S., 2022. Unmixing Convolutional Features for Crisp Edge Detection. IEEE Trans. Pattern Anal. Mach. Intell. 44 (10), 6602–6609.
Jia, S., Wang, Z., Li, Q., Jia, X., Xu, M., 2022. Multiattention Generative Adversarial Network for Remote Sensing Image Super-Resolution. IEEE Trans. Geosci. Remote Sens. 60, 1–15.
Jiang, K., Wang, Z., Yi, P., Wang, G., Lu, T., Jiang, J., 2019. Edge-Enhanced GAN for Remote Sensing Image Superresolution. IEEE Trans. Geosci. Remote Sens. 57 (8), 5799–5812.
Joze, H.R.V., Zharkov, I., Powell, K., Ringler, C., Liang, L., Roulston, A., Lutz, M., Pradeep, V., 2020. ImagePairs: Realistic Super Resolution Dataset via Beam Splitter Camera Rig. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2190–2200.
Kang, X., Li, J., Duan, P., Ma, F., Li, S., 2022. Multilayer Degradation Representation-Guided Blind Super-Resolution for Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 60, 1–12.
Kingma, D.P. and Ba, J.J.a.p.a., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Lai, W.-S., Huang, J.-B., Ahuja, N., Yang, M.-H., 2017. Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 624–632.
Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., Schindler, K., 2018. Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network. ISPRS J. Photogramm. Remote Sens. 146, 305–319.
Lei, S., Shi, Z., Zou, Z., 2017. Super-Resolution for Remote Sensing Images via Local-Global Combined Network. IEEE Geosci. Remote Sens. Lett. 14 (8), 1243–1247.
Lei, S., Shi, Z., Mo, W., 2022. Transformer-Based Multistage Enhancement for Remote Sensing Image Super-Resolution. IEEE Trans. Geosci. Remote Sens. 60, 1–11.
Lepcha, D.C., Goyal, B., Dogra, A., Goyal, V., 2023. Image super-resolution: A comprehensive review, recent trends, challenges and applications. Information Fusion 91, 230–260.
Li, Y., Du, Z., Wu, S., Wang, Y., Wang, Z., Zhao, X., Zhang, F., 2021. Progressive split-merge super resolution for hyperspectral imagery with group attention and gradient guidance. ISPRS J. Photogramm. Remote Sens. 182, 14–36.
Li, X., Wang, W., Hu, X. and Yang, J., 2019. Selective Kernel Networks, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 510-519.
Li, X., Liu, G., Jinlin, N., 1999. Autofocusing of ISAR images based on entropy minimization. IEEE Trans. Aerosp. Electron. Syst. 35 (4), 1240–1252.
Li, Y., Mavromatis, S., Zhang, F., Du, Z., Sequeira, J., Wang, Z., Zhao, X., Liu, R., 2022. Single-Image Super-Resolution for Remote Sensing Images Using a Deep Generative Adversarial Network With Local and Global Attention Mechanisms. IEEE Trans. Geosci. Remote Sens. 60, 1–24.
Liang, J., Cao, J., Sun, G., Zhang, K., Gool, L.V. and Timofte, R., 2021. SwinIR: Image Restoration Using Swin Transformer, 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pp. 1833-1844.
Lim, B., Son, S., Kim, H., Nah, S. and Mu Lee, K., 2017. Enhanced deep residual networks for single image super-resolution, Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 136-144.
Liu, A., Liu, Y., Gu, J., Qiao, Y., Dong, C., 2022a. Blind Image Super-Resolution: A Survey and Beyond. IEEE Trans. Pattern Anal. Mach. Intell, 1–19.
Liu, M., Shi, Q., Marinoni, A., He, D., Liu, X., Zhang, L., 2022b. Super-Resolution-Based Change Detection Network With Stacked Attention Module for Images With Different Resolutions. IEEE Trans. Geosci. Remote Sens. 60, 1–18.
Lu, Z., Li, J., Liu, H., Huang, C., Zhang, L. and Zeng, T., 2022. Transformer for single image super-resolution, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 457-466.
Ma, C., Rao, Y., Lu, J., Zhou, J., 2022. Structure-Preserving Image Super-Resolution. IEEE Trans. Pattern Anal. Mach. Intell. 44 (11), 7898–7911.
Maeda, S., 2020. Unpaired Image Super-Resolution Using Pseudo-Supervision, 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 288-297.
Meng, X., Shen, H., Yuan, Q., Li, H., Zhang, L., Sun, W., 2019. Pansharpening for Cloud-Contaminated Very High-Resolution Remote Sensing Images. IEEE Trans. Geosci. Remote Sens. 57 (5), 2840–2854.
Mittal, A., Soundararajan, R., Bovik, A.C., 2013. Making a "Completely Blind" Image Quality Analyzer. IEEE Signal Process Lett. 20 (3), 209–212.
Molini, A.B., Valsesia, D., Fracastoro, G., Magli, E., 2019. DeepSUM: Deep neural network for Super-resolution of Unregistered Multitemporal images. IEEE Trans. Geosci. Remote Sens. 58 (5), 3644–3656.
Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., Zhang, K., Cao, X., Shen, H., 2020. Single Image Super-Resolution via a Holistic Attention Network. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-.-M. (Eds.), Computer Vision – ECCV 2020. Springer International Publishing, Cham, pp. 191–207.
Pan, Z., Ma, W., Guo, J., Lei, B., 2019. Super-Resolution of Single Remote Sensing Image Based on Residual Dense Backprojection Networks. IEEE Trans. Geosci. Remote Sens. 57 (10), 7918–7933.
Pouliot, D., Latifovic, R., Pasher, J., Duffe, J., 2018. Landsat Super-Resolution Enhancement Using Convolution Neural Networks and Sentinel-2 for Training. Remote Sens. 10 (3), 394.
Robinson, D.J., Redding, N.J. and Crisp, D.J., 2002. Implementation of a Fast Algorithm for Segmenting SAR Imagery. Electron. Res. Lab., Salisbury, SA, Australia, Tech. Rep. DSTO-TR-1242.
Sdraka, M., Papoutsis, I., Psomas, B., Vlachos, K., Ioannidis, K., Karantzalos, K., Gialampoukidis, I., Vrochidis, S., 2022. Deep Learning for Downscaling Remote Sensing Images: Fusion and Super-Resolution. IEEE Geosci. Remote Sens. Mag, 2–55.
Shen, H., Qiu, Z., Yue, L., Zhang, L., 2022. Deep-Learning-Based Super-Resolution of Video Satellite Imagery by the Coupling of Multiframe and Single-Frame Models. IEEE Trans. Geosci. Remote Sens. 60, 1–14.
Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D. and Wang, Z., 2016. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1874-1883.
Tai, Y., Yang, J. and Liu, X., 2017. Image super-resolution via deep recursive residual network, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3147-3155.
Tu, J., Mei, G., Ma, Z., Piccialli, F., 2022. SWCGAN: Generative Adversarial Network Combining Swin Transformer and CNN for Remote Sensing Image Super-Resolution. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 5662–5673.
Vivone, G., Alparone, L., Chanussot, J., Mura, M.D., Garzelli, A., Licciardi, G.A., Restaino, R., Wald, L., 2015. A Critical Comparison Among Pansharpening Algorithms. IEEE Trans. Geosci. Remote Sens. 53 (5), 2565–2586.
Wald, L., 2002. Data fusion: definitions and architectures: fusion of images of different spatial resolutions. Presses des MINES.
Wang, P., Bayram, B., Sertel, E., 2022a. A comprehensive review on deep learning based remote sensing image super-resolution methods. Earth Sci. Rev. 232, 104110.
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 (4), 600–612.
Wang, Z., Chen, J., Hoi, S.C.H., 2021b. Deep Learning for Image Super-Resolution: A Survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (10), 3365–3387.
Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., Yan, W., Zhao, J., Cai, Q. and Min, H., 2022b. A Review of Image Super-Resolution Approaches Based on Deep Learning and Applications in Remote Sensing, Remote Sens.
Wang, J., Gao, K., Zhang, Z., Ni, C., Hu, Z., Chen, D., Wu, Q., 2021a. Multisensor Remote Sensing Imagery Super-Resolution with Conditional GAN. Journal of Remote Sensing 2021, 9829706.
Wang, Q., Shi, W., Atkinson, P.M., Pardo-Igúzquiza, E., 2015. A new geostatistical solution to remote sensing image downscaling. IEEE Trans. Geosci. Remote Sens. 54 (1), 386–396.
Wei, Y., Gu, S., Li, Y., Timofte, R., Jin, L. and Song, H., 2021. Unsupervised Real-world Image Super Resolution via Domain-distance Aware Training, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13380-13389.
Xiong, Y., Guo, S., Chen, J., Deng, X., Sun, L., Zheng, X., Xu, W., 2020. Improved SRGAN for Remote Sensing Image Super-Resolution Across Locations and Sensors. Remote Sens. 12 (8).
Yang, J., Wright, J., Huang, T.S., Ma, Y., 2010. Image Super-Resolution Via Sparse Representation. IEEE Trans. Image Process. 19 (11), 2861–2873.
Yin, G., Wang, W., Yuan, Z., Ji, W., Yu, D., Sun, S., Chua, T.S., Wang, C., 2022. Conditional Hyper-Network for Blind Super-Resolution With Multiple Degradations. IEEE Trans. Image Process. 31, 3949–3960.
Yue, Z., Zhao, Q., Xie, J., Zhang, L., Meng, D. and Wong, K.Y.K., 2022. Blind Image Super-resolution with Elaborate Degradation Modeling on Noise and Kernel, 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2118-2128.
Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., Zhang, L., 2016. Image super-resolution: The techniques, applications, and future. Signal Process. 128, 389–408.
Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B. and Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks, Proceedings of the European Conference on Computer Vision (ECCV), pp. 286-301.
Zhang, K., Gao, X., Tao, D., Li, X., 2012. Single Image Super-Resolution With Non-Local Means and Steering Kernel Regression. IEEE Trans. Image Process. 21 (11), 4544–4556.

Zhang, Z., Gao, K., Wang, J., Min, L., Ji, S., Ni, C., Chen, D., 2022b. Gradient Enhanced Dual Regression Network: Perception-Preserving Super-Resolution for Multi-Sensor Remote Sensing Imagery. IEEE Geosci. Remote Sens. Lett. 19, 1–5.
Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y., 2021. Residual Dense Network for Image Restoration. IEEE Trans. Pattern Anal. Mach. Intell. 43 (7), 2480–2495.
Zhang, N., Wang, Y., Zhang, X., Xu, D., Wang, X., Ben, G., Zhao, Z., Li, Z., 2022a. A Multi-Degradation Aided Method for Unsupervised Remote Sensing Image Super Resolution With Convolution Neural Networks. IEEE Trans. Geosci. Remote Sens. 60, 1–14.
Zhang, L., Wu, X., 2006. An edge-guided image interpolation algorithm via directional filtering and data fusion. IEEE Trans. Image Process. 15 (8), 2226–2238.
Zheng, X., Huan, L., Xia, G.-S., Gong, J., 2020. Parsing very high resolution urban scene images by learning deep ConvNets with edge-aware loss. ISPRS J. Photogramm. Remote Sens. 170, 15–28.
Zhou, J., Civco, D.L., Silander, J.A., 1998. A wavelet transform method to merge Landsat TM and SPOT panchromatic data. Int. J. Remote Sens. 19 (4), 743–757.
Zhu, Y., Geiß, C., So, E., 2021. Image super-resolution with dense-sampling residual channel-spatial attention networks for multi-temporal remote sensing image classification. Int. J. Appl. Earth Obs. Geoinf. 104, 102543.