



## Article

# Remote Sensing Image Super-Resolution via Multi-Scale Texture Transfer Network

Yu Wang <sup>1,2</sup> , Zhenfeng Shao <sup>1,\*</sup> , Tao Lu <sup>3</sup> , Xiao Huang <sup>4</sup> , Jiaming Wang <sup>3</sup> , Xitong Chen <sup>3</sup> , Haiyan Huang <sup>1</sup> and Xiaolong Zuo <sup>1</sup>

<sup>1</sup> State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

<sup>2</sup> School of General Aviation, Jingchu University of Technology, Jingmen 448000, China

<sup>3</sup> Hubei Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China

<sup>4</sup> Department of Environmental Sciences, Emory University, Atlanta, GA 30322, USA

\* Correspondence: shaozhenfeng@whu.edu.cn

**Abstract:** As the degradation factors of remote sensing images become increasingly complex, it becomes challenging to infer the high-frequency details of remote sensing images compared to ordinary digital photographs. For super-resolution (SR) tasks, existing deep learning-based single remote sensing image SR methods tend to rely on texture information, leading to various limitations. To fill this gap, we propose a remote sensing image SR algorithm based on a multi-scale texture transfer network (MTTN). The proposed MTTN enhances the texture feature information of reconstructed images by adaptively transferring texture information according to the texture similarity of the reference image. The proposed method adopts a multi-scale texture-matching strategy, which promotes the transmission of multi-scale texture information of remote sensing images and obtains finer-texture information from more relevant semantic modules. Experimental results show that the proposed method outperforms state-of-the-art SR techniques on the Kaggle open-source remote sensing dataset from both quantitative and qualitative perspectives.

**Keywords:** remote sensing; multi-scale; texture transfer; super-resolution; deep residual network



**Citation:** Wang, Y.; Shao, Z.; Lu, T.; Huang, X.; Wang, J.; Chen, X.; Huang, H.; Zuo, X. Remote Sensing Image Super-Resolution via Multi-Scale Texture Transfer Network. *Remote Sens.* **2023**, *15*, 5503. <https://doi.org/10.3390/rs15235503>

Academic Editors: Junmin Liu, Tiejong Zeng, Xile Zhao, Bin Zhao and Claudia Paris

Received: 18 October 2023

Revised: 21 November 2023

Accepted: 23 November 2023

Published: 25 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Remote sensing is an important ground observation technology that has developed rapidly in recent years. Given its unique advantages of extensive coverage, regular temporal regularity, and high accessibility, it has been widely used in disaster detection and early alerting [1], resource exploration [2], and land cover classification [3–6]. There are broad application prospects in environmental testing and other fields [7–12]. Due to space imaging equipment and communication bandwidth limitations, the resolution of images captured from satellite sensors is limited. In addition, the image quality is subject to various factors, including atmospheric turbulence, imaging process noise, and motion blur, leading to further reduced resolution and quality. Super-resolution (SR) algorithms can enhance acquired remote sensing images, thus, bypassing hardware limitations.

In the past few decades, SR technology has been widely applied to enhance the quality and resolution of remote sensing images. Existing remote sensing SR methods can be categorized into two main types: methods based on traditional algorithms [13] and methods based on deep learning [14,15]. The traditional learning-based super-resolution methods can learn prior knowledge from a data sample library by constructing learning models, thereby establishing a mapping relationship between low-resolution (LR) and high-resolution (HR) images. This process generates high-frequency details and texture information lost during the downscaling. Zheng et al. [16] first applied sparse representation theory to the field of remote sensing image SR, establishing overcomplete coupled high–low resolution dictionaries and solving sparse coefficients using an orthogonal matching pursuit strategy.

Hou et al. [17] introduced a global joint dictionary model and non-local self-similarity for global and local constraints, establishing the intrinsic connections between image blocks and global information. Zhang et al. [18] proposed a sparse residual dictionary learning method, directly reconstructing residual images with a residual dictionary to alleviate the ringing effects of unstructured pixel-domain high-low resolution dictionaries. Wu et al. [19] proposed a SR framework based on a multi-feature dictionary, constructing multiple LR dictionaries for representation and overlay in the HR space.

With the widespread use of convolutional neural networks (CNN) in SR tasks, the deep learning-based remote sensing SR method has become the mainstream research direction of image SR methods. Deep learning-based methods usually learn complex mapping functions between LR and HR patches through CNN. Considering its outstanding learning ability, Shi et al. [20] built a sub-pixel CNN that builds a mapping mechanism from LR to HR. Kim et al. [21] proposed a novel network method with deep residual learning to ease the training process. Lai et al. [22] designed a pyramid network that fuses multi-scale residuals in the feature domain. Luo et al. [23] achieved excellent results by replacing zero paddings with self-similarity to prevent useless information. Wang et al. [24] proposed an SR method that preserves the temporal correlation between frames. Ledig et al. [25] introduced perceptual loss and proposed an SR method for generative adversarial networks (GAN) to obtain images with more realistic textures. The generation network guides image reconstruction with a perceptual loss so that the reconstruction result is close to the original HR image. The discriminant network compares the reconstructed image to the original HR image, and the reconstructed image is output if it can pass the discriminant network.

Current SR methods primarily target natural images, but they fall short when applied to the diverse shooting angles and multi-scale characteristics present in large-scale remote sensing data. Existing methods lack consideration for these factors, making it crucial to develop an SR method specifically tailored for accurately reconstructing remote sensing images with intricate textures. Drawing inspiration from the effectiveness of multi-angle priors in enriching prior information [26], this paper introduces reference images (Ref) to enhance the prior information of training samples. The algorithm is adapted for the reconstruction of remote sensing images by aligning feature texture information. The proposed network initially matches local texture information between the Ref image and the LR image, then conveys the identified texture features to the reconstruction network via the depth residual module. The final output is high-quality remote sensing images with detailed texture information. Extensive experimental results demonstrate the algorithm's superiority over other state-of-the-art methods, both quantitatively and qualitatively.

We summarize the major contributions of this work as follows:

- (1) We propose a novel remote sensing image RefSR method based on a multi-scale texture transfer network (MTTN), which can effectively address the issue of insufficient texture detail information in the reconstruction process of the single remote sensing image SR method.

- (2) We propose a novel multi-scale texture transfer method. The proposed method uses residual blocks with different convolution kernel sizes to extract texture information of different scales from LR and Ref remote sensing images and take advantage of the texture details provided by Ref images to facilitate the reconstruction of LR images.

- (3) We conduct extensive experiments on the challenging Kaggle open dataset, and the experimental results demonstrate the effectiveness and adaptability of our suggested MTTN method. In addition, the results also point to its great potential in subsequent remote sensing image classification tasks.

The remainder of this article is structured as follows. Section 2 presents related work, including existing SISR of remote sensing images and reference-based SR methods. Section 3 describes the principal method of this model in detail. Section 4 introduces the experimental results. Section 5 concludes this study and Section 6 discusses future work.

## 2. Related Work

Over the past decades, deep learning-based SR methods have demonstrated excellent reconstruction performance in evaluation indicators or visual quality compared to traditional methods. This section introduces the single image super-resolution (SISR) and reference-based super-resolution (RefSR) of remote sensing images.

### 2.1. SISR of Remote Sensing Images

The SISR method for remote sensing images has recently drawn increasing attention due to the unique value of high-quality remote sensing images in application scenarios. In addition to relying on advanced hardware, the SR method offers an efficient way to enhance the quality of remote sensing images. Recently, with the rapid popularity of convolutional neural networks in acquiring high-quality images, many researchers have proposed a large number of deep learning-based SISR methods for remote sensing images. Lei et al. [27] designed a CNN-based network framework to learn the local details and global prior information of remote sensing images. Xu et al. [28] proposed a deep memory connection network that reconstructs high-quality remote sensing images through local and global memory connections, so the reconstruction performance of remote sensing images is effectively improved. Lu et al. [29] employed a novel multi-scale residual neural network method. This method extracts image blocks of different sizes as multi-scale information and then fuses multi-scale high-frequency information to reconstruct images. Jiang et al. [30] designed an edge enhancement-based SR network to enhance the features of remote sensing images using edge detail information extracted by the network. Dong et al. [31] designed a multi-sensory attention method to obtain multi-level information of remote sensing features through the proposed feature extraction module. Haut et al. [32] designed a network to extract more related features through combining residual units and skipping connections. Qin et al. [33] designed a deep gradient-aware network with image-specific enhancement to preserve more gradient details in reconstructed images. Zhang et al. [34] proposed a novel attention module to capture finer remote sensing image texture information through this module. Lei et al. [35] created a new remote sensing image SR framework by introducing a transformer to extract multi-level features and better fuse the features of different dimensions. Liu et al. [36] designed a graph neural network (GNN) based SISR method, which can better consider the ill-conditioned phenomena in remote sensing images.

Despite the great progress in aforementioned SISR methods, most of these methods rely on paired images and local information. However, when confronted with certain specific and complex textures, they fail to capture sufficient prior information. Therefore, the SISR methods that handle remote sensing imagery have been criticized for their limitations in handling multi-source data.

### 2.2. Reference-Based Super-Resolution

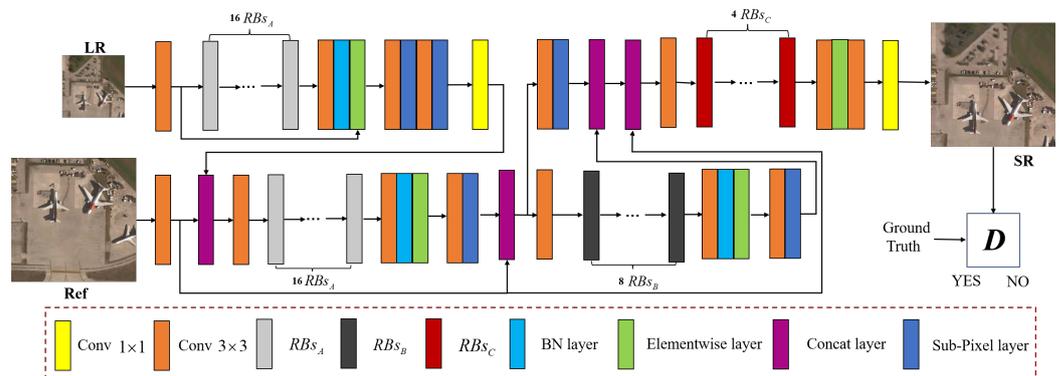
In contrast to the SISR method, the RefSR method achieves the SR reconstruction process by introducing other related images. For the Ref image to effectively aid the SR reconstruction process, texture information similar to the LR image is typically required. Consequently, the Ref image is usually selected from related video frames [37,38] or photos taken from various angles [39]. Recently, SR methods based on Ref images have made great progress in general image methods. Zheng et al. [40] used cross-scale warping to create an end-to-end fully convolutional deep neural network. This network improves the accuracy of reconstructed images by performing end-to-end spatial alignment at the pixel level. Zhang et al. [41] employed a novel RefSR model that restores the target image through texture migration of the Ref image. Xie et al. [42] designed a matching and exchange module for the RefSR task to extract similar texture and high-frequency features from the Ref image by appropriately assigning the gradient to the previous feature encoding module. Yang et al. [43] designed a novel texture transformer model that transmits HR texture from the Ref image through a texture converter and attention module. Huang et al. [44] proposed a novel RefSR framework, which includes a texture extraction module and a texture transfer

module to more fully transfer textures in an adaptive manner. Dong et al. [45] employed a GAN-based RefSR network, mainly using the relevant attention module that improves the network’s reconstruction ability through gradient-assisted alignment. Cai et al. [46] designed a texture transformer for remote sensing image SR algorithm, which reduces the SR model’s dependence on Ref images through U-Transformer.

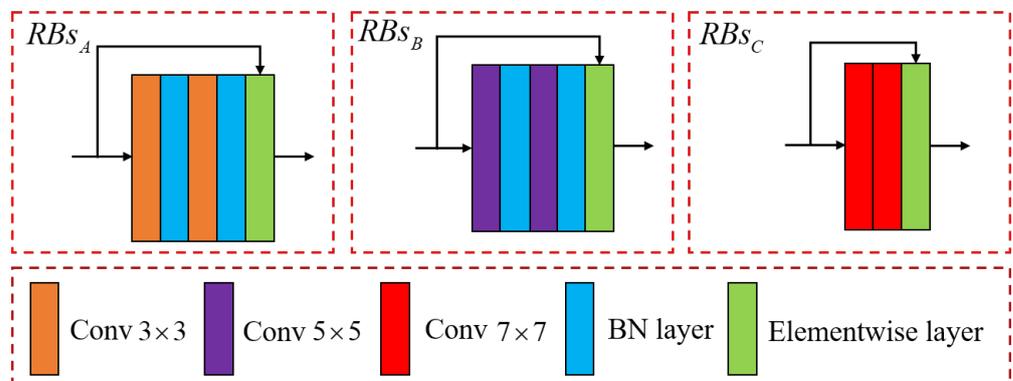
The above-mentioned RefSR method introduces a new perspective for addressing remote sensing SR tasks. While the RefSR method offers more possibilities for remote sensing image super-resolution compared to SISR methods, existing RefSR methods often overlook the spatial diversity features of remote sensing images. These features contain valuable information that can significantly benefit remote sensing SR tasks. To address this gap, we have developed an innovative multi-scale texture transfer model to enhance the performance and robustness of RefSR. By extracting texture information from reference remote sensing images, this model further aids in extracting multi-scale information from the original low-resolution images, thereby reconstructing more refined remote sensing images.

### 3. Method

The main idea of the proposed method is to reconstruct SR images  $I^{SR}$  from LR remote sensing images  $I^{LR}$  and similar Ref images  $I^{Ref}$ , and then synthesize a reasonable texture subject to  $I^{Ref}$ , while maintaining consistency with the  $I^{LR}$  content. The proposed model’s overall network structure is shown in Figure 1. The multi-scale texture transfer method takes into account the semantic and textural similarity of  $I^{LR}$  and  $I^{Ref}$  images and can better suppress irrelevant textures while transferring related remote sensing image texture features.



**Figure 1.** Framework of the proposed multi-scale texture transfer network (MTTN). The specific structure of  $RBs_A$ ,  $RBs_B$ , and  $RBs_C$  are shown in Figure 2.



**Figure 2.** Illustration of the residual blocks (RBs) in the proposed MTTN model. The  $RBs_A$ ,  $RBs_B$ , and  $RBs_C$  represent RBs with different convolution kernel sizes.

### 3.1. Feature Swapping

The first step in our MTTN is to swap remote sensing image features. The design algorithm searches the entire remote sensing reference image for locally similar multi-scale textures, and these extracted multi-scale textures can be used to exchange the texture features of  $I^{LR}$  with enhanced remote sensing SR image reconstruction. The multi-scale texture is then transferred directly to the final output  $I^{SR}$  via feature search in HR space coordinates. Specifically, we first extract the corresponding feature map on the  $I^{LR}$ , and then sample the extracted feature map twice to obtain an enlarged LR image  $I^{LR\uparrow}$  with the same resolution as the  $I^{HR}$ . We also apply the same upscaling factor for  $I^{Ref}$  before performing bicubic interpolation sampling to obtain a Ref image block  $I^{Ref\downarrow\uparrow}$  matching the up-sampled  $I^{LR\uparrow}$ . As the LR and Ref images' data sources may differ, we compute texture similarity in the high-resolution feature space  $\phi(I)$  to emphasize structure and texture information, and we apply the inner product method to compute feature similarity between image pairs:

$$h_{i,j} = \left\langle K_i(\phi(I^{LR\uparrow})), \frac{K_j(\phi(I^{Ref\downarrow\uparrow}))}{\|K_j(\phi(I^{Ref\downarrow\uparrow}))\|} \right\rangle, \quad (1)$$

where  $K_i()$  denotes the  $i$ -th patch obtained from the high-resolution space,  $\| \cdot \|$  denotes feature normalization. The Ref feature block performs the best matching for the reference image features through a normalization function. Then we calculate the similarity between image pairs via convolution operations on the LR patch. The kernel operation expression of the Ref patch follows:

$$N_j = \phi(I^{LR\uparrow}) * \frac{K_j(\phi(I^{Ref\downarrow\uparrow}))}{\|K_j(\phi(I^{Ref\downarrow\uparrow}))\|}, \quad (2)$$

where  $N_j$  represents the Ref patch's similarity graph.  $N_j(x, y)$  represents the texture similarity between LR and Ref. We represent LR texture patches based on similarity scores by creating swapped feature maps. The image patches in the feature map are represented as:

$$K_{\omega(x,y)} = K_{j^*}(\phi(I^{Ref})), j^* = \arg \max_j N_j(x, y), \quad (3)$$

where  $\omega()$  corresponds to the patch center and the patch index. Note that while  $I^{Ref\downarrow\uparrow}$  is used for matching (Equation (2)), the original Ref  $I^{Ref}$  is used for exchange (Equation (3)) to retain the high-frequency feature information. The final generated swap feature map is then fed to the subsequent multi-scale texture transfer by calculating the average between the swap feature maps.

### 3.2. Multi-Scale Texture Transfer

Our multi-scale texture transfer method merges the transferred texture data into the generation network of various feature layers corresponding to various scales after the layer-by-layer transfer of the feature maps extracted from the feature extraction network and high-resolution reference images. The above-described procedure builds the exchanging feature map  $M_l$  for multi-scale layer  $l$ , where the texture encoder  $\psi_l$  corresponds to the scale being used.

To build the fundamental generative network, we use residual blocks (RBs) with various convolution kernel sizes. The output  $\phi_l$  of layer  $l$  can be recursively expressed as:

$$\psi_l = [\text{Res}(\psi_{l-1} \nabla M_{l-1}) + \psi_{l-1}] \uparrow_{2\times}, \quad (4)$$

where  $\text{Res}()$  represents the RBs,  $\nabla$  represents the channel cascade, and  $\uparrow_{2\times}$  represents the use of sub-pixel convolution layer for  $2\times$  magnification. The remote sensing reconstruction image is generated once the  $L$  layer has reached the desired HR resolution:

$$I^{SR} = \text{Res}(\psi_{l-1} \nabla M_{l-1}) + \psi_{l-1}. \quad (5)$$

Unlike the traditional SISR method that only considers the texture difference between  $I^{SR}$  and  $I^{HR}$ , the proposed MTTN network considers the texture variation between  $I^{SR}$  and  $I^{Ref}$ . In other words, the texture of  $I^{SR}$  must be similar to the swap feature map  $M_l$ . Therefore, we define the texture loss  $L_{tex}$  as:

$$L_{tex} = \sum_l \lambda_l \|G_r(\phi_l(I^{SR}) \times N_l^*) - G_r(M_l \times N_l^*)\|_F, \quad (6)$$

where  $G_r()$  calculates the gamma matrices, and  $\lambda_l$  is the normalized exponent of the  $l$  layer.  $N_l^*$  represents the weight map of the input LR remote sensing image patch. The difference between this weight map and  $I^{LR}$  is the fact that texture owns a lower weight, leading to lower texture loss. Thus, the texture transferred from  $I^{Ref}$  to  $I^{SR}$  can be enforced adaptively based on the image quality of Ref image, resulting in more robust texture features.

### 3.3. Loss Function

To improve the visual quality of remote sensing reconstructed images, we design four loss functions. In addition to introducing texture loss  $L_{tex}$ , our algorithm also introduces reconstruction loss  $L_{rec}$ , perceptual loss  $L_{per}$ , and adversarial loss  $L_{adv}$ . The  $L_{rec}$  is a commonly used loss function in SR, and the  $L_{per}$  and the  $L_{adv}$  are the losses introduced by the generation confrontation network in order to reconstruct a more realistic image. We describe these loss functions below.

#### 3.3.1. Reconstruction Loss

The  $L_{rec}$  is typically defined by the mean square error. The function aims to achieve a higher peak signal-to-noise ratio (PSNR), and the L1-norm is adopted:

$$L_{rec} = \|I^{HR} - I^{SR}\|_1, \quad (7)$$

#### 3.3.2. Perceptual Loss

Many SR networks have introduced perceptual loss to achieve more realistic visual quality. Therefore, our network also uses the perceptual loss function ( $L_{per}$ ):

$$L_{per} = \frac{1}{V} \sum_{i=1}^C \|\phi_i(I^{HR}) - \phi_i(I^{SR})\|_F, \quad (8)$$

where  $V$  and  $C$  denote the volume and channel of the feature map, respectively.  $\phi_i$  indicates the  $i$ -th channel of the output feature map, and  $\|\cdot\|_F$  represents the Frobenius norm.

#### 3.3.3. Adversarial Loss

The  $L_{adv}$  function can improve the sharpness of SR reconstructed images. Our network uses the improved  $L_{adv}$  function in WGAN [26]:

$$L_{adv} = -\mathbb{E}_{I^{HR}} \left[ \log \left( D \left( I^{HR}, I^{SR} \right) \right) \right] - \mathbb{E}_{I^{SR}} \left[ \log \left( 1 - D \left( I^{SR}, I^{HR} \right) \right) \right], \quad (9)$$

where  $D$  is a relativistic average discriminator [47].

## 4. Experiments

This section mainly introduces the experiments of the proposed MTTN network, including the dataset in Section 4.1, the evaluation indicators in Section 4.2, the implementation details of the experiments in Section 4.3, the quantitative and qualitative analyses with different methods in Section 4.4, and the ablation studies in Section 4.5.

#### 4.1. Dataset

In this study, we use the Kaggle open-source remote sensing competition dataset (<https://www.kaggle.com/c/draper-satellite-image-chronology/data>). For the RefSR-based remote sensing image problem, the texture similarity between input remote sensing image pairs poses a significant impact on the SR results. In general, for training and evaluating RefSR algorithms, Ref images with various similarities to LR images are preferred. However, the Kaggle open-source dataset contains a wide variety of numbers and subject matter, including 324 different scenarios. Therefore, we constructed a dataset of reference images with various similarities based on the Kaggle dataset. We use the SIFT [48] feature-matching method to collect image pairs with different similarities. This feature-matching method collects Ref remote sensing images that are consistent with local texture-matching targets. We also segment these images with a size of  $320 \times 320$ , leading to a total of 1200 remote sensing images, of which 600 are used as input images, and the other 600 are used as reference images for training. Note that the Ref image fed to the algorithm is the same scene as the input image but with different angles. Figure 3 presents some selected samples of our dataset.



**Figure 3.** Selected samples of RefSR remote sensing datasets. (High-resolution image (**left**) and Reference image (**right**)).

#### 4.2. Evaluation Indicators

In this experiment, we use five popular objective evaluation metrics, including peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [49], feature similarity (FSIM) [50], visual information fidelity (VIF) [51], and relative dimensionless global error in synthesis (ERGAS) [52].

Peak Signal-to-Noise Ratio (PSNR), with decibel (dB) as unit, evaluates the similarity between two images from the pixel similarity level. It is one of the most commonly used objective measures to evaluate image quality:

$$PSNR = 10 \log_{10} \frac{(2^N - 1)^2}{MSE} \quad (10)$$

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W [X(i, j) - Y(i, j)]^2 \quad (11)$$

where  $MSE$  is the mean square error between the original image  $X$  and the reconstructed image  $Y$ ,  $N$  is the number of bits of image pixels (most general optical images have 8 bits). The image's width and height are denoted by  $W$  and  $H$ , respectively. The PSNR ranges from  $[0, +\infty]$ , where a higher value indicates more similarity between the images.

Structural Similarity (SSIM) is an objective evaluation index designed based on the human visual system. By decomposing the image, it comprehensively evaluates the degree of similarity between two images based on image brightness, contrast, and structure. For images  $I_x$  and  $I_y$ , SSIM can be calculated as:

$$SSIM(I_x, I_y) = \frac{(2\mu_{I_x}\mu_{I_y} + C_1)(2\sigma_{I_x I_y} + C_2)}{(\mu_{I_x}^2 + \mu_{I_y}^2 + C_1)(\sigma_{I_x}^2 + \sigma_{I_y}^2 + C_2)} \quad (12)$$

where  $\mu_{I_x}$  and  $\mu_{I_y}$  are the mean and luminance components between image pairs, respectively, and  $\sigma_{I_x}^2$  and  $\sigma_{I_y}^2$  are the variance and contrast components of the two images, respectively.  $\sigma_{I_x I_y}$  is the covariance of images  $I_x$  and  $I_y$ , used for structural contrast components.  $C_1$ ,  $C_2$ , and  $C_3$  are the balance parameters. The value range of SSIM is  $[0, 1]$ . The larger the value, the greater the degree of image similarity.

Feature Similarity Index Measure (FSIM) uses low-level image features to replace image statistical features in SSIM and evaluates the feature similarity between two images from three feature levels of image phase consistency and gradient magnitude. For images  $I_x$  and  $I_y$ , FSIM can be calculated as follows:

$$FSIM(I_x, I_y) = \frac{\sum_{(I_x, I_y) \in \Omega} S_L(I_x, I_y) \cdot PC_m(I_x, I_y)}{\sum_{(I_x, I_y) \in \Omega} PC_m(I_x, I_y)} \quad (13)$$

where  $\Omega$  represents the pixel domain of the whole image,  $PC_m(\cdot)$  and  $S_L(\cdot)$  represent the phase consistency and fusion similarity of the two images, respectively. The value range of FSIM is  $[0, 1]$ . The higher the value, the more similar the images are.

Visual Information Fidelity (VIF) is based on statistical models of natural scenes, image distortions, and the human visual system. For images  $I_x$  and  $I_y$ , VIF can be calculated as:

$$VIF(I_x, I_y) = \frac{\sum I(\bar{C}^N; \bar{F}^N | S^N)}{\sum I(\bar{C}^N; \bar{E}^N | S^N)} \quad (14)$$

where  $I(\bar{C}^N; \bar{F}^N | S^N)$  and  $I(\bar{C}^N; \bar{E}^N | S^N)$  represent the information extracted from specific subbands of the input image.  $N$  is the number of sub-bands in the wavelet domain,  $C$  is the set of random vector fields,  $E$  is the random coefficient vector of the wavelet sub-band of the original image,  $F$  is the random coefficient vector of the wavelet sub-band of the distorted image, and  $S$  is an independent scalar factor.

The Erreur Relative Globale Adimensionnelle de Synthèse (ERGAS) provides a global fusion evaluation. The smaller the value (the closer to zero), the better the performance. ERGAS is defined as:

$$ERGAS = 100 \frac{h}{l} \sqrt{\frac{1}{p} \sum_{i=1}^L \left( \frac{RMSE^2(i)}{MEAN(i)} \right)} \quad (15)$$

where  $h$  and  $l$  represent the spatial resolution of the panchromatic image and the multispectral image, respectively,  $p$  represents the number of bands of the fusion map,  $MEAN(i)$  is the  $i$ -band of the mean Ref multispectral image, and  $RMSE(i)$  is the Ref image and the fusion map MSE.

### 4.3. Experimental Details

In this study, we adopt VGG19 [53] as the pre-training model of the network and design an end-to-end remote sensing SR network. We first downsample the original remote sensing images to LR images of  $40 \times 40$  pixels, and then perform data augmentation by random horizontal flipping and 90-degree rotation during training, and adopt a learning rate decay strategy during training. The batch size of the network is set to 9, and the learning rate is initially set to  $1 \times 10^{-4}$  with the reduction rate of each epoch of 0.9. We use the ADAM optimizer [54] for network optimization, and the momentum parameter is set to  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ .

For training and testing, we use a server with NVIDIA 1080Ti GPU and Intel I7-8700 CPU. Our MTTN network is implemented based on the TensorFlow framework under Ubuntu 16.04, CUDA 8.0 and CUDNN 5.1 systems. In our experiments, it takes about 20 h to train our model with a 1080Ti GPU under the above settings for 50 epochs.

### 4.4. Quantitative and Qualitative Comparison with Different Methods

To demonstrate the superiority of MTTN, we compare it to recent SISR and RefSR remote sensing imagery algorithms. The SR methods to be compared include Bicubic, SRCNN [55], VDSR [21], SRResnet [25], SRFBN [56], SRRFN [57], SEAN [58], HAN [59], MHAN [34], SRNTT [41], DLGNN [36], and HSENet [60]. Among them, DLGNN [36] and HSENet [60] are the latest remote sensing SR algorithms that demonstrate great reconstruction results. The competing algorithm also includes a RefSR method, i.e., SRNTT [41], a classic RefSR method.

All of the competing methods were trained and tested on the Kaggle public dataset that was recreated in this study to ensure a fair comparison. Table 1 displays the objective evaluation indicators for all the algorithms, and the best outcomes are highlighted in red. It can be seen from Table 1 that the proposed MTTN shows great performance on the five objective evaluation indicators, i.e., PSNR, SSIM, FSIM, VIF, and ERGAS.

The above results prove that certain SR models designed for natural images are not suitable for remote sensing images. The multi-scale texture transfer network we designed is more effective in learning the texture and structure of remote sensing images, as it uses texture information from Ref remote sensing images to assist in reconstructing texture detail information in LR images.

Figures 4 and 5 presents a visual comparison of the experimental results and further the advantage of the proposed MTTN network. We selected typical scenes, including playgrounds, parking lots, ports, airports, buildings and roads, to evaluate the superiority of the MTTN network. From the zoomed-in details in Figures 4 and 5, we notice the CNN-based approach can reconstruct some texture details. However, as some models are directly transferred to remote sensing SR tasks, the reconstructed remote sensing images contain notable blurred contours. The GAN-based method presents better visual effects, but the generated artifacts deteriorate the reconstructed remote sensing image. In comparison, our proposed MTTN algorithm produces sharper edges and finer details than other competing algorithms. We also observe that the proposed MTTN further enhances the visual performance of the reconstructed image by combining the multi-scale residual information of Ref images.

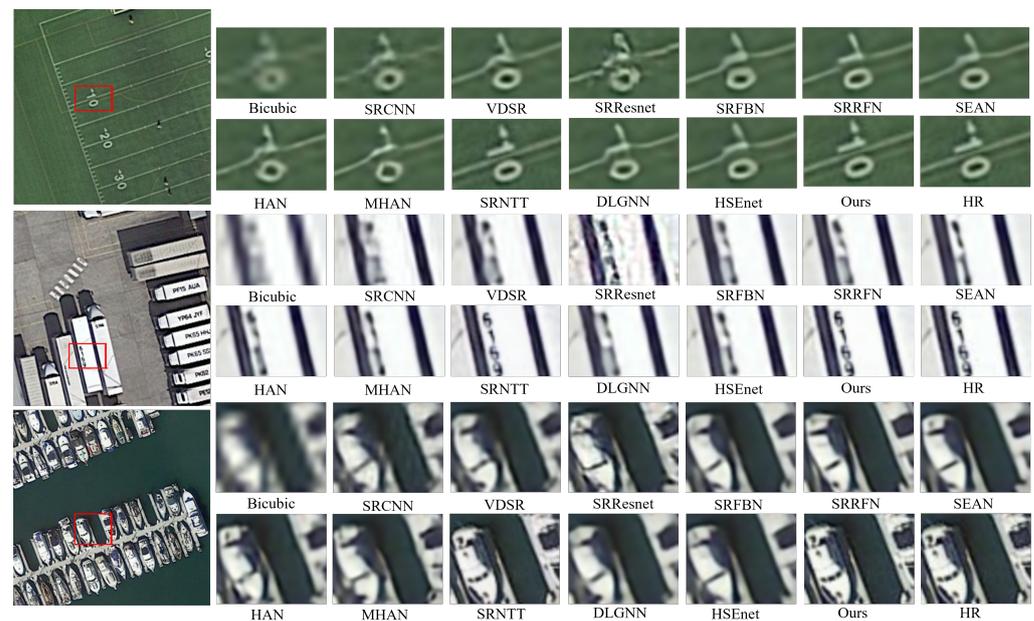
### 4.5. Ablation Studies

#### 4.5.1. Effectiveness of Reference Image

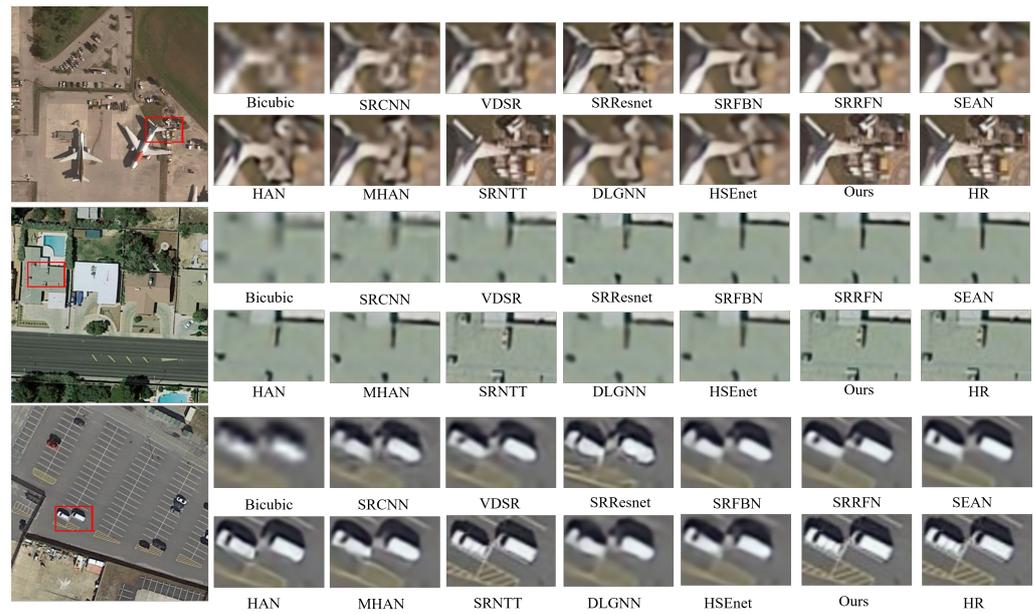
To verify the effectiveness of Ref images on MTTN networks, we set up ablation experiments with different configurations. Table 2 lists the experimental results of various ablation experiments. The SISR model performs worse than the original network by 0.17 dB in terms of PSNR. However, our model still achieves comparable performance to some SISR algorithms without introducing Ref into remote sensing images. The MSE comparison chart in Figure 6 shows that after adding Ref remote sensing images, the reconstructed remote sensing images present richer texture details, indicating that SR performance can be enhanced when incorporating Ref images.

**Table 1.** Comparison of average PSNR, SSIM, FSIM, VIF, and ERGAS results obtained from the Kaggle Open Source dataset.  $\uparrow$  means that the larger the value, the better the reconstruction effect.  $\downarrow$  means that the smaller the value, the better the reconstruction effect.

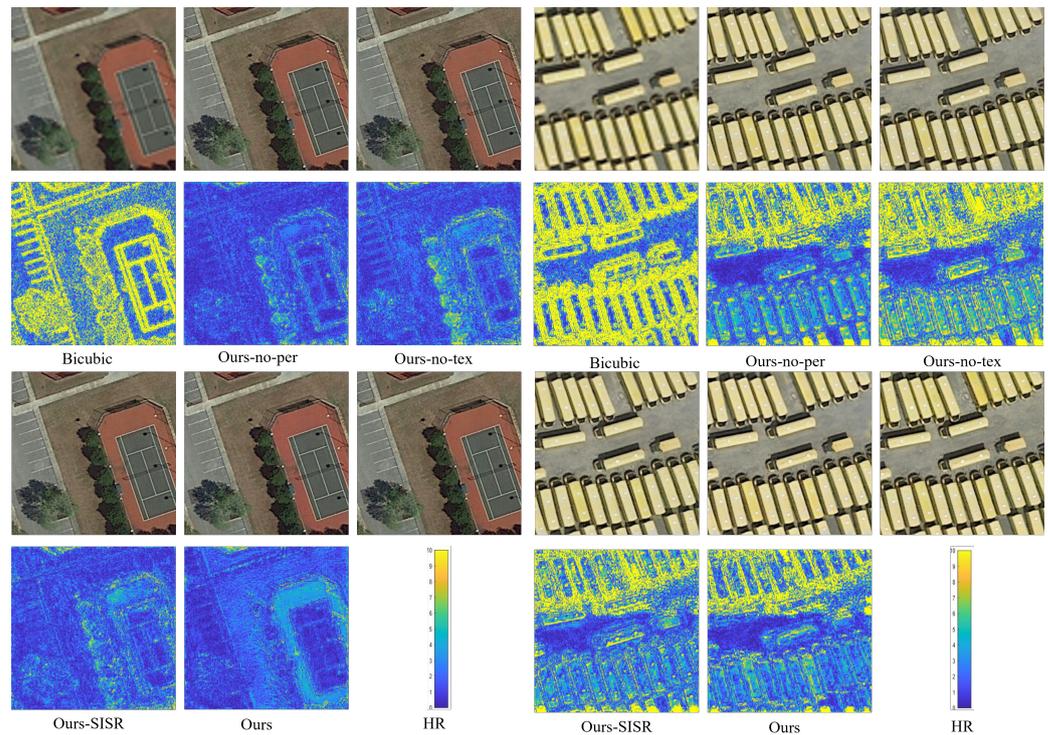
Algorithm/Metrics	PSNR/dB $\uparrow$	SSIM $\uparrow$	FSIM $\uparrow$	VIF $\uparrow$	ERGAS $\downarrow$
Bicubic	24.73	0.6972	0.7544	0.2558	2.7915
SRCNN	26.56	0.7675	0.8292	0.3368	2.2890
VDSR	27.20	0.8001	0.8461	0.3759	2.1461
SRResnet	26.42	0.7790	0.8480	0.3634	2.3560
SRFBN	28.02	0.8264	0.8617	0.4121	1.9713
SRRFN	28.61	0.8415	0.8794	0.4350	1.8559
SEAN	28.41	0.8374	0.8716	0.4261	1.8852
HAN	28.02	0.8259	0.8743	0.4157	1.9714
MHAN	27.86	0.8265	0.8687	0.4106	2.0071
SRNTT	30.28	0.8983	0.9243	0.6207	1.5731
DLGNN	28.04	0.8246	0.8609	0.4106	1.9688
HSEnet	28.42	0.8357	0.8746	0.4275	1.8907
MTTN (Ours)	30.48	0.9020	0.9333	0.6352	1.5355



**Figure 4.** Subjective comparison of our method with other algorithms on the Kaggle open dataset (playgrounds, parking lots and ports). From left to right, top to bottom, they are the result of Bicubic, SRCNN [55], VDSR [21], SRResnet [25], SRFBN [56], SRRFN [57], SEAN [58]), HAN [59], MHAN [34], SRNTT [41], DLGNN [36], and HSEnet [60], proposed methods and HR. Best view via zoomed-in view.



**Figure 5.** Subjective comparison of our method with other algorithms on the Kaggle open dataset (airports, buildings and roads). From left to right, top to bottom, they are the result of Bicubic, SRCNN [55], VDSR [21], SRResnet [25], SRFBN [56], SRRFN [57], SEAN [58]), HAN [59], MHAN [34], SRNTT [41], DLGNN [36], and HSEnet [60], proposed methods and HR. Best view via zoomed-in view.



**Figure 6.** The MSE maps display the disparities between reconstructions and ground truths in various ablation experiments. The color of the MSE map corresponds to the magnitude of the error, which varies as the images' dissimilarity increases or decreases.

**Table 2.** Comparative results of different modes of ablation experiments on the Kaggle Open Source dataset with a scale factor of 4.

Algorithm/Metrics	PSNR/dB $\uparrow$	SSIM $\uparrow$	FSIM $\uparrow$	VIF $\uparrow$	ERGAS $\downarrow$
MTTN without Ref images	30.31	0.8977	0.9319	0.6319	1.5866
MTTN without texture loss	30.10	0.8958	0.9314	0.6259	1.6130
MTTN without perceptual loss	29.73	0.8919	0.9268	0.6311	1.7056
MTTN (Ours)	30.48	0.9020	0.9333	0.6352	1.5355

#### 4.5.2. Effectiveness of Loss Function

Our proposed MTTN model uses a novel texture loss, making it distinct from most SR methods. Different from other texture transfer methods, the MTTN network transfers the most relevant texture of the reference image to the target image. Specifically, the local matching method maintains the texture consistency of remote sensing images; the adaptive neural migration improves texture transfer; and the spatial regularization achieves better global spatial consistency. The test results for texture loss are shown in Table 2. The PSNRs of MTTN with and without texture loss tested on the Kaggle dataset are 30.48 dB and 30.10 dB, respectively. It proves that the texture information from Ref cannot be effectively transferred to the target image without texture loss.

We also incorporate perceptual loss into this network to enhance the visual appeal of reconstructed remote sensing images. To verify the effectiveness of this loss, we conducted relevant ablation experiments. The experimental results are shown in Table 2. Regarding the test PSNR results on the Kaggle dataset, the PSNR values after introducing perceptual loss are significantly higher than the test results without perceptual loss. From Figure 6, we notice that the two loss functions in our proposed MTTN effectively reduce the disparity between SR images and the corresponding ground-truth images.

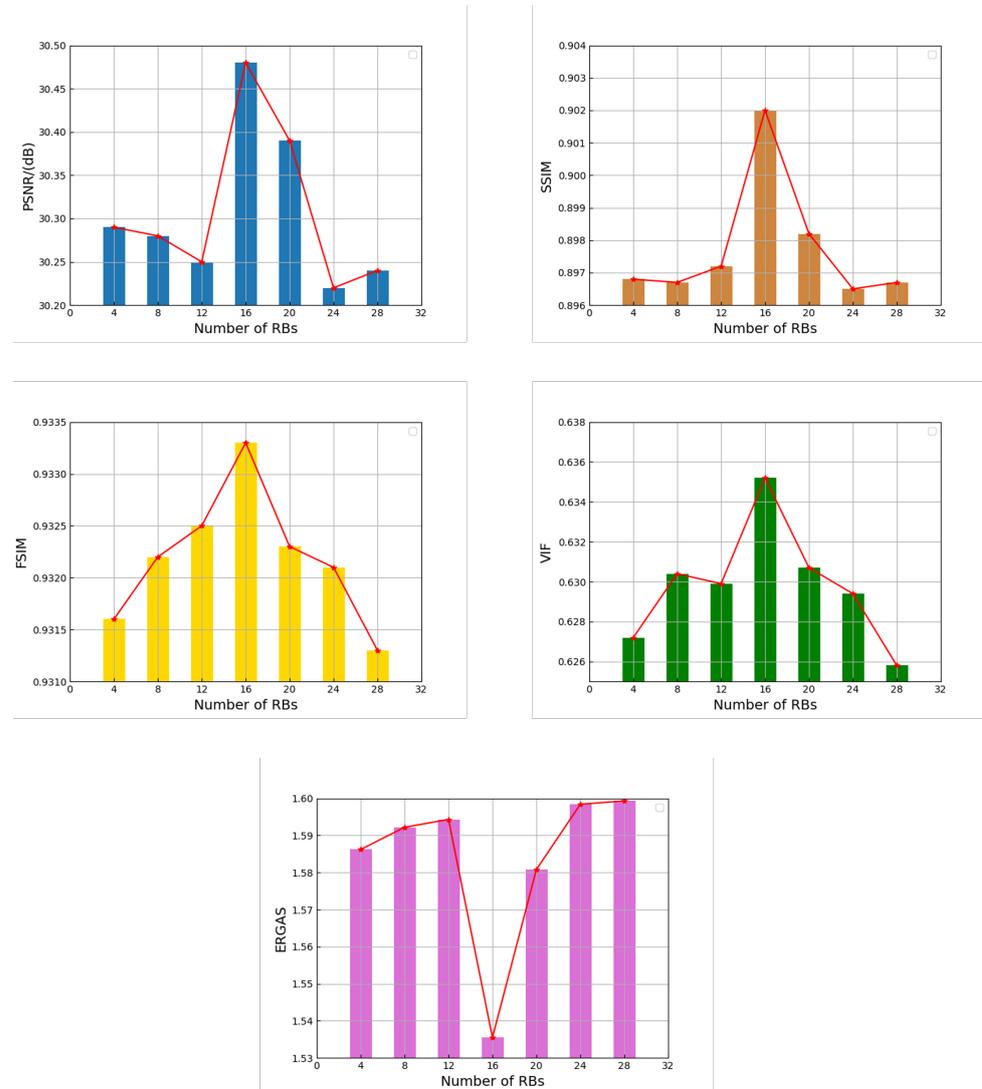
#### 4.5.3. Effectiveness of Residual Blocks

In this study, we extract multi-scale texture information from remote sensing images using different numbers of residual blocks (RBs). Similar to other works [61,62], in order to make full use of the multi-scale texture information of remote sensing images, our different residual blocks adopt a combination of different convolution kernel sizes, and the specific structure is shown in Figure 2. According to the experience sharing of other work [41], we set the number of RBs at different scales to different numbers ( $3 \times 3$  convolution uses  $16 RBs_A$ ,  $5 \times 5$  convolution uses  $8 RBs_B$  and  $7 \times 7$  convolution uses  $4 RBs_C$ ), which proves that the MTTN network can ensure that the parameter quantity of the network does not increase additionally, and at the same time, it also guarantees a good reconstruction ability of the network. To verify the effects of different numbers of design methods of  $RBs_A$  on the performance of MTTN, we designed a set of ablation experiments. As shown in Figure 7, the reconstruction performance of our network decreases as the number of residual blocks increases. Therefore, considering the network's parameters and the reconstruction performance of remote sensing images, the MTTN network adopts the above settings to obtain better reconstruction performance when transmitting texture information.

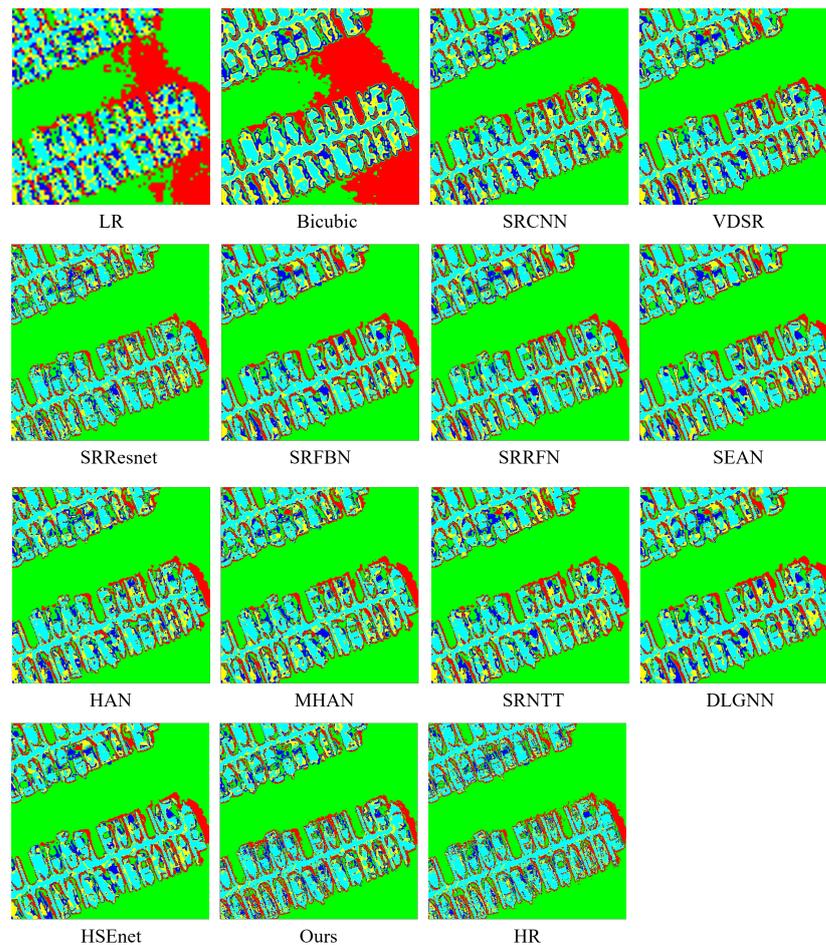
#### 4.5.4. Effectiveness of Remote Sensing Scene Classification

The SR algorithm for remote sensing images is considered a pre-processing step for some downstream tasks, e.g., change detection and semantic segmentation. To further support the efficacy of the suggested MTTN, we evaluate the reconstruction outcomes of various networks using an unsupervised semantic segmentation algorithm (ISODATA). We set the number of categories to five and the maximum number of iterations to five [63]. It is particularly challenging for ISODATA to obtain precise classification results because the results of the comparison experiments show significant spectral distortions and fuzzy patterns. The comparison of the experimental results of the classification algorithm is

shown in Figure 8. In particular, we notice the misclassification of ships and the lack of texture information from the results of DLGNN and HSEnet, while SRNTT fails to recover detailed ship edge information. The results demonstrate that the reconstructed images by the MTTN network contain fine details, evidenced by the ISODATA classification results.



**Figure 7.** Performance comparison of MTTN with the number of  $RBs_A$ .



**Figure 8.** The classification outcomes achieved by various algorithms using the ISODATA classification method.

## 5. Conclusions

In this study, we propose MTTN, a novel RefSR method for remote sensing images with multi-scale residual transfer. Specifically, the proposed MTTN extends the prior information of training samples by introducing additional texture information from reference images. In addition, the use of multi-scale methods to transfer texture information also enriches the prior information required by the network at different stages. The MTTN network introduces multi-scale residual blocks to enhance the image edge and texture information. We perform a comprehensive evaluation of the experimental results in the Kaggle remote sensing competition dataset. Extensive experimental results demonstrate that MTTN can enhance the texture details and visual effects of reconstructed images. Quantitatively, the MTTN network outperforms other competing algorithms in all selected objective evaluation metrics. The self-reference ablation experiments demonstrate that MTTN achieves outstanding reconstructed results with and without reference images.

## 6. Discussion

This paper has conducted relevant research on remote sensing super-resolution algorithms based on reference images, laying a certain foundation for future related work. However, in the practical acquisition scenarios of remote sensing images, the process of obtaining low-resolution remote sensing images is complex, and there are factors such as noise, downsampling, and blur, making existing methods unsuitable for direct application in the super-resolution reconstruction of remote sensing images in real-world scenarios. Therefore, future research directions may consider investigating remote sensing image

super-resolution reconstruction in complex real-world scenarios, aiming to enhance the model's generalization capability and applicability.

**Author Contributions:** Conceptualization, Y.W., Z.S. and T.L.; methodology, Y.W., T.L. and J.W.; software, Y.W. and Z.S.; validation, Y.W., J.W. and X.C.; formal analysis, Y.W., H.H. and X.Z.; investigation, Y.W., J.W. and X.C.; resources, Y.W., Z.S. and T.L.; writing—original draft preparation, Y.W. and J.W.; writing—review and editing, Y.W. and X.H.; visualization, H.H. and X.Z.; supervision, Z.S. and T.L.; project administration, T.L.; funding acquisition, Z.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This article is supported in part by the National Natural Science Foundation of China (42090012), Guangxi Science and Technology Plan Project (Guike 2021AB30019), Hubei Province Key R&D Project (2022BAA048), Sichuan Province Key R&D Project (2022YFN0031, 2023YFN0022, 2023YFS0381), Zhuhai Industry-University-Research Cooperation Project (ZH22017001210098PWC), Shanxi Provincial Science and Technology Major Special Project (202201150401020), Guangxi Key Laboratory of Spatial Information and Surveying and Mapping Fund Project (21-238-21-01), Opening Fund of Hubei Key Laboratory of Intelligent Robot under Grant (HBIR202103), and Supported by Hubei Provincial Natural Science Foundation of China (2023AFB158).

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author, upon reasonable request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Bredemeyer, S.; Ulmer, F.G.; Hansteen, T.H.; Walter, T.R. Radar path delay effects in volcanic gas plumes: The case of Láscar Volcano, Northern Chile. *Remote Sens.* **2018**, *10*, 1514. [[CrossRef](#)]
- Li, C.; Ma, Y.; Mei, X.; Liu, C.; Ma, J. Hyperspectral unmixing with robust collaborative sparse regression. *Remote Sens.* **2016**, *8*, 588. [[CrossRef](#)]
- Jiang, J.; Ma, J.; Chen, C.; Wang, Z.; Cai, Z.; Wang, L. SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4581–4593. [[CrossRef](#)]
- He, N.; Fang, L.; Li, S.; Plaza, A.; Plaza, J. Remote sensing scene classification using multilayer stacked covariance pooling. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6899–6910. [[CrossRef](#)]
- Fang, L.; Liu, G.; Li, S.; Ghamisi, P.; Benediktsson, J.A. Hyperspectral image classification with squeeze multibias network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1291–1301. [[CrossRef](#)]
- Li, R.; Zheng, S.; Duan, C.; Wang, L.; Zhang, C. Land cover classification from remote sensing images based on multi-scale fully convolutional network. *Geo-Spat. Inf. Sci.* **2022**, *25*, 278–294. [[CrossRef](#)]
- Zhang, S.; Shao, Z.; Huang, X.; Bai, L.; Wang, J. An internal-external optimized convolutional neural network for arbitrary orientated object detection from optical remote sensing images. *Geo-Spat. Inf. Sci.* **2021**, *24*, 654–665. [[CrossRef](#)]
- Shao, Z.; Wu, W.; Li, D. Spatio-temporal-spectral observation model for urban remote sensing. *Geo-Spat. Inf. Sci.* **2021**, *24*, 372–386. [[CrossRef](#)]
- Liu, J.; Xiang, J.; Jin, Y.; Liu, R.; Yan, J.; Wang, L. Boost Precision Agriculture with Unmanned Aerial Vehicle Remote Sensing and Edge Intelligence: A Survey. *Remote Sens.* **2021**, *13*, 4387. [[CrossRef](#)]
- Ren, J.; Wang, R.; Liu, G.; Wang, Y.; Wu, W. An SVM-Based Nested Sliding Window Approach for Spectral–Spatial Classification of Hyperspectral Images. *Remote Sens.* **2021**, *13*, 114. [[CrossRef](#)]
- Bai, T.; Wang, L.; Yin, D.; Sun, K.; Chen, Y.; Li, W.; Li, D. Deep learning for change detection in remote sensing: A review. *Geo-Spat. Inf. Sci.* **2022**, *26*, 262–288. [[CrossRef](#)]
- Yu, X.; Pan, J.; Wang, M.; Xu, J. A curvature-driven cloud removal method for remote sensing images. *Geo-Spat. Inf. Sci.* **2023**, 1–22. [[CrossRef](#)]
- Li, X.; Hu, Y.; Gao, X.; Tao, D.; Ning, B. A multi-frame image super-resolution method. *Signal Process.* **2010**, *90*, 405–414. [[CrossRef](#)]
- Wang, J.; Shao, Z.; Huang, X.; Lu, T.; Zhang, R.; Ma, J. Enhanced image prior for unsupervised remote sensing super-resolution. *Neural Netw.* **2021**, *143*, 400–412. [[CrossRef](#)]
- Wang, Y.; Shao, Z.; Lu, T.; Liu, L.; Huang, X.; Wang, J.; Jiang, K.; Zeng, K. A lightweight distillation CNN-transformer architecture for remote sensing image super-resolution. *Int. J. Digit. Earth* **2023**, *16*, 3560–3579. [[CrossRef](#)]
- Zhihui, Z.; Bo, W.; Kang, S. Single remote sensing image super-resolution and denoising via sparse representation. In Proceedings of the 2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping, Xiamen, China, 10–12 January 2011; pp. 1–5.
- Hou, B.; Zhou, K.; Jiao, L. Adaptive super-resolution for remote sensing images based on sparse representation with global joint dictionary model. *IEEE Trans. Geosci. Remote Sens.* **2017**, *56*, 2312–2327. [[CrossRef](#)]

18. Zhang, Y.; Wu, W.; Dai, Y.; Yang, X.; Yan, B.; Lu, W. Remote sensing images super-resolution based on sparse dictionaries and residual dictionaries. In Proceedings of the 2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing, Chengdu, China, 21–22 December 2013; pp. 318–323.
19. Wu, W.; Yang, X.; Liu, K.; Liu, Y.; Yan, B. A new framework for remote sensing image super-resolution: Sparse representation-based method by processing dictionaries with multi-type features. *J. Syst. Archit.* **2016**, *64*, 63–75. [[CrossRef](#)]
20. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
21. Kim, J.; Kwon Lee, J.; Mu Lee, K. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
22. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.
23. Luo, Y.; Zhou, L.; Wang, S.; Wang, Z. Video satellite imagery super resolution via convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2398–2402. [[CrossRef](#)]
24. Wang, Z.; Yi, P.; Jiang, K.; Jiang, J.; Han, Z.; Lu, T.; Ma, J. Multi-memory convolutional neural network for video super-resolution. *IEEE Trans. Image Process.* **2018**, *28*, 2530–2544. [[CrossRef](#)]
25. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
26. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5767–5777.
27. Lei, S.; Shi, Z.; Zou, Z. Super-resolution for remote sensing images via local–global combined network. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1243–1247. [[CrossRef](#)]
28. Xu, W.; Guangluan, X.; Wang, Y.; Sun, X.; Lin, D.; Yirong, W. High quality remote sensing image super-resolution using deep memory connected network. In Proceedings of the IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 8889–8892.
29. Lu, T.; Wang, J.; Zhang, Y.; Wang, Z.; Jiang, J. Satellite image super-resolution via multi-scale residual deep neural network. *Remote Sens.* **2019**, *11*, 1588. [[CrossRef](#)]
30. Jiang, K.; Wang, Z.; Yi, P.; Wang, G.; Lu, T.; Jiang, J. Edge-enhanced GAN for remote sensing image superresolution. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5799–5812. [[CrossRef](#)]
31. Dong, X.; Xi, Z.; Sun, X.; Gao, L. Transferred multi-perception attention networks for remote sensing image super-resolution. *Remote Sens.* **2019**, *11*, 2857. [[CrossRef](#)]
32. Haut, J.M.; Paoletti, M.E.; Fernandez-Beltran, R.; Plaza, J.; Plaza, A.; Li, J. Remote sensing single-image superresolution based on a deep compendium model. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1432–1436. [[CrossRef](#)]
33. Qin, M.; Mavromatis, S.; Hu, L.; Zhang, F.; Liu, R.; Sequeira, J.; Du, Z. Remote sensing single-image resolution improvement using a deep gradient-aware network with image-specific enhancement. *Remote Sens.* **2020**, *12*, 758. [[CrossRef](#)]
34. Zhang, D.; Shao, J.; Li, X.; Shen, H.T. Remote sensing image super-resolution via mixed high-order attention network. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5183–5196. [[CrossRef](#)]
35. Lei, S.; Shi, Z.; Mo, W. Transformer-Based Multistage Enhancement for Remote Sensing Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5615611. [[CrossRef](#)]
36. Liu, Z.; Feng, R.; Wang, L.; Han, W.; Zeng, T. Dual Learning-Based Graph Neural Network for Remote Sensing Image Super-Resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5628614. [[CrossRef](#)]
37. Liu, C.; Sun, D. A bayesian approach to adaptive video super resolution. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 209–216.
38. Caballero, J.; Ledig, C.; Aitken, A.; Acosta, A.; Totz, J.; Wang, Z.; Shi, W. Real-time video super-resolution with spatio-temporal networks and motion compensation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4778–4787.
39. Wang, Y.; Lu, T.; Xu, R.; Zhang, Y. Face Super-Resolution by Learning Multi-view Texture Compensation. In *MultiMedia Modeling, Proceedings of the 26th International Conference, MMM 2020, Daejeon, Republic of Korea, 5–8 January 2020, Proceedings, Part II 26*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 350–360.
40. Zheng, H.; Ji, M.; Wang, H.; Liu, Y.; Fang, L. Crossnet: An end-to-end reference-based super resolution network using cross-scale warping. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 88–104.
41. Zhang, Z.; Wang, Z.; Lin, Z.; Qi, H. Image super-resolution by neural texture transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7982–7991.
42. Xie, Y.; Xiao, J.; Sun, M.; Yao, C.; Huang, K. Feature representation matters: End-to-end learning for reference-based image super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 230–245.

43. Yang, F.; Yang, H.; Fu, J.; Lu, H.; Guo, B. Learning Texture Transformer Network for Image Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5791–5800.
44. Huang, Y.; Zhang, X.; Fu, Y.; Chen, S.; Zhang, Y.; Wang, Y.F.; He, D. Task Decoupled Framework for Reference-Based Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5931–5940.
45. Dong, R.; Zhang, L.; Fu, H. RRSGAN: Reference-based super-resolution for remote sensing image. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–17. [[CrossRef](#)]
46. Cai, D.; Zhang, P. T<sup>3</sup>SR: Texture Transfer Transformer for Remote Sensing Image Superresolution. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7346–7358. [[CrossRef](#)]
47. Jolicoeur-Martineau, A. The relativistic discriminator: A key element missing from standard GAN. *arXiv* **2018**, arXiv:1807.00734.
48. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; Volume 2, pp. 1150–1157.
49. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
50. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [[CrossRef](#)]
51. Sheikh, H.R.; Bovik, A.C. Image information and visual quality. *IEEE Trans. Image Process.* **2006**, *15*, 430–444. [[CrossRef](#)]
52. Ranchin, T.; Wald, L. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation. *Photogramm. Eng. Remote Sens.* **2000**, *66*, 49–61.
53. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv arXiv:1409.1556* **2014**.
54. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
55. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In *Computer Vision—ECCV 2014, Proceedings of the 13th European Conference, Zurich, Switzerland, 6–12 September 2014, Proceedings, Part IV 13*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 184–199.
56. Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; Wu, W. Feedback network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3867–3876.
57. Li, J.; Yuan, Y.; Mei, K.; Fang, F. Lightweight and Accurate Recursive Fractal Network for Image Super-Resolution. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; pp. 3814–3823.
58. Fang, F.; Li, J.; Zeng, T. Soft-Edge Assisted Network for Single Image Super-Resolution. *IEEE Trans. Image Process.* **2020**, *29*, 4656–4668. [[CrossRef](#)]
59. Niu, B.; Wen, W.; Ren, W.; Zhang, X.; Yang, L.; Wang, S.; Zhang, K.; Cao, X.; Shen, H. Single image super-resolution via a holistic attention network. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 191–207.
60. Lei, S.; Shi, Z. Hybrid-scale self-similarity exploitation for remote sensing image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5401410. [[CrossRef](#)]
61. Wang, Y.; Shao, Z.; Lu, T.; Wu, C.; Wang, J. Remote sensing image super-resolution via multiscale enhancement network. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 5000905. [[CrossRef](#)]
62. Wang, Y.; Lu, T.; Wu, Z.; Wu, Y.; Zhang, Y. Face super-resolution via hierarchical multi-scale residual fusion network. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **2021**, *104*, 1365–1369. [[CrossRef](#)]
63. Chang, Y.; Yan, L.; Fang, H.; Zhong, S.; Liao, W. HSI-DeNet: Hyperspectral image restoration via convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 667–682. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.