Multistage Progressive Interactive Fusion Network for Sentinel-2: High Resolution for All Bands

Xin Liu, Xiangchao Meng, *Member, IEEE*, Qiang Liu, Xu Chen, Rui Zhao, *Member, IEEE*, and Feng Shao, *Member, IEEE*

***Abstract*—Sentinel-2 satellite remote sensing images have been widely used in various fields, such as change detection and resource monitoring. However, Sentinel-2 provides multispectral bands with inconsistent spatial resolutions (i.e., 60 m for three bands, 20 m for six bands, and 10 m for four bands), which has greatly limited the application values, especially for cooperative analysis or application on different bands at a unified resolution. In this article, we proposed a multistage progressive interactive fusion network to generate all 10-m high-resolution bands. Specifically, a refined multistage spatial resolution enhancement model in a multistage way is developed to progressively improve the low-resolution bands to preserve the spectral information of the enhanced bands. Moreover, an information interaction module is proposed for the three branches of high- (10 m), medium- (20 m), and low-resolution (60 m) bands to achieve effective information interaction. The experimental results show that our method is superior to other existing state-of-the-art methods, and it can be applied to the reconstruction of the high-resolution vegetation index.**

***Index Terms*—Convolutional neural network, deep learning, image super-resolution, Sentinel-2.**

I. Introduction

SENTINEL-2 optical remote sensing image, with a width of 290 km and 10-day revisiting interval (Sentinel 2A and 2B combined for 5 days) [1], has been widely used in many fields, such as land use change detection and vegetation health monitoring [2], [3], [4], [5], [6], [7]. However, the inconsistent spatial resolution of different spectral bands, as shown in Table I, has greatly limited the application value. On one hand, the spatial resolution of near-infrared bands and short-wave infrared (SWIR) bands are relatively low, which restricts the development of vegetation health information, coastal water, and aerosol monitoring. On the other hand, the inconsistent spatial resolution makes it difficult to cooperatively analyze at a unified resolution. Therefore, it is significant to improve the 20- and 60-m resolution bands to 10 m, to obtain Sentinel-2 multispectral images with all high spatial resolution bands of 10 m.

How to improve the 20- and 60-m bands of Sentinel-2 image to 10 m has attracted attention in recent years. On the whole, there are two main strategies: single image super-resolution [8], [9] and multisource image fusion. The single image super-resolution is to improve the spatial resolution of the image by taking full advantage of the image itself. The typical methods include the learning of a deep convolutional network for image super-resolution [10] proposed in 2015, the very deep convolutional networks [11] inspired by VGG-Net [12], the deep recursive convolutional network [13] using jump connections, and the enhanced deep super-resolution network [14]. However, single-image super-resolution cannot take full advantage of the complementary high spatial resolution information from other bands of Sentinel-2 images; correspondingly, the spatial detail enhancement is limited. Multisource image fusion aims to take advantage of the complementary information of multisource images to achieve better performance. Therefore, how to take full advantage of the 10-m bands of Sentinel-2 to improve the spatial resolution of 20- and 60-m bands has attracted more attention. The most typical multisource image fusion method is pansharpening [15], [16], [17], [18], [19], i.e., improving the spatial resolution of a multispectral image via a high spatial resolution panchromatic image. However, different from pansharpening, the spectral range between the low-resolution bands (i.e., 20 m and 60 m) and the high-resolution bands (i.e., 10 m) of Sentinel-2 images is totally different, which brings a significant challenge. For the Sentinel-2 image fusion, Wang et al. [1] proposed the regional point-to-point regression Kriging method to effectively improve the spatial resolution of 20-m bands to 10 m. Park et al. [20] proposed a series of modifications to optimize frequency band selection and synthesis, which were used for pansharpening with component substitution and multiresolution analysis. Du et al. [21] tested four popular pansharpening methods, including principle component analysis [22], intensity hue saturation, high-pass filter (HPF) [23], and à trous wavelet transform to sharpen the B11 SWIR band of Sentinel-2 for calculating the high-resolution normalized differential water index. Gasparovic and Jogun [24] evaluated five common super-resolution methods to evaluate the impact of super-resolution results on land cover classification. The experiment found that the overall classification accuracy was improved after the spatial resolution improvement of the low-resolution bands of the Sentinel-2 satellite, especially for

Manuscript received 8 August 2023; revised 25 September 2023; accepted 9 October 2023. Date of publication 23 October 2023; date of current version 17 November 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42171326 and Grant 62071261; in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR23D010001 and Grant LY22F010014; and in part by the Ningbo Natural Science Foundation under Grant 2022J076. *(Corresponding author: Xiangchao Meng.)*

The authors are with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China (e-mail: 15058453975@163.com; mengxiangchao@nbu.edu.cn; qiangliu0722@163.com; chenxu5251@163.com; 759572276@qq.com; shaofeng@nbu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3326965

TABLE I
DETAIL INFORMATION OF DIFFERENT SENTINEL-2 SATELLITE BANDS

Bands	Resolution (m)	Central wavelength (um)	Description
B1: Coastal aerosol	60	0.433	Monitoring of aerosols in nearshore water and the atmosphere
B2: Blue	10	0.490	Visible spectrum
B3: Green	10	0.560	
B4: Red	10	0.665	
B5: Vegetation red edge1	20	0.705	Monitoring vegetation health information
B6: Vegetation red edge2	20	0.740	
B7: Vegetation red edge3	20	0.783	
B8: NIR	10	0.842	Near infrared band (wide)
B8A: Narrow NIR	20	0.865	Near infrared band (narrow)
B9: Water vapor	60	0.945	Water vapor band
B10: SWIR-Cirrus	60	1.375	Short wave infrared band
B11: SWIR1	20	1.610	
B12: SWIR2	20	2.190	

the detection of forest categories, low vegetation, bare soil, and coastline.

To date, deep-learning-based methods have been the mainstream. Lanaras et al. [25] proposed Dens2 network, which is one of the earliest known articles using deep-learning residual networks to complete Sentinel-2 band super-resolution. The 20- and 60-m super-resolution network was divided into two models for separate training. Each model used six residual blocks as the main body, and the jump addition after the original image interpolation was added at the end of the network to carry out the spectral fidelity of the super-resolution image. This spectral fidelity approach inspired the later works, and many methods have been used since. Gargiulo et al. [26] believed that the super-resolution task with 20-m resolution bands of Sentinel-2 was very valuable, but the network structure of the original methods was relatively deep and the parameters were very large. So they further proposed a lightweight network [26] with four convolutional layers to complete the feature extraction and reconstruction of the Sentinel-2 images, and the network structure is similar to the classic network PanNet [27] in the super-resolution task. Wu et al. [28] proposed a parallel input network, which divided the image into three modules: parallel spatial residual learning, spatial feature fusion, and spectral feature mapping. Multiple resblocks were used for feature learning on each input stream, then the learned feature maps were added in channel dimensions, and the fully connected layer was finally used for image reconstruction. Zhang et al. [29] used the self-attention mechanism in image feature extraction. Salgueiro et al. [30] proposed a holistic Sentinel-2 band super-resolution network. Although the training was carried out in two steps, both 20- and 60-m super-resolution results could be output during the test at the same time. In addition, this method for the first time used 20-m super-resolution result output as the input of 60-m bands super-resolution network to assist 60-m bands super-resolution. The main body of the network uses a large number of residual connections and dense connections. Although the network is very deep, it also proves that the residual connections and dense connections are very helpful for image super-resolution.

Generally, several problems still exist in the existing methods. On the one hand, most existing methods generally interpolate low-resolution images to high-resolution images, and then connect them for feature extraction and reconstruction. In this way, the unique information of different resolution bands is ill-considered and relatively rough. On the other hand, since the super-resolution task contains different resolution bands of 20 and 60 m, most existing methods often fail to take into account the relationship between the two different resolution bands, the information between the bands cannot be effectively interacted, leading to the insufficient feature information utilization.

1) We propose a multistage progressive interactive fusion network to generate all high-resolution 10-m bands for Sentinel-2. In the proposed method, a refined multistage spatial resolution enhancement model is designed to progressively improve the spatial resolution of low-resolution bands while achieving the high spectral fidelity.
2) We propose a multiscale residual space spectral attention module by considering the unique features of different resolution bands to effectively extract the spatial and spectral information of images at different scales. Furthermore, an interesting information interaction module is designed to carry out effective dynamic information interaction according to the contribution degree of different branches.

The rest of this article is organized is as follows. In Section II, the proposed method is described. Section III shows the experiments and analysis. Finally, Section IV concludes this article.

II. METHODOLOGY

In this article, we propose a multistage parallel network to solve the problem of Sentinel-2 image super-resolution.

Fig. 1 shows the network structure of the proposed method. The network consists of parallel inputs of Sentinel-210-, 20-, and 60-m bands with different resolutions. According to the input images with different resolutions, three branches for the 10-, 20-, and 60-m bands are, respectively, named as high-resolution branch, medium-resolution branch, and low-resolution branch from top to bottom. The architecture of the proposed multistage progressive interactive fusion network (MPIFNet) includes three parts: the multistage super-resolution module, the information interaction module, and the fused image reconstruction part. The main purpose of the multistage super-resolution module is to improve the resolution of the low-resolution image step by step.

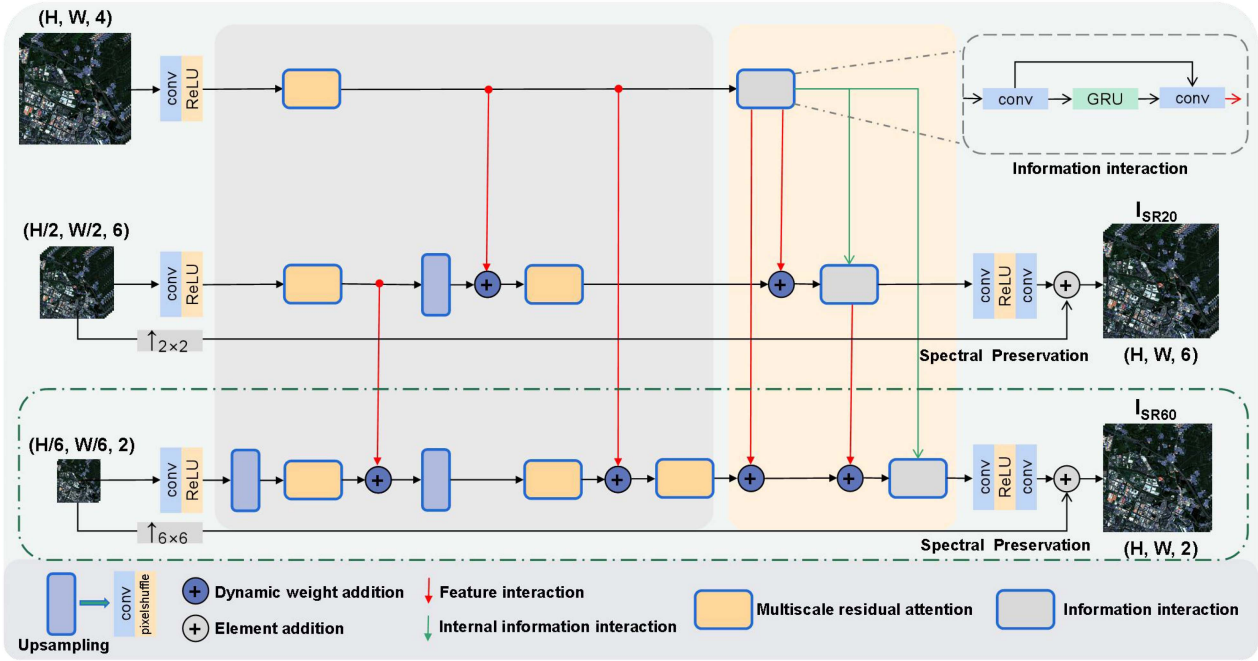


Fig. 1. Flowchart of the proposed MPIFNet.

In addition, this part also uses the multiscale residual space spectrum attention module to extract the image's fine features, and the detailed features of the higher resolution branches are injected into the lower resolution branches continuously. It is aimed to improve accuracy and delicacy in image super-resolution and feature extraction. The second part is the information interaction module based on the gate recurrent unit (GRU) [31], [32] to integrate the feature information of different branches dynamically. Finally, the super-resolution image is reconstructed through the fused image reconstruction part.

A. *Multistage Super-Resolution Module*

For the multistage super-resolution module, in order to achieve effective resolution enhancement, it is necessary to go through a fine feature extraction module containing attention modules [33], [34], [35], [36]. After obtaining the corresponding image features, the feature of the high-resolution branch will be injected into the low-resolution branch, so that we can ensure that the injected features are high-resolution image's effective information. In this experiment, the convolution layer and the pixel shuffle were selected for image upsampling.

Inspired by the convolution block attention module (CBAM) [37], a modified multiscale residual space-spectrum attention module is proposed. The module first passes the input feature image parallel through three feature extraction branches with different receptive fields (3∗3, 5∗5, and 7∗7), and then the ReLU activation function is followed for nonlinear transformation. The purpose of using convolution kernels of different sizes is to comprehensively extract multiscale features considering the characteristic of the remote sensing images. For example, a large convolution kernel can extract more global features, while a small convolution kernel can extract more detailed features. Then, the features extracted by convolution kernels of different sizes enter into the spectral attention module, respectively. Different from the channel attention module of the CBAM, a 1∗1∗C feature map is obtained after maximum pooling and average pooling. Before passing through multilayer perceptron, the feature map of the original C channels should be reduced to the number of channels of the original input image to match the features of input images with different resolutions. For example, the number of characteristic channels in the spectral attention module of the multiscale residual space-spectrum attention module of the high-resolution branch should be reduced to four channels, which is the same as the input of the high-resolution branch, and then increased to the corresponding channel number. Similarly, the number of channels in the spectral attention module of the medium-resolution branch and the low-resolution branch should be changed to 6 and 2. The subsequent spatial attention mechanism is to make a channel-based maximum pooling and average pooling to obtain two H∗W∗1 feature graphs. The sequence of maximum pooling and average pooling has been fully verified in CBAM experiments. Then, the two feature maps are based on the channel connection operation. After a 7∗7 convolution operation, the number of channels is reduced to H∗W∗1. Finally, the result of the attention module passing through the channel is obtained through a Sigmoid activation function.

After the channel connection of the feature graphs output by three different branches, the residual input of the input features should be carried out, which can better retain the original spectral features. Finally, through a simple convolution layer output, this is a complete introduction to the multiscale residual space-spectrum attention module, whose structure is shown in Fig. 2.

For the information injection module from high-resolution branch to low-resolution branch, dynamic weight addition [38]

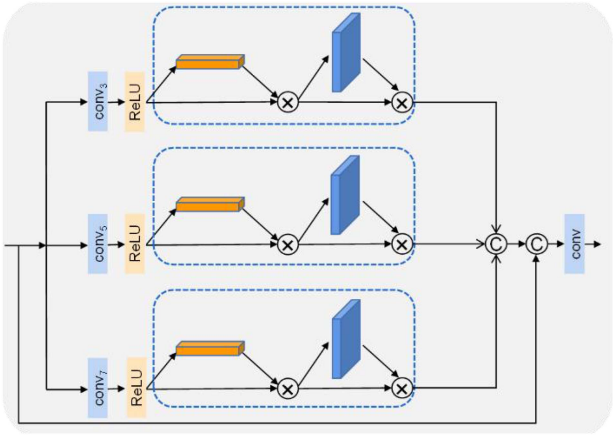


Fig. 2. Multiscale residual space spectrum attention module.

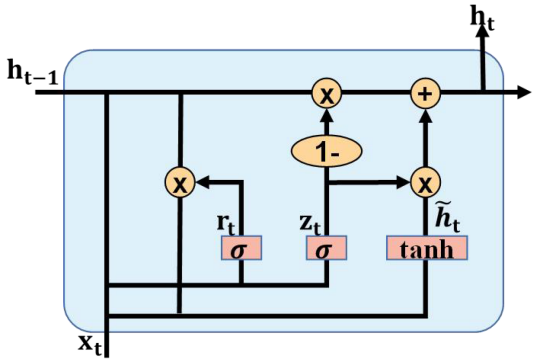


Fig. 3. Structure diagram of GRU.

is designed shown as follows, whose structure is similar to the Softmax function, that is the proportion of each input to the total input is calculated, and then weighted addition is performed:

$$\text{Output} = \sum_{\text{i}} \frac{W_i}{\sum_j W_j + \varepsilon} \times Input_{\text{i}} \tag{1}$$

where $w_i > 0$ are the dynamic learned weight. The value of ε is setting to 0.0001.

B. Information Interaction Module

The information interaction module is carried out for the information interaction among different branches. We aim to inject high spatial resolution features from high-resolution branch into the medium- and low-resolution branches to improve the spatial resolution of relatively low-resolution images. The GRU module is chosen as the main body of the information interaction module. Convolution layers are added before and after the GRU module to better extract features and integrate features. The jump connection is combined to slow down the disappearance of gradient and enhance the usability of features.

The GRU is the most widely used variant of recurrent neural network (RNN), which aims to solve the problem of gradient disappearance appearing. GRU can also be regarded as a variant of long short-term memory (LSTM) [39], [40]. Although inspired by LSTM, it is easier to calculate and can achieve the corresponding effect with LSTM. Besides, GRU has fewer parameters, so the training speed is faster and the overfitting risk can be reduced to a greater extent. Correspondingly, the GRU module is used to solve the long dependency problem in the RNN network [41], [42]. The current moment and the previous moment can be converted into feature information of different resolutions. The GRU module is used to judge the importance of different information. Adjusting the dependency relationship, to control the flow and interaction of information. The structure of GRU is shown in the Fig. 3 below:

Here, X_t is the input of the current moment, h_{t-1} is the hidden state of the previous moment, and h_t is the calculated hidden state of the current moment. According to Fig. 1, the input state X_t in the current moment is the feature information obtained for each branch. Every image that passes through the GRU module has the same spatial resolution. In addition, h_{t-1} denotes the feature information obtained from higher resolution branches. The last hidden state of the GRU module of the medium and low variable branches are the current output state h_t of the GRU module of the high-resolution branch. Since the GRU module of the high-resolution branch has only one input, its status at the previous time is set to 0. As can be seen from the figure above, GRU is composed of reset gate and update gate. Before hiding the state of the current moment, GRU will first calculate a candidate state $\widetilde{h}_t$, and the calculation of this candidate state needs the calculation of the reset gate, the calculation formula of reset door r_t is formula 2. Where, W_r is the parameter of the reset gate, and σ is the Sigmoid activation function, which can limit the value of the update gate between [0, 1]

$$r_t = \sigma(W_r[\mathrm{h_{t-1}}, \mathrm{x_t}]). \tag{2}$$

If the reset gate is close to 0, candidate state $\widetilde{h}_t$ will ignore the hidden state of h_{t-1} at the previous time, and use the current input state x_t to do the calculation. In other words, the feature information from the higher resolution branches will be ignored. In this way, it could allow hidden state h_t at the current moment to avoid any irrelevant information, using only relevant information to focus the calculation. The candidate state $\widetilde{h}_t$ is calculated as

$$\tilde{h}_t = \tanh(W_h[\mathrm{r_t} \times \mathrm{h_{t-1}}, \mathrm{x_t}]). \tag{3}$$

After the candidate state is calculated, the update gate is used to control how much information in the previous state can be transferred to the current state. For our experiment, the update gate can control the degree of information interaction. The calculation formula of the update gate is as follows. The calculation formula of the reset door is the same, but the weight matrix is changed

$$z_t = \sigma(W_r[\mathrm{h_{t-1}}, \mathrm{x_t}]). \tag{4}$$

Finally, the current state of the moment can be calculates as

$$h_t = (1 - \mathrm{z_t}) \times \mathrm{h_{t-1}} + z_t \times \tilde{\mathrm{h}}_{\mathrm{t}}. \tag{5}$$

According to the above formula, there is a competitive relationship between the corresponding branch and higher branches. As z_t gets bigger, output from GRU module h_t will be taken

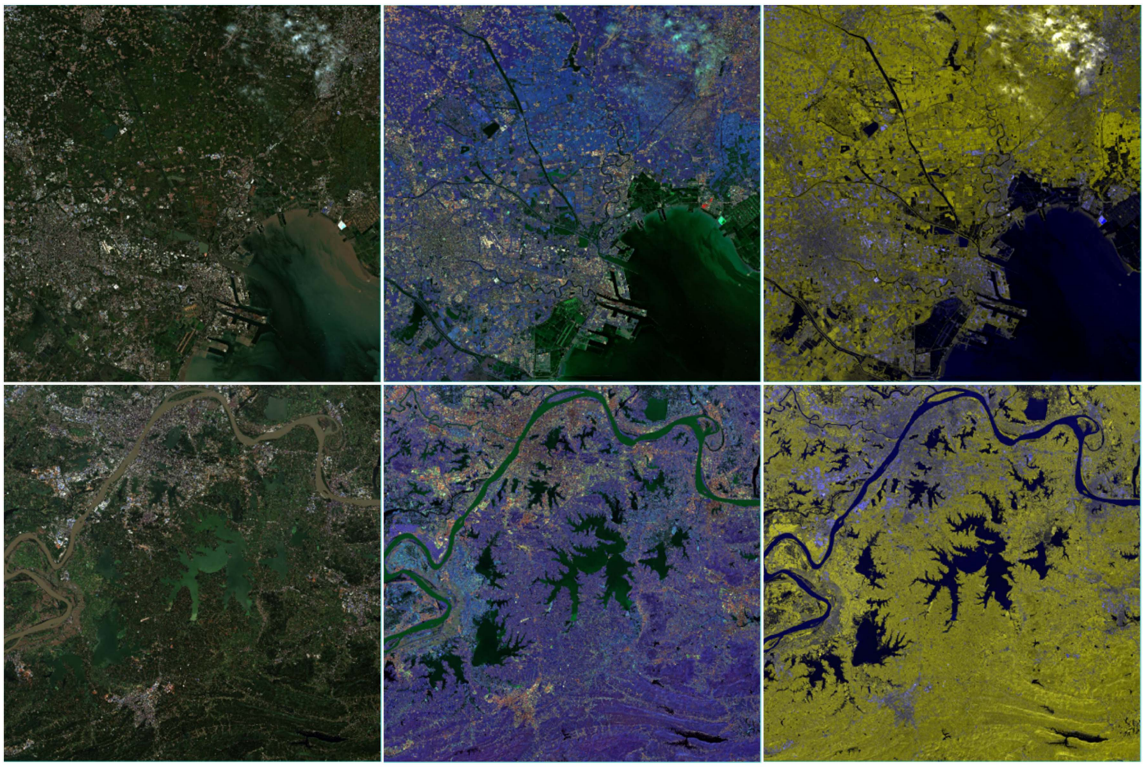


Fig. 4. Data used in the experiment. The first is in Tianjin and the second is in Wuhan. (a) RGB true color display for 10-m resolution image. (b) {B12, B5, B6} false color display for 20-m resolution image. (c) {B9, B9, B1} false color display for 60-m resolution image. (a) 10 m. (b) 20 m. (c) 60 m.

more from this branch; when z_t is 0, the current moment state h_t is completely taken from the higher resolution branches.

C. Fused Image Reconstruction

The experiment uses convolution layers and activation layer for image reconstruction. The main purpose is to convert the features in the hidden layer into real images, and the second is to control the number of channels to output images. After the network reconstruction, it enters the spectral fidelity stage.

The main part of the network is to improve the spatial resolution of low-resolution images. For remote sensing image enhancement, it is not only necessary to improve the spatial resolution but also to ensure that the enhanced image can retain the spectral information of the original image. For the preservation of spectral information, we choose to use a long skip connection, which will add the original interpolated image and the reconstructed result graph, so as to obtain the final output image, which is the enhanced result graph

$$X_f = X_{up} + X_h \tag{6}$$

where X_{up} represents the upsampled multispectral image, X_h represents reconstruction images, and X_f denotes the super-resolution image.

To continuously optimize the network parameters, $L1$-loss is used as a loss function in this experiment. $L1$-loss is called the mean absolute error (MAE), which refers to the average distance between the predicted value X_i of the model and the real value Y_i. The formula is as follows:

$$L1\_loss(X, Y) = \frac{1}{n}\sum_{i=1}^{n} |\mathrm{Y_i} - \mathrm{X_i}| \tag{7}$$

where X is super-resolution image, Y is reference image, and n is sample size. The lower the value, the better the model effect.

III. Experiments

A. Data

In the experiment, two Sentinel-2 L1C images were selected. These images are shown in the Fig. 4. These images were acquired in September 2021, and the spatial dimension was 10 980∗10 980. One of the scene data is located in Tianjin, China. There are very rich types of ground objects, including urban buildings, farmland, and sea area. The rich types of ground objects can better verify the universality of the method. Another scene is located in Wuhan, China, which contains the city, plain farmland, numerous barrier lakes, and part of the Yangtze River. These two scenes of data were used, respectively, in the reduced-resolution experiments and the full-resolution experiments.

B. Experimental Details

The reduced-resolution and full-resolution experiments were conducted. For the reduced resolution experiment of 20-m band super-resolution task, the Wald at al.'s protocol [43] was used to reduce the original 10- and 20-m images to 20 and 40 m,

respectively. Similarly, for the 60-m band super-resolution task, the Wald's protocol was also used to reduce the original 10-, 20-, and 60-m images to 60, 120, and 360 m, respectively. Then, the original image was separated with 20% data as the testing set and 80% data as the training set. Finally, the dataset is partitioned. For the reduced resolution experiment of 20-m super-resolution task, the training set was cut into 5642 blocks, every training block's size is 60∗60. The testing set was cut into 1725 blocks, every testing block's size is 144∗144. For the 60-m super-resolution task, the training set was cut into 600 blocks, every training block's size is 60∗60. The testing set was cut into 168, every testing block's size is 144∗144. Each network undergoes 500 times of preliminary training, and in the respective training process, these training blocks were input to the network with 32 inputs per batch. The number of network parameters is 6.25 m and flops is 448.23 M. The experiment used the Adam optimizer [45] and $L1$-norm as the loss function, and the initial learning rate was $1*10^{-4}$. After every 100 training cycles, the learning rate was half of the original.

For the full-resolution experiment, the parameters of the training network are the same as those of the full-resolution experiment, but the block processing on the testing set is different. Since the full-resolution experiment does not need to use Wald's protocol to downsample the original image, in order to make the original image more realistic, the test block is trimmed to a slightly larger size. In the 20- and 60-m super-resolution networks, every block's size is 288∗288.

C. Quantitative Evaluation Metrics

In order to quantitatively evaluate the output results, seven reference-based evaluation indexes of structural similarity (SSIM), peak signal-to-noise ratio (PSNR), universal image quality index (UIQI), Erreur Relative Globale Adimensionnelle De Systhese (ERGAS), CC, spatial correlation coefficient (SCC), and spectral angle mapper (SAM) were used in the reduced-resolution experiments, and three nonreference evaluation indexes of quality w/no reference index (QNR), D_λ, and D_S were applied in the full-resolution experiments. Here, *X* denotes super-resolution image and *Y* denotes reference image.

SSIM [44]: SSIM is an index to measure the similarity of two images. The value range of SSIM is [−1,1], and the larger the value is, the more similar the two images is

$$\mathrm{SSIM} = \frac{(2u_{\mathrm{X}}u_{\mathrm{Y}} + c_1)(2\sigma_{\mathrm{XY}} + c_2)}{(u_{\mathrm{X}}^2 + u_{\mathrm{Y}}^2 + c_1)(\sigma_{\mathrm{X}}^2 + \sigma_{\mathrm{Y}}^2 + c_2)} \tag{8}$$

where u_x is the mean value of X, u_Y is the mean value of Y, σ_X^2 is the variance of X, σ_Y^2 is the variance of Y, and σ_{XY} is the covariance of X and Y. $c_{1=(K_1L)}^2$and $c_{2=(K_2L)}^2$ are two coefficients that are used to keep things stable. L is a dynamic indicator of the pixel value. $K_1 = 0.01$ and $K_2 = 0.03$.

PSNR: PSNR is an important index to measure the quality of reconstructed images, where MSE represents the mean square error of the super-resolution image and the reference image

$$\mathrm{MSE} = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(\mathrm{X}(i,j) - \mathrm{Y}(i,j))^2 \tag{9}$$

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{MAX_{\mathrm{Y}}^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX_{\mathrm{Y}}}{\sqrt{MSE}}\right) \tag{10}$$

where MAX_Y indicates the largest pixel in the reference image. The larger the value of PSNR, the better the quality of the super-resolution image.

UIQI: Its calculation formula is as follows:

$$\mathrm{UIQI} = \frac{4\sigma_{\mathrm{XY}}u_{\mathrm{X}}u_{\mathrm{Y}}}{(\sigma_{\mathrm{X}}^2 + \sigma_{\mathrm{Y}}^2)\left[(u_{\mathrm{X}})^2 + (u_{\mathrm{Y}})^2\right]}. \tag{11}$$

ERGAS:

$$\mathrm{ERGAS} = \frac{100}{S}\sqrt{\frac{1}{B}\sum_{i=1}^{B}\left[\frac{\sqrt{MSE(\mathrm{X_i},\mathrm{Y_i})}}{E(\mathrm{Y_i})}\right]^2} \tag{12}$$

where *B* represents the number of channels of the image, *S* is the scale factor of the two images, and *E* is the calculated mean. The smaller the value, the better the quality.

SAM [45]: It is an evaluation index used to measure the spectral fidelity between two images. Its calculation formula is as follows:

$$\mathrm{SAM} = \cos^{-1}\left(\frac{\mathrm{XY}}{\|\mathrm{X}\|_2\|\mathrm{Y}\|_2}\right) \tag{13}$$

where $\cos^{-1}$ is arccosine. The smaller the value, the better the spectral quality of the image.

SCC: The similarity of two images in spatial details can be better determined, which is the same as CC. The larger the value, the higher the spatial similarity between the two images

$$\mathrm{SCC} = \frac{\sigma_{HP_{\mathrm{X}}HP_{\mathrm{Y}}}}{\sqrt{\sigma_{HP_{\mathrm{X}}}^2\sigma_{HP_{\mathrm{Y}}}^2}} \tag{14}$$

where HP denotes HPF.

QNR [46]: The QNR is a popular nonreference quality evaluation index. It is composed of two parts: the spectral index D_λ and spatial D_S index. The ideal values of the D_λ and D_S are 0, and the corresponding ideal index for QNR is 1

$$\mathrm{QNR} = (1 - \mathrm{D}_\lambda)^\alpha(1 - \mathrm{D_s})^\beta. \tag{15}$$

For spectral information evaluation

$$D_\lambda \triangleq \sqrt[p]{\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|Q\left(\hat{\mathrm{G}}_\mathrm{i},\hat{\mathrm{G}}_\mathrm{j}\right) - Q\left(\tilde{\mathrm{G}}_\mathrm{i},\tilde{\mathrm{G}}_\mathrm{j}\right)\right|^p} \tag{16}$$

where $\{\widehat{G}_i\}_i^N$ denotes super-resolution image bands, $\{\widetilde{G}_i\}_i^N$ denotes low-resolution image bands, and *N* denotes number of multispectral image bands. *Q* is used to calculate the correlation between two bands and *p* is generally set to 1.

For spatial information evaluation

$$D_s = \sqrt[q]{\frac{1}{N}\sum_{i=1}^{N}\left|Q(\widehat{\mathrm{G}_\mathrm{i}},\mathrm{P}) - Q(\tilde{\mathrm{G}}_\mathrm{i},\tilde{\mathrm{P}})\right|^q}. \tag{17}$$

Among them, *P* is the PAN image and $\widehat{P}$ is PAN image with the same resolution of multispectral image. *q* is also usually set to 1. The smaller the value of D_λ and D_S, the better the quality of the reconstructed image.

TABLE II
QUANTITATIVE EVALUATION RESULTS FOR 20-M BAND SUPER-RESOLUTION

	SSIM↑	PSNR↑	UIQI↑	ERGAS↓	SCC↑
Bicubic	0.8289	27.7093	0.9628	6.8911	0.5510
Dsen2$_{2018\ ISPRS}$	0.9850	38.2953	0.9969	2.0727	0.9524
SPR-Net$_{2020\ RS}$	0.9829	38.1633	0.9968	2.1072	0.9508
Sen2-RDSR$_{2021\ RS}$	**0.9893**	38.6491	0.9971	1.9911	0.9562
Our propose	0.9887	**38.6578**	**0.9972**	**1.9902**	**0.9563**

The best results were marked in bold and the second-best results were marked in underline.

TABLE III
QUANTITATIVE EVALUATION RESULTS FOR 60-M BAND SUPER-RESOLUTION

	SSIM↑	PSNR↑	UIQI↑	ERGAS↓	SCC↑
Bicubic	0.4310	21.2280	0.9000	9.0862	0.1502
Dsen2$_{2018\ ISPRS}$	0.8930	30.7759	0.9887	3.0727	0.8166
SPR-Net$_{2020\ RS}$	0.8889	30.7883	0.9888	3.0748	0.8127
Sen2-RDSR$_{2021\ RS}$	0.8925	31.1472	0.9895	2.9102	0.8432
Our propose	**0.8965**	**31.2112**	**0.9896**	**2.8876**	**0.8542**

The best results were marked in bold and the second-best results were marked in underline.

TABLE V
FULL-RESOLUTION RESULTS FOR 20-M BAND SUPER-RESOLUTION

	QNR↑	D_{λ} ↓	D_S ↓
Dsen2$_{2018\ ISPRS}$	0.3345	0.1048	0.6262
SPR-Net$_{2020\ RS}$	0.3337	0.1046	0.6273
Sen2-RDSR$_{2021\ RS}$	0.3374	0.1042	0.6232
Our propose	**0.3543**	**0.1026**	**0.6052**

The best results were marked in bold and the second-best results were marks in underline

TABLE IV
EVALUATION RESULTS OF RED-EDGED VEGETATION INDEX

	CC↑	ERGAS↓	RMSE↑
Bicubic	0.8783	14.7373	0.0667
Dsen2$_{2018\ ISPRS}$	0.9719	7.0852	0.0320
SPR-Net$_{2020\ RS}$	0.9704	7.2746	0.0329
Sen2-RDSR$_{2021\ RS}$	0.9735	6.8886	0.0312
Our propose	**0.9736**	**6.8412**	**0.0309**

The best results were marked in bold and the second-best results were marks in underline.

D. Experimental Results

1) Reduced-Resolution Experiment: Some state-of-the-art deep-learning-based methods were carried out for comparison, so as to demonstrate the feasibility of the proposed method. The experimental results for 20- and 60-m band super-resolution are shown in Tables II and III, respectively.

According to Tables II and III, it can be seen that the proposed method achieves the best results in four indexes for the 20-m bands super-resolution experiment, and has a significant improvement compared with other methods. Another index SSIM also achieved the second-best result. Similarly, for the 60-m bands super-resolution experiment, the proposed method achieves the best results for all the indexes, indicating that the proposed method can achieve good performance in the improvement of spatial details and the fidelity of spectral information.

In order to better understand the differences between the proposed and other comparison methods, MAE residuals were provided, as shown in Figs. 5 and 6. Among them, the more yellow the color, the greater the difference from the reference image, and the bluer the color, the smaller the difference from the reference image, and the higher the quality of the reconstructed image. The experimental results are as follows.

According to the figures above, it can be clearly seen that the method proposed in this article has the best performance in the 20- and 60-m band super-resolution experiments. The blue area occupies a larger area and the color is darker, indicating that the method proposed in this article is effective for the Sentinel-2 super-resolution task.

2) Experimental Evaluation on Red-Edge Vegetation Index: To comprehensively verify the performance of the proposed method, the evaluation of super-resolution results cannot only stay in the two aspects of spatial details and spectral resolution, and the practical application of enhanced images should also be added to the image evaluation. Accordingly, we conducted the evaluation of the different fusion results from application viewpoint on red-edge vegetation index [30], which is calculated is as follows:

$$\mathrm{NDVI-RE} = \frac{\mathrm{B7} - \mathrm{B5}}{\mathrm{B7} + \mathrm{B5}} \tag{18}$$

where *B*5 and *B*7 are both bands within the red-edge range, which is effective for monitoring vegetation health.

Finally, evaluation indexes, such as CC, ERGAS, and RMSE, were calculated by using the red-edge vegetation index map of each method and reference image. According to the quality of the evaluation index, the validity of the super-resolution image for vegetation identification and detection can be judged.

Some details of NDVI-RE results were shown in Fig. 7. The blue part is the high vegetation coverage area, and the red part is the low vegetation coverage area. As can be seen from the results in Fig. 7, this method is more accurate in the division of vegetation area and nonvegetation area, which proves that the spatial details and spectral information of the proposed fusion image are more consistent with the reference image. The quantitative evaluation results on the NDIV-RE are shown in Table IV. It shows that the proposed super-resolution image has competitive performance in the reconstruction of the red-edge vegetation index.

3) Full-Resolution Experiment: The full-resolution experiments were conducted on original Sentinel-2 data. The high-resolution branch input a band combination map with a resolution of 10 m, the medium-resolution branch input a band combination map with a resolution of 20 m, and the low-resolution branch input a band combination map with a resolution of 60 m. The experimental results are shown in Fig. 8. In addition, quantitative evaluation results with the widely used nonreference indexes of QNR, D_{λ}, and D_S were shown in Tables V and VI.

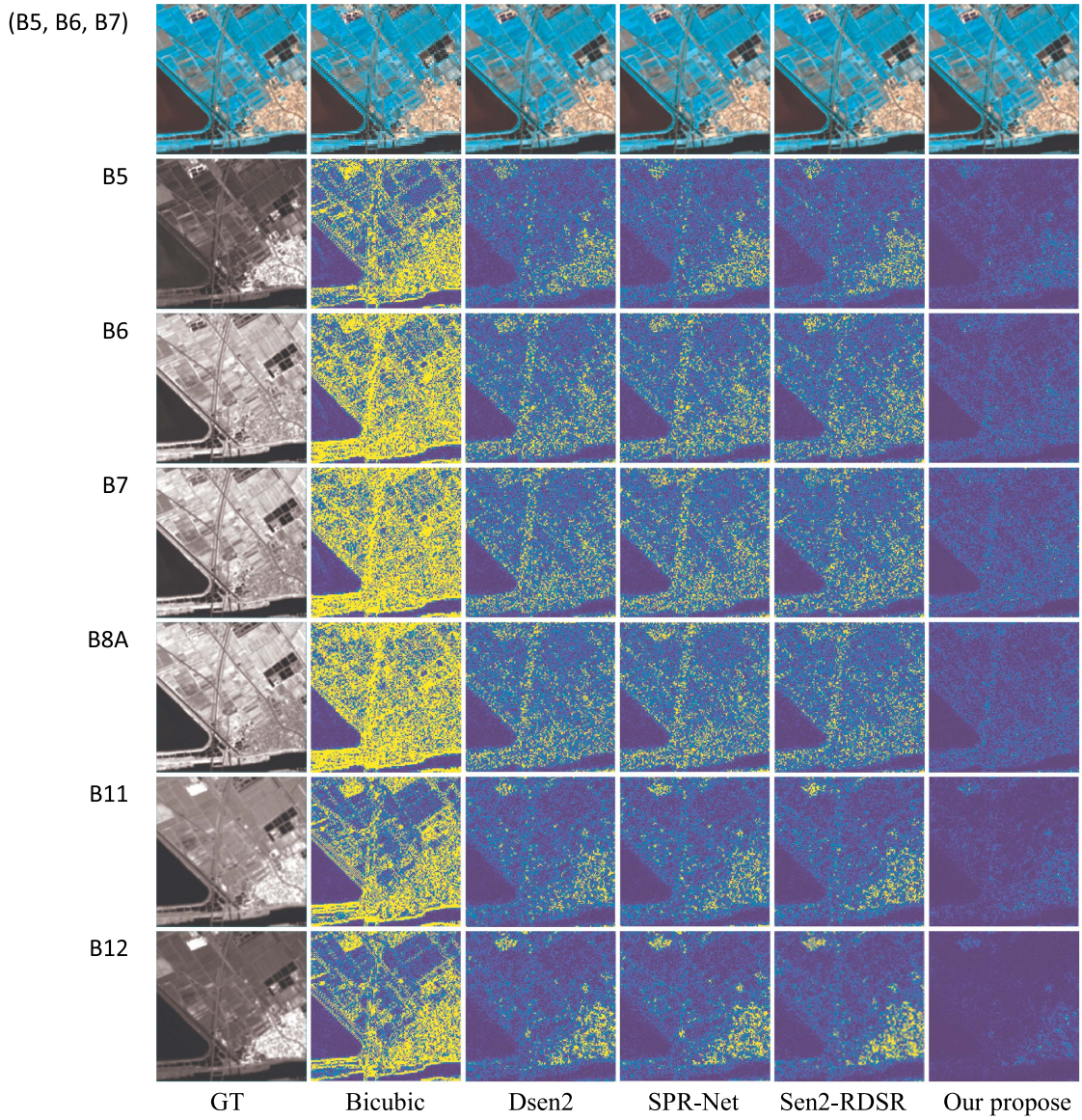


Fig. 5. Residual map of super-resolution results at 20-m bands.

TABLE VI
FULL-RESOLUTION RESULTS FOR 60-M BAND SUPER-RESOLUTION

	QNR↑	D_{λ} ↓	D_S ↓
Dsen2$_{2018\ ISPRS}$	0.9596	**0.0193**	0.0216
SPR-Net$_{2020\ RS}$	0.9562	0.0226	0.0218
Sen2-RDSR$_{2021\ RS}$	0.9535	0.0243	0.0229
Our propose	**0.9601**	0.0218	**0.0187**

The best results were marked in bold and the second-best results were marks in underline.

As can be seen from the above table, the proposed method achieves the best results in all three full-resolution indexes in the 20-m band super-resolution task, especially in QNR and D_S. For the full-resolution experiment of 60-m bands, this experiment also achieved the best results in two indexes, and D_{λ} ranked the second place. To sum up, our method has achieved competitive results in the experiments of 20-m bands and 60-m bands, which proves that the proposed network can balance the construction of spatial details and the preservation of spectral information and obtain high-quality full-resolution images.

E. Discussion

1) Ablation Experiment and Analysis: In order to verify the validity of some modules in this method, ablation experiments for the multistage progressive super-resolution stage, the information interaction module and the dynamic weight addition module were performed. For convince, we simply named the multistage progressive super-resolution stage, information interaction module, and the dynamic weight addition module as MP, IIM, and DWM, respectively. The experimental results were shown in Tables VII and VIII. We remove a pixelshuffle from 60-m super-resolution network and increased the resolution of

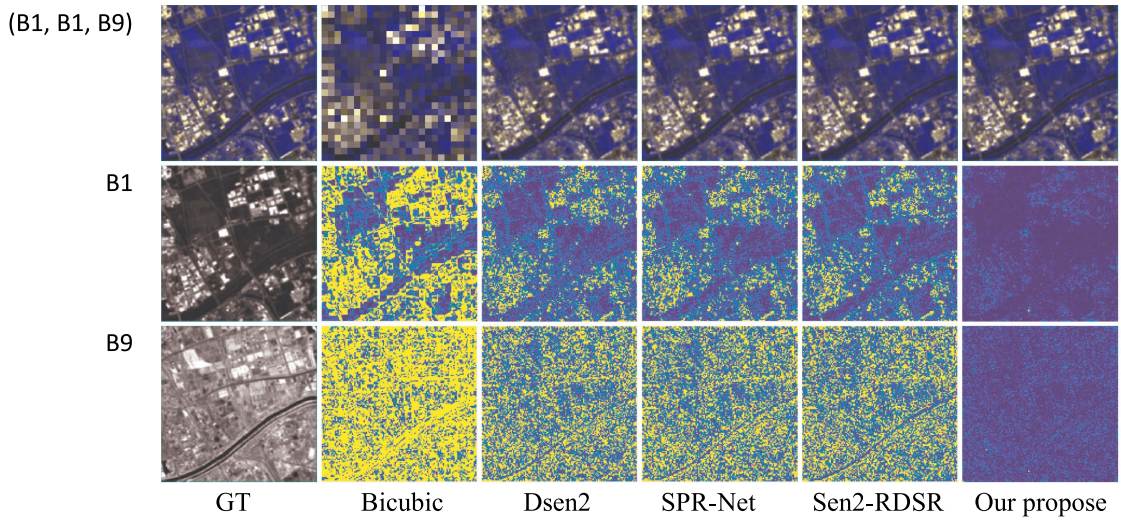


Fig. 6. Residual map of super-resolution results at 60-m bands.

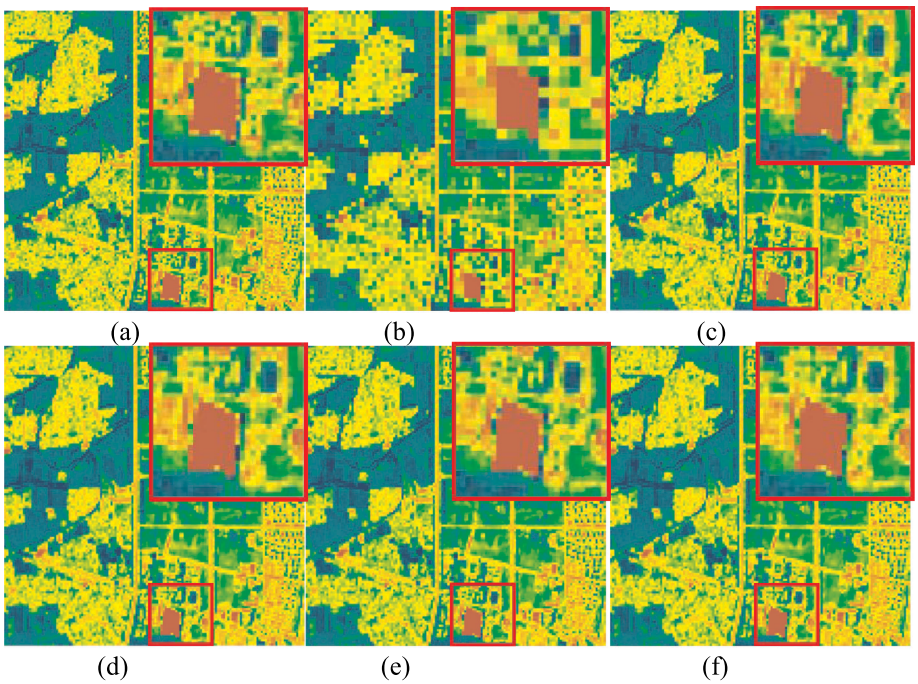


Fig. 7. NDVI-RE experimental result. (a) GT. (b) Bicubic. (c) Dens2. (d) SPR-Net. (e) Sen2-RDSR. (f) Our propose.

TABLE VII
ABLATION EXPERIMENTAL RESULTS OF 20-M BAND SUPER-RESOLUTION

	SSIM$_\uparrow$	PSNR$_\uparrow$	UIQI$_\uparrow$	ERGAS$_\downarrow$	SAM$_\downarrow$	CC$_\uparrow$	SCC$_\uparrow$
w/o MP & IIM & DWM	0.9854	38.4689	0.9963	2.0305	1.5396	0.9934	0.9553
w/o IIM & DWM	<u>0.9867</u>	38.4773	0.9970	2.0288	1.5380	0.9939	0.9553
w/o DWM	0.9863	<u>38.5623</u>	<u>0.9971</u>	<u>2.0103</u>	<u>1.5217</u>	<u>0.9940</u>	<u>0.9554</u>
W IIM & DWM & MP	**0.9887**	**38.6578**	**0.9972**	**1.9902**	**1.4964**	**0.9942**	**0.9563**

The best results were marked in bold and the second-best results were marks in underline.

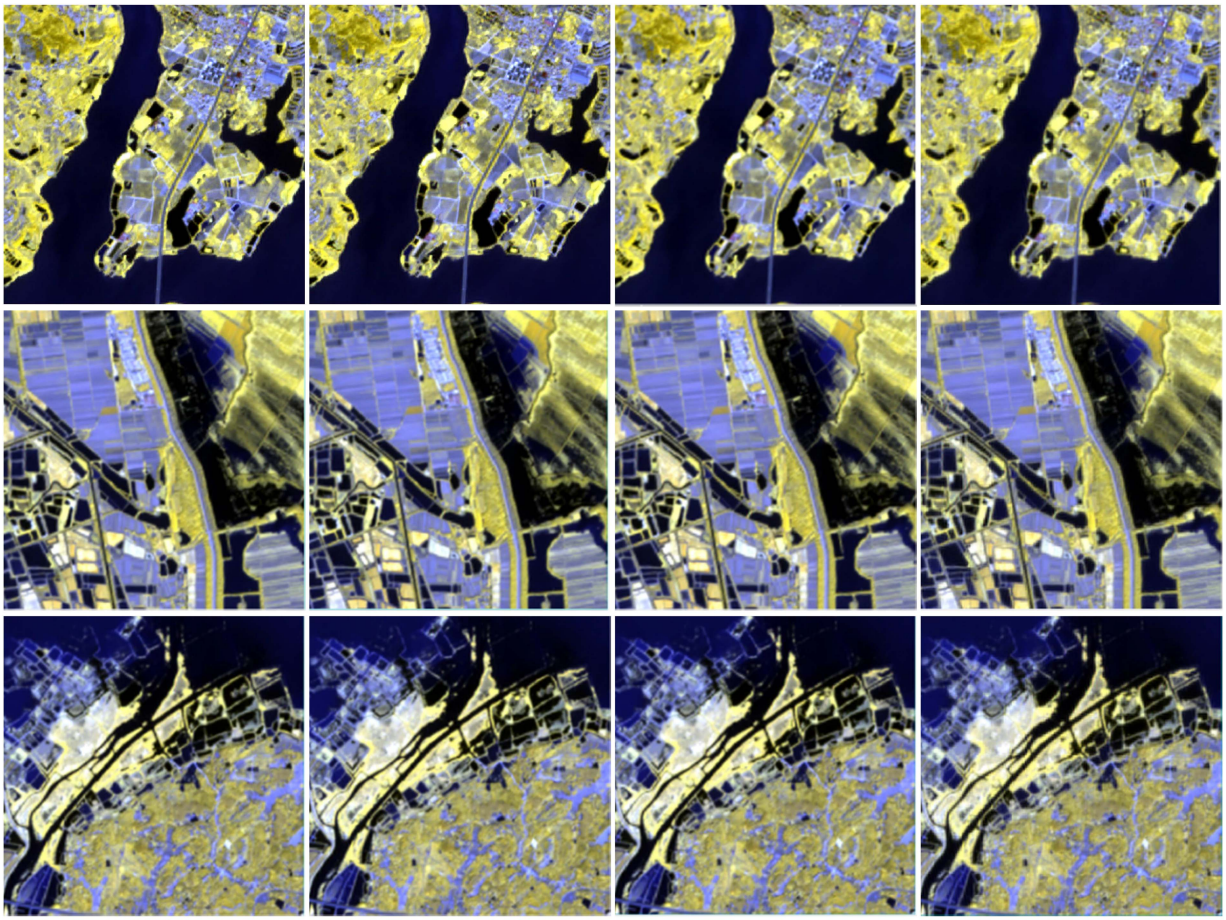


Fig. 8. Full-resolution experimental results of 20-m bands. {B7, B6, B5} false color display. (a) Dens2. (b) SPR-Net. (c) Sen2-RDSR. (d) our.

TABLE VIII
ABLATION EXPERIMENTAL RESULTS OF 60-M BAND SUPER-RESOLUTION

	SSIM$_\uparrow$	PSNR$_\uparrow$	UIQI$_\uparrow$	ERGAS$_\downarrow$	SAM$_\downarrow$	CC$_\uparrow$	SCC$_\uparrow$
w/o MP & IIM & DWM	0.8895	30.7586	0.9881	2.9175	2.1490	0.9544	0.8559
w/o IIM & DWM	0.8903	30.9129	0.9886	2.9101	2.0312	0.9568	**0.8566**
w/o DWM	<u>0.8942</u>	<u>31.0732</u>	<u>0.9890</u>	<u>2.9059</u>	<u>1.9635</u>	<u>0.9572</u>	<u>0.8563</u>
W IIM & DWM & MP	**0.8965**	**31.2112**	**0.9896**	**2.8867**	**1.9137**	**0.9588**	0.8542

The best results were marked in bold and the second-best results were marks in underline

60-m images to 10 m in one go. We display the result of removing all three parts in the first column. Compared with the second row, the results without MP were relatively poor, which notes the multistage progressive super-resolution stage improves the effect of super-resolution. In addition, compared with the second row, the information interaction module is added in the third row. According to the following table, after the addition of the information interaction module, various indicators have been improved correspondingly, indicating that the information interaction module is effective for information communication and detail injection of branches. And in the fourth row of the table, the multistage progressive super-resolution stage, the information interaction module and dynamic weight addition module were added at the same time. Compared with the network without the three modules, the performance of the network has been improved, especially in the SAM index of the evaluation of spectral information.

2) Advantages and Disadvantages:

1) The advantage of the proposed method is the full consideration of the relationship between different resolution bands, and the MPIFNet is designed to gradually enhance the low-resolution bands, to preserve the spectral information of the enhanced bands as much as possible. In addition, the feature extraction module composed of the attention mechanism and the information interaction module is proposed to improve the performance.

2) However, the super-resolution of 20/60 m images with two different resolutions is still divided into two independent works. The advantage of this is that the super-resolution effect can be greatly improved, but the network training stage needs more time. In fact, if Wald's protocol is used for degradation, two independent tasks will inevitably be separated. This is also the disadvantage of current Sentinel-2 satellite super-resolution. Therefore, an effective unsupervised super-resolution work should pay attention in the future.

IV. Conclusion

In this article, we proposed a multistage super-resolution network for Sentinel-2 images, which inputs observations with three different resolutions in parallel. In each stage of resolution improvement, feature extraction was carried out based on spectral and spatial attention modules and detailed feature injection of high-resolution images was received. When low-resolution images were upgraded to the same resolution as high-resolution images, information interaction based on the GRU module was carried out in the three-flow branches, for better low-resolution image detail reconstruction. The proposed method was comprehensively tested and verified from different aspects, including reduced- and full-resolution experiments on quantitative and qualitative evaluation and also the application viewpoint on vegetation index. The experimental results showed that the proposed method achieved superior performance to other existing deep-learning methods.

References

[1] Q. Wang, W. Shi, Z. Li, and M. P. Atkinson, "Fusion of Sentinel-2 images," *Remote Sens. Environ.*, vol. 187, pp. 241–252, 2016.

[2] M. Lange, H. Feilhauer, I. Kühn, and D. Doktor, "Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series," *Remote Sens. Environ.*, vol. 277, 2022, Art. no. 112888.

[3] W. Frampton, J. Dash, G. Watmough, and E. Milton, "Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 82, pp. 83–92, 2013.

[4] Sentinel-2 User Handbook, 2013. Accessed: Mar. 21, 2021. [Online]. Available: http://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook

[5] A. Hornero et al., "Monitoring the incidence of Xylella fastidiosa infection in olive orchards using ground-based evaluations, airborne imaging spectroscopy and Sentinel-2 time series through 3-D radiative transfer modelling," *Remote Sens. Environ.*, vol. 236, 2020, Art. no. 111840.

[6] C. Li, L. Zhou, and W. Xu, "Estimating aboveground biomass using Sentinel-2 MSI data and ensemble algorithms for grassland in the Shengjin Lake Wetland, China," *Remote Sens.*, vol. 13, no. 8, 2021, Art. no. 1595.

[7] J. Tan et al., "Inducing flooding index for vegetation mapping in water-land ecotone with Sentinel-1 & Sentinel-2 images: A case study in Dongting Lake, China," *Ecol. Indicators*, vol. 144, 2022, Art. no. 109448.

[8] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *Trans. Image Process.*, vol. 23, no. 6, pp. 2569–2582, 2014.

[9] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNet: A fast and lightweight network for single-image super resolution," *Trans. Cybern.*, vol. 51, no. 3, pp. 1443–1453, 2021.

[10] C. Dong, C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[11] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[13] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1637–1645.

[14] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1132–1140.

[15] X. Meng, H. Shen, H. Li, L. Zhang, and R. Fu, "Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges," *Inf. Fusion*, vol. 46, pp. 102–113, 2019.

[16] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, 2016.

[17] J.-S. Choi, Y. Kim, and M. Kim, "S3: A spectral-spatial structure loss for pan-sharpening networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 892–833, May 2020.

[18] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.

[19] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 594.

[20] H. Park, J. Choi, N. Park, and S. Choi, "Sharpening the VNIR and SWIR bands of Sentinel-2A imagery through modified selected and synthesized band schemes," *Remote Sens.*, vol. 9, no. 10, 2017, Art. no. 1080.

[21] Y. Du, Y. Zhang, F. Ling, Q. Wang, W. Li, and X. Li, "Water bodies mapping from Sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the SWIR band," *Remote Sens.*, vol. 8, no. 4, 2016, Art. no. 354.

[22] V. Shah, N. Younan, and R. King, "An efficient pan-sharpening method via a combined adaptive-PCA approach and contourlets," *Remote Sens.*, vol. 46, no. 5, pp. 1323–1335, 2008.

[23] M. Fahnestock, T. Scambos, T. Moon, A. Gardner, H. Terry, and K. Marin, "Rapid large-area mapping of ice flow using Landsat 8," *Remote Sens. Environ.*, vol. 185, pp. 89–94, 2016.

[24] M. Gasparovic and T. Jogun, "The effect of fusing Sentinel-2 bands on land-cover classification," *Remote Sens.*, vol. 39, no. 3, pp. 822–841, 2018.

[25] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 146, pp. 305–319, 2018.

[26] M. Gargiulo, A. Mazza, R. Gaetano, G. Ruello, and G. Scarpa, "Fast super-resolution of 20 m Sentinel-2 bands using convolutional neural networks," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2635.

[27] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1753–1761.

[28] J. Wu, Z. He, and J. Hu, "Sentinel-2 sharpening via parallel residual network," *Remote Sens.*, vol. 12, 2020, Art. no. 279.

[29] R. Zhang, G. Cavallaro, and J. Jitsev, "Super-resolution of large volumes of Sentinel-2 images with high performance distributed deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 617–620.

[30] L. Salgueiro, J. Marcello, and V. Vilaplana, "Single-image super-resolution of Sentinel-2 low resolution bands with residual dense convolutional neural networks," *Remote Sens.*, vol. 13, no. 24, 2021, Art. no. 5007.

[31] Y. Hao, Y. Sheng, and J. Wang, "Variant gated recurrent units with encoders to preprocess packets for payload-aware intrusion detection," *IEEE Access*, vol. 7, pp. 49985–49998, 2019.

[32] A. Agarap, "A neural network architecture combining gated recurrent unit (GRU) and support vector machine (SVM) for intrusion detection in network traffic data," in *Proc. 10th Int. Conf. Mach. Learn. Comput.*, 2017, pp. 26–30.

[33] I. Sutskever, O. Vinyals, and Q. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, vol. 37, pp. 3104–3112.

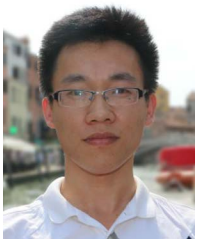
[34] K. Xu, J. Ba, and R. Kiros, "Show, attend and tell: Neural image caption generation with visual attention," *Comput. Sci.*, vol. 37, pp. 2048–2057, 2015.

[35] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, *arXiv:1508.04025*.

[36] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.
[37] S. Woo, J. Park, J. Lee, and I. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 8–14.
[38] X. Wu, T.-Z. Huang, L.-J. Deng, and T.-J. Zhang, "Dynamic cross feature fusion for remote sensing pansharpening," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 14667–14676.
[39] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, pp. 1735–1780, 1997.
[40] S. Sun, L. Mu, L. Wang, and P. Liu, "L-UNet: An LSTM network for remote sensing image change detection," *Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[41] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
[42] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.
[43] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetry Eng. Remote Sens.*, vol. 63, pp. 691–699, 1997.
[44] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[45] R. Yuhas, A. Goetz, and J. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.
[46] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogrammetry Eng. Remote Sens.*, vol. 74, pp. 193–200, 2008.

Xin Liu received the B.S. degree in internet of things engineering in 2020 from the Ningbo University, Ningbo, China, where he is currently working toward the M.S. degree in communication engineering.

His research interests include deep learning for image processing and multisource image fusion.



Xiangchao Meng (Member, IEEE) received the B.S. degree in geographic information system from the Shandong University of Science and Technology, Qingdao, China, in 2012, and the Ph.D. degree in cartography and geography information system from the Wuhan University, Wuhan, China, in 2017.

He is currently an Associate Professor with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China. His research interests include multisource data fusion and machine learning.

Qiang Liu received the B.S. degree from the Shenyang Ligong University, Shenyang, China, in 2016, and the M.S. degree from the School of Ningbo University, Ningbo, China, in 2019. He is currently working toward the Ph.D. degree with the Ningbo University.

His research interests include deep learning for image processing, artificial intelligence, and multisource image fusion.

Xu Chen received the B.S. degree in communication engineering in 2021 from the Ningbo University, Ningbo, China, where he is currently working toward the M.S. degree in communication engineering.

His research interests include hyperspectral image processing and multisource data fusion.

Rui Zhao (Member, IEEE) received the B.S. degree in photogrammetry and remote sensing from the Wuhan University, Wuhan, China, in 2012, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2017.

He worked as a Postdoctor with the School of Earth and Space Sciences, Peking University, Beijing, China, from 2020 to 2022. He is currently a Lecturer with the Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, China. His research interests include hyperspectral image processing, artificial intelligence, pattern recognition, deep learning, and machine learning.

Dr. Zhao serves as a Reviewer for Science Citation Index magazines including IEEE TGRS, JSTARS, and GRSL.

Feng Shao (Member, IEEE) received the B.S. and Ph.D. degrees in electronic science and technology from the Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively.

He was a Visiting Fellow with the School of Computer Engineering, Nanyang Technological University, Singapore, from 2012 to 2012. He is currently a Professor with the Faculty of Information Science and Engineering, Ningbo University, Ningbo, China. He has authored/coauthored more than 100 technical articles in refereed journals and proceedings in the areas of 3-D video coding, 3-D quality assessment, and image perception.

Dr. Shao was the recipient of the Excellent Young Scholar Award from the NSF of China, in 2016.