# Landsat-8 and Sentinel-2 Image Fusion Based on Multiscale Smoothing-Sharpening Filter

Peng Wang, *Senior Member, IEEE*, Mingxuan Huang, Shupeng Shi, Bo Huang, *Member, IEEE*, Bilian Zhou, Gang Xu, Liguo Wang, and Henry Leung, *Fellow, IEEE*

***Abstract*—With the increasing demand for high temporal and spatial resolution multispectral image sequences, many studies have been carried out on fusion on Landsat-8 and Sentinel-2 images to obtain image sequences with a revisit cycle of 2 and 3 days and a spatial resolution of 10 m. However, current fusion methods suffer from complex computation and loss of spectral and spatial information. To address these issues, a Landsat-8 and Sentinel-2 image fusion based on multiscale smoothing-sharpening filter (MSSF) method is proposed. MSSF combines well the initial spatial prediction obtained from the Landsat-8 image at the target date and the detailed image extracted from the Sentinel-2 image at the reference date. Thin plate spline interpolation with morphological opening-closing algorithm is implemented on the Landsat-8 image at the target date, and the Laplacian of Gaussian enhancement algorithm is applied to the Sentinel-2 image at the reference date in the preprocessing stage. Smoothing-sharpening filter (SSIF) is employed to separate the high and low frequency components of the two preprocessed images. The multiscale SSIF is then utilized to migrate the details from the preprocessed Sentinel-2 image to the preprocessed Landsat-8 image. The performance of MSSF and five compared methods was evaluated qualitatively and quantitatively. Experiments on three remote sensing data sets gathered from different experimental sites confirm that the proposed MSSF method could efficiently generate Sentinel-2-like images with high spatial and spectral resolution.**

***Index Terms*—Image filtering, image fusion, Landsat-8, remote sensing image, Sentinel-2.**

Received 10 September 2024; revised 18 September 2024; accepted 23 September 2024. Date of publication 27 September 2024; date of current version 11 October 2024. This work was supported in part by the Technology Innovation Center for Collaborative Applications of Natural Resources Data in GBA, MNR under Grant 2024NRDK01, in part by the Hunan Key Laboratory of Remote Sensing Monitoring of Ecological Environment in Dongting Lake Area under Grant DTH Key Lab.2024-11, in part by the Central Universities in Nanjing University of Aeronautics and Astronautics under Grant NS2023020, in part by the Open Fund of Wenzhou Future City Research Institute under Grant WL2023004, in part by the National Natural Science Foundation of China under Grant 61801211, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20221478, in part by the Hong Kong Scholars Program under Grant XJ2022043, in part by the Postgraduate Research and Practice Innovation Program of NUAA under Grant xcxjh20230406, and in part by the Graduate Education and Teaching Reform Research Project in NUAA under Grant 2024YJXGG19. *(Corresponding authors: Bo Huang; Shupeng Shi.)*

Peng Wang is with the Hunan Key Laboratory of Remote Sensing Monitoring of Ecological Environment in Dongting Lake Area, Changsha 410004, China, also with the Technology Innovation Center for Collaborative Applications of Natural Resources Data in GBA, Ministry of Natural Resources, Guangdong 510062, China, and also with the Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China.

Mingxuan Huang is with the Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China.

Shupeng Shi is with the Hunan Key Laboratory of Remote Sensing Monitoring of Ecological Environment in Dongting Lake Area, Changsha 410004, China (e-mail: hngdbhdc@126.com).

Bo Huang is with the Department of Geography, The University of Hong Kong, Hong Kong (e-mail: bohuang@hku.hk).

Bilian Zhou is with the Land and Resources Technical Center of Guangdong Province, Guangdong 510062, China, and also with the Technology Innovation Center for Collaborative Applications of Natural Resources Data in GBA, Ministry of Natural Resources, Guangdong 510062, China.

Gang Xu is with Wenzhou Future City Research Institute, Wenzhou 325016, China, and also with the Zhejiang College of Security Technology, Wenzhou 325016, China.

Liguo Wang is with the College of Information and Communication Engineering, Dalian Minzu University, Liaoning 116600, China.

Henry Leung is with the Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: leungh@ucalgary.ca).

Digital Object Identifier 10.1109/JSTARS.2024.3469974

## I. Introduction

SURFACE reflectance image sequences acquired by medium spatial resolution satellites play an important role [1], [2], [3] in crop yield estimation [4], [5], natural disaster monitoring [6], environmental pollution monitoring [7], [8], climate change research [9], [10], and mineral resources exploration [11]. Among the medium spatial resolution data, Landsat-8 operational land imager and Sentinel-2 multispectral imager have been widely used. Landsat-8 provides data in 11 bands, of which 10 bands have a spatial resolution of 30 m and one panchromatic band has a spatial resolution of 15 m [12]. This satellite can achieve global coverage once every 16 days. Sentinel-2 consists of two satellites, named Sentinel-2A and Sentinel-2B, with a five-day revisit cycle [13]. It provides data in 13 bands, including 4 bands at 10 m spatial resolution, 6 bands at 20 m spatial resolution, and 3 bands at 60 m spatial resolution [14]. Although Sentinel-2 data combine high spatial and temporal resolution, weather factors such as clouds and rain can render a large amount of data unusable. In addition, the 5-day revisit period makes it impossible to monitor rapid surface changes [15], [16]. Combining Landsat-8 and Sentinel-2 data can reduce the revisit period to 2 and 3 days [17], greatly expanding the application value of the data series. Moreover, Landsat-8 and Sentinel-2 images have similar wavelengths and the same geographic coordinate system [18], [19], which makes it feasible to monitor surface changes at higher frequencies by fusing Landsat-8 and Sentinel-2 data [20].

Many methods have been developed for Landsat-8 and Sentinel-2 image fusion. Wang et al. [21] extend area-to-point regression kriging (ATPRK) to Landsat-8 and Sentinel-2 image fusion and utilized the 15 m panchromatic band of the Landsat-8

image and 20 m bands of the Sentinel-2 image as auxiliary references. Another type of method is based on the classical spatiotemporal image fusion (STIF) models. For example, the spatial and temporal adaptive reflectance fusion model (STARFM) proposed by Gao et al. [22] is the first weight function-based STIF model, which selects similar neighboring pixels within a moving window and then calculates the reflectance value of the center pixel using a linear weighting function. The flexible spatiotemporal data fusion (FSDAF) method proposed by Zhu et al. [23] combines weighting and unmixing to predict temporal variation through unmixing, followed by distributing residuals guided by spatial interpolation results. Guan et al. [24] propose an object-based STIF framework using an object-restricted similar pixel selection method that significantly improves the ability of the similar pixel-based STIF models to fuse Landsat-8 and Sentinel-2 images. But most classical STIF methods require at least three input images, i.e., a low-resolution image at the target date and two high-resolution images at the reference date. Since the coverage dates of Landsat-8 and Sentinel-2 barely overlap, it makes this input requirement difficult to fulfill. To solve this problem, Wu et al. [25] perform image degradation and radiometric correction processes on the input images, enabling classical STIF models to accomplish Landsat-8 and Sentinel-2 image fusion with only two input images. In recent years, some methods based on deep learning have been developed [26], [27]. Wu et al. [28] develop a degradation-term constrained spatiotemporal fusion network (DSTFN), which uses a structure combining residual dense blocks and attention mechanism modules to enhance feature-level image fusion. Cheng et al. [29] propose a multiscale and attention mechanism-based residual spatiotemporal fusion network (MARSTFN), combining features obtained from the multiscale and attention mechanisms, residual networks and skip connections.

Although the studies mentioned earlier have produced good performance, they still encounter some problems in practice. Machine learning-based methods usually have trouble in preserving spatial and spectral information and suffer from spectral distortion, over-smoothing, and over-sharpening. It has been shown that ATPRK suffers from over-sharpening [30]. Classical STIF models are mainly used to accomplish Landsat and MODIS image fusion, where the zoomed-in scale factor of spatial resolution is 16–20 times. However, the scale factor in Landsat-8 and Sentinel-2 image fusion is three times. Moreover, compared to the target of STIF, i.e., Landsat images, the resolution of Sentinel-2 images is higher, which means finer structures to be reconstructed. Therefore, applying classical STIF models directly to Landsat-8 and Sentinel-2 image fusion could cause spectral distortions and lack of detailed features in the fused images [31], [32]. Deep learning-based methods generally achieve better fusion results, but they require many training samples with high computational complexity.

To solve the abovementioned problems, this article proposes a Landsat-8 and Sentinel-2 image fusion based on multiscale smoothing-sharpening filter (MSSF) method. We focus on the fusion of Landsat-830-m bands and Sentinel-210-m bands. The proposed MSSF mainly utilizes a smoothing-sharpening filter (SSIF) [33], which controls smoothing and sharpening by tuning parameter with desirable edge-awareness property. In addition, preprocessing is designed in MSSF to enhance the quality of input images. MSSF preserves the main features of the Landsat-8 image at the target date and transfers detailed information of the Sentinel-2 image at the reference date. The contributions of this work are as follows.

1) The proposed MSSF contains a novel Landsat-8 and Sentinel-2 image fusion framework by designing a multiscale SSIF. The high and low frequency components are separated through SSIF. A multiscale SSIF is designed to extract and transfer the missing details of the Landsat-8 image from the high frequency components of the Sentinel-2 image, enhancing the spatial details of the fused image.
2) MSSF boosts two input images by using denoising and enhancement in preprocessing. The Landsat-8 image is upsampled via thin plate spline (TPS) interpolation [34] followed by morphological opening-closing algorithm [35] to reduce noise. The Sentinel-2 image is enhanced with Laplacian of Gaussian (LOG) enhancement algorithm [36] to reduce noise and enhance spatial details.

The rest of this article is organized as follows. Section II describes the proposed method in detail. Section III conducts the experiments on three data sets. Section IV presents discussions on parameter selection and performance of the proposed MSSF. Section V draws the conclusions.

## II. Proposed Method

As shown in Fig. 1, the proposed method utilizes a Sentinel-2 image $S$ at the reference date $t_0$ and a Landsat-8 image $L$ at the target date $t_p$ to produce a fused image $S_{proposed}$ at $t_p$ with the resolution of the Sentinel-2 image. $L$ is upsampled via TPS interpolation, and denoised via morphological opening-closing algorithm. While $S$ is denoised, the detailed information is enhanced through LOG enhancement algorithm in the preprocessing step. SSIF is then employed to extract high frequency components of the preprocessed Landsat-8 image and Sentinel-2 image. The high frequency components are processed by multiscale SSIF to migrate the missing detailed image of the Landsat-8 image from the Sentinel-2 image.

### A. Preprocessing

The proposed method employs TPS to upsample the original Landsat-8 image at $t_p$ to the size of the Sentinel-2 image, because TPS can produce smooth results and retains most of the spatial information. TPS is a robust spatial data interpolation algorithm based on spatial correlation, which builds a surface passing through all control points via TPS function and minimizes curvature amount of that surface, measured by an energy function [37]. The interpolation result of pixel $(x, y)$ in band $k$ is calculated as follows:

$$L_{up}^{k}(x,y) = a_0 + a_1 x + a_2 y + \sum_{l=1}^{L} b_l d_l^2 \log d_l \qquad (1)$$

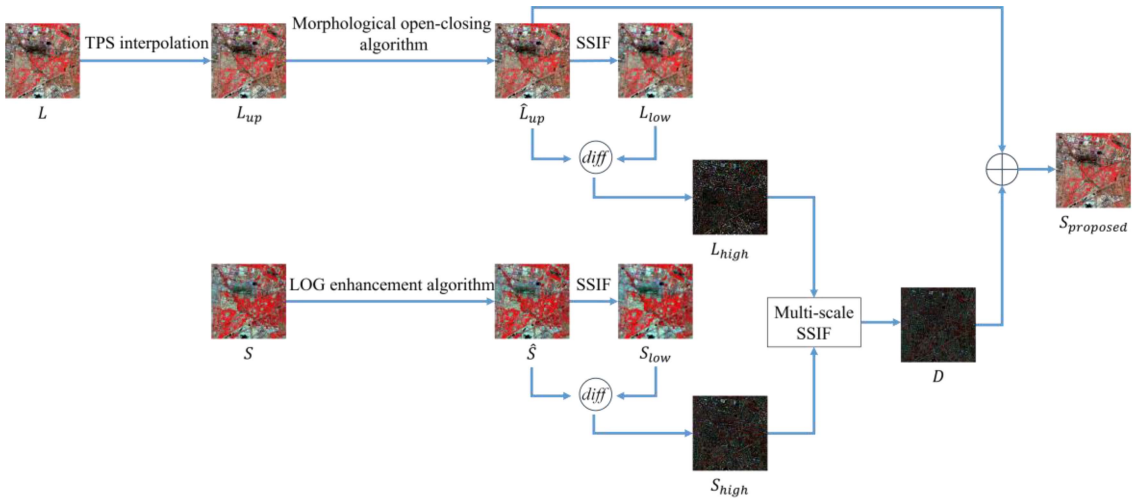



Fig. 1. Flowchart of MSSF model.

where $a_0$, $a_1$, $a_2$, and $b_l$ are parameters determined by minimizing the energy function, and $L$ is the number of known control points $(x_l, y_l)$, $d_l = \sqrt{(x - x_l)^2 + (y - y_l)^2}$. $b_l$ should satisfy the constraints in the following equation:

$$\sum_{l=1}^{L} b_l = \sum_{l=1}^{L} b_l x_l = \sum_{l=1}^{L} b_l y_l = 0. \tag{2}$$

Noise and artifacts may exist in the original image due to sensor and environmental conditions, which creep into the fused image and reduce the accuracy of the fusion result [38]. Since upsampling also tends to cause noise amplification and blurring [39], it is necessary to reduce the influence of noise on Landsat-8 and Sentinel-2 images in the preprocessing step. The proposed method applies a simple and effective denoising procedure to the upsampled Landsat-8 image by morphological opening-closing algorithm as shown in (3). The opening algorithm serves to remove tiny objects and unnecessary edges, while the closing algorithm tends to fill tiny voids within the object [40].

$$\hat{L}^k_{up}(x, y) = [L^k_{up}(x, y) \circ S(m, n)] \bullet S(m, n) \tag{3}$$

where $S(m, n)$ is the square structuring element. $\circ$ indicates the opening algorithm which is morphological erosion followed by morphological dilation, $\bullet$ indicates the closing algorithm which is morphological dilation followed by morphological erosion. The morphological erosion and morphological dilation are calculated as follows:

$$L^k_{up}(x, y) \odot S(m, n) = \min_{m,n}(L^k_{up}(x + m, y + n) - S(m, n)) \tag{4}$$

$$L^k_{up}(x, y) \oplus S(m, n) = \max_{m,n}(L^k_{up}(x - m, y - n) + S(m, n)) \tag{5}$$

where $\odot$ represents the morphological erosion, and $\oplus$ represents the morphological dilation.

Spatial details in the Sentinel-2 image at $t_0$ play an important role in the fusion process. The optimized image $\hat{S}$ is obtained by applying LOG enhancement algorithm to the Sentinel-2 image $S$, where the noise is reduced by the Gaussian filter and the spatial details are enhanced by the Laplace operator. The optimized image $\hat{S}$ processed by LOG operator with center point $(0, 0)$ and standard deviation $\sigma$ is generated as follows:

$$\hat{S}(x, y) = S(x, y) + S(x, y) * \frac{x^2 + y^2 - 2\sigma^2}{2\pi\sigma^6} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{6}$$

where $*$ is the convolution operator.

### B. SSIF

While the TPS interpolation captures the main spatial information in the Landsat-8 image, such as land cover changes and local variations, considerable texture and edge information is missing due to resolution deficiencies. These details are mainly present in the high frequency components of an image, and it is reasonable to assume that they do not change between $t_0$ and $t_p$ [41]. Therefore, the proposed method extracts the high frequency components of the two images and migrates the details from the Sentinel-2 image at $t_0$ band by band.

This is realized by SSIF since it is an edge-aware filter and the level of smoothing or sharpening can be easily adjusted by only one tuning parameter. The guidance image can be the input image or another image. Similar to the classical guided filter [42], SSIF develops a patch interpolation model. There are $M = (2r + 1)^2$ pixels contained in a patch with a radius of $r$, and the $i$th patch is denoted as $\Omega_i$. Let $I_i(q)$ represent the pixel value at position $q$ in $\Omega_i$, the output $J_i(q)$ of SSIF is

$$J_i(q) = \mu_i + \text{sign}(\phi_i)\alpha_i(G_i(q) - v_i) \tag{7}$$

where $\mu_i$ is the mean of the input image patch, $\phi_i$ is the covariance of the input and guidance image patch, $G_i(q)$ is the pixel value at position $q$ in $\Omega_i$ in the guidance image, and $v_i$ is the mean of the guidance image patch. $\alpha_i$ is the coefficient of the local linear model between $J_i(q)$ and $G_i(q)$. $\alpha_i$ is a constant for which an appropriate value needs to be determined to minimize the model error. The negative log-likelihood for the patch is

calculated as follows:

$$-\log p(\{J_i(q)\}|\alpha_i) = \frac{1}{2\tau^2 M}\sum_{q\in\Omega_i}\left(I_i(q)-J_i(q)\right)^2 = \frac{\varsigma_i^2}{2}\alpha_i^2 - |\phi_i|\,\alpha_i \tag{8}$$

where $\varsigma_i^2 = \frac{1}{M}\sum_{q\in\Omega_i}\left(G_i(q)-v_i\right)^2$ is the covariance of the guidance image patch, and $\tau$ is set to 1. Considering that $\alpha_i$ satisfies the generalized Gamma distribution [43] as a priori, the cost function $D(\alpha_i)$ can be obtained.

$$D(\alpha_i) = -\log p(\{J_i(q)\}|\alpha_i) - \log p(\alpha_i) = \frac{\varsigma_i^2}{2}\alpha_i^2 - |\phi_i|\,\alpha_i + \frac{1}{2\theta^2}\alpha_i^2 - \eta\log\alpha_i \tag{9}$$

where $\theta > 0$ is a scale parameter and $\eta \geq 0$ is a parameter to control the shape of the distribution. The parameter settings are adjusted to $\epsilon = \frac{1}{\theta}$, $\eta = \kappa\epsilon$ to ensure that the selection between smoothing ($0 \leq \kappa < 1$) and sharpening ($\kappa > 1$) is controlled only by $\kappa$. Let $\frac{\partial D}{\partial \alpha_i} = 0$, $\alpha_i$ can be obtained as follows:

$$\alpha_i = \frac{1}{2}\left\{\frac{|\phi_i|}{\varsigma_i^2+\epsilon} + \sqrt{\left(\frac{|\phi_i|}{\varsigma_i^2+\epsilon}\right)^2 + \frac{4\kappa\epsilon}{\varsigma_i^2+\epsilon}}\right\}. \tag{10}$$

The pixel at position $q$ belongs to $M$ overlapping patches because blocks containing $M$ pixels are built centered on each pixel. The final result is a weighted average of the outputs produced by each patch, specifically, the filtered pixel value at position $q$ is

$$J(q) = G(q)\sum_{i=1}^{M} w_i \text{sign}(\phi_i)\alpha_i + \sum_{i=1}^{M} w_i(\mu_i - \text{sign}(\phi_i)\alpha_i v_i) \tag{11}$$

where $G(q)$ represents the pixel value at position $q$ in the guidance image, and the weight $w_i$ is calculated by

$$w_i = \frac{c_i}{1+\left(\varsigma_i^2/(s\bar{\varsigma}^2)\right)^2} \tag{12}$$

where $c_i$ is a regularization parameter which ensures $\sum_{i=1}^{M} w_i = 1$, $s$ is a user-defined scale parameter and $\bar{\varsigma}^2$ is the average of all $\varsigma_i^2$. SSIF$(\cdot)$ denotes the SSIF operator in this article.

The smoothing function of SSIF serves as a low-pass filter with an edge-preserving effect, which extracts the low frequency components of the preprocessed Landsat-8 and Sentinel-2 images. The high frequency components $L_{high}$ and $S_{high}$ are obtained by subtracting the low frequency components from the corresponding filter inputs.

$$L_{high} = \hat{L}_{up} - \text{SSIF}\left(\hat{L}_{up}\right) \tag{13}$$

$$S_{high} = \hat{S} - \text{SSIF}\left(\hat{S}\right). \tag{14}$$

### C. Multiscale SSIF Structure

The high frequency components of the Sentinel-2 image are employed as the initial input image, and the high frequency components of the Landsant-8 image are employed as the guidance image. SSIF is utilized to extract the missing detail parts of $\hat{L}_{up}$. In order to bring the detailed image closer to the land cover at $t_p$ rather than $t_0$, a multiscale SSIF structure [44], [45] is designed where filtered outputs are processed repeatedly by SSIF. A dual-scale example of the proposed multiscale SSIF is shown in Fig. 2.

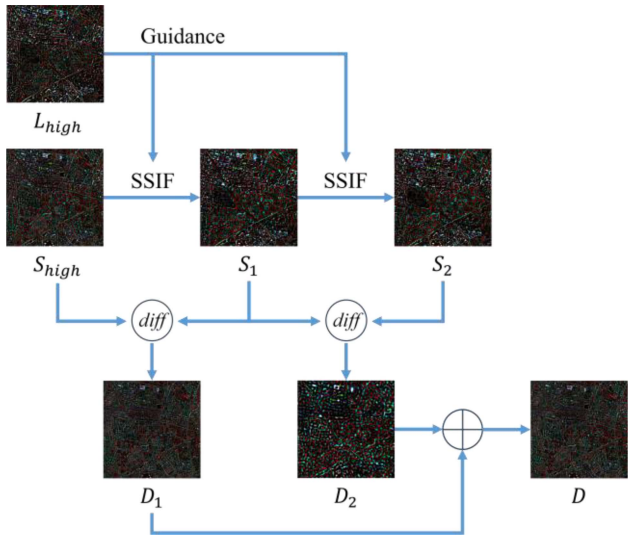



Fig. 2. Flowchart of a dual-scale example of multiscale SSIF.

$$S_j = \text{SSIF}\left(S_{j-1}\right) \tag{15}$$

where $j$ is the number of repetitions, and each repetition increases the scale by one, yielding a blurrier filtered image and a detailed image which is the difference between two neighboring filtered images calculated by (15). When $j = 1$, $S_{j-1} = S_{high}$.

$$D_j = S_{j-1} - S_j. \tag{16}$$

The total detailed image is the sum of detailed images at all scales

$$D = \sum_{j=1}^{N} D_j \tag{17}$$

where $N$ is the number of scales. The final fusion result is obtained by summing the total detailed image $D$ with $\hat{L}_{up}$

$$S_{proposed} = \hat{L}_{up} + D. \tag{18}$$

## III. Experiments

Three remote sensing data sets consisting of Landsat-8 Level-2 and Sentinel-2 Level-1C products collected from two regions are used in our experiments. Four bands are selected in the experiments, including the blue band, the green band, the red band, and the near-infrared (NIR) band. The composite of all images uses NIR-red-green as RGB. The performance of the proposed method is compared with four methods, namely ATPRK, STARFM, FSDAF, the regression model fitting, spatial filtering and residual compensation (Fit-FC) [46] method, and DSTFN. The parameters in all compared methods are configured according to the corresponding original papers. All classical

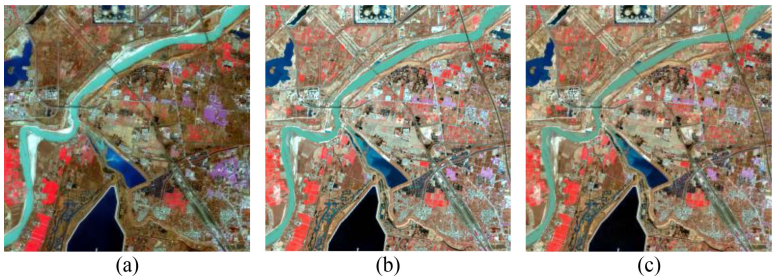



Fig. 3. (a) Sentinel-2 image of Dezhou captured on December 16, 2018, (b) Landsat-8 image, and (c) Sentinel-2 image captured on March 16, 2018.

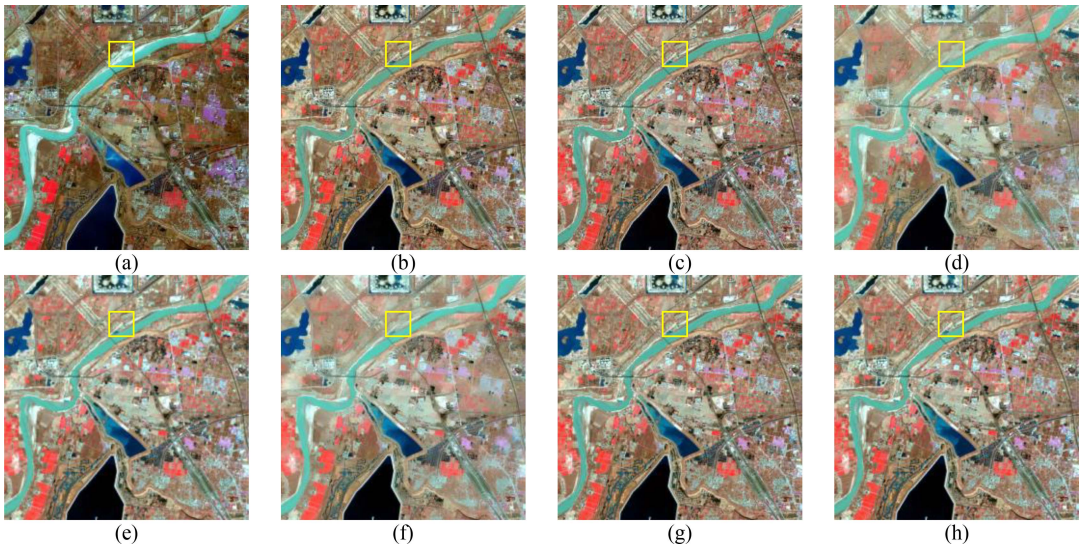



Fig. 4. (a) Actual Sentinel-2 image at the Dezhou study area observed on December 16, 2018. (b) Sentinel-2 image observed on March 16, 2018, and the fused images generated by (c) ATPRK, (d) STARFM, (e) FSDAF, (f) fit-FC, (g) DSTFN, and (h) MSSF.

STIF-based methods, i.e., STARFM, FSDAF, and Fit-FC, adopt corresponding simplified versions to accommodate scenes with only two input images. DSTFN is trained using a data set collected at the Dezhou site. The scale of the multiscale SSIF in the proposed MSSF is uniformly set to 2. Structure similarity (SSIM) [47], peak signal-to-noise ratio (PSNR) [48], root mean square error (RMSE) [49], correlation coefficient (CC) [50], and spectral angle mapping (SAM) [51] are employed as quality indices. Higher SSIM, PSNR and CC, as well as lower RMSE and SAM indicate superior fusion results.

The data sets used in Experiments 1 and 2 were collected in Dezhou, Shandong Province, China. Each Sentinel-2 image covers $990 \times 990$ pixels, while each Landsat-8 image covers $330 \times 330$ pixels. Changes in the farmlands in this region are synchronized with the growth cycles of two kinds of crop, namely winter wheat, which is sowed in the previous year and matures in early June, and maize, which is sowed in late June and matures in early October. In addition, the land cover types of the region include urban regions and some decentralized villages.

The data set used in Experiments 3 was collected at the border across Hailar and Qiqihar in Northeast China. Each Sentinel-2 image covers $1320 \times 1320$ pixels, while each Landsat-8 image covers $440 \times 440$ pixels. The main land cover types in the area involve extensive farmlands, woodlands and lakes, and a small number of residential areas. Among them, woodlands are homogeneous with slow phenological changes, while farmlands are heterogeneous with significant phenological changes, which are influenced by human activities. For convenient description, the data sets used in the three experiments are abbreviated as Scenario 1, Scenario 2, and Scenario 3, respectively, in the subsequent sections.

### *A. Experiment 1-Images From a Region Experiencing Land Cover Changes*

The Sentinel-2 image on December 16, 2018 [Fig. 3(a)] and the Landsat-8 image on March 16, 2018 [Fig. 3(b)] were used to generate a 10-m image on March 16, 2018, and the actual Sentinel-2 image [Fig. 3(c)] was employed as a comparison image. Fig. 4 presents the two input images, the actual Sentinel-2 image and the fused images obtained by different methods. The enlarged images of the area indicated in Fig. 4 are depicted in Fig. 5.

This area contains various land cover types with fine structure and undergoes land cover changes and phenological changes, which makes fusion difficult. All methods tested are able to detect major land cover changes. However, the fusion results of STARFM and Fit-FC suffer from inaccurate pixel values and

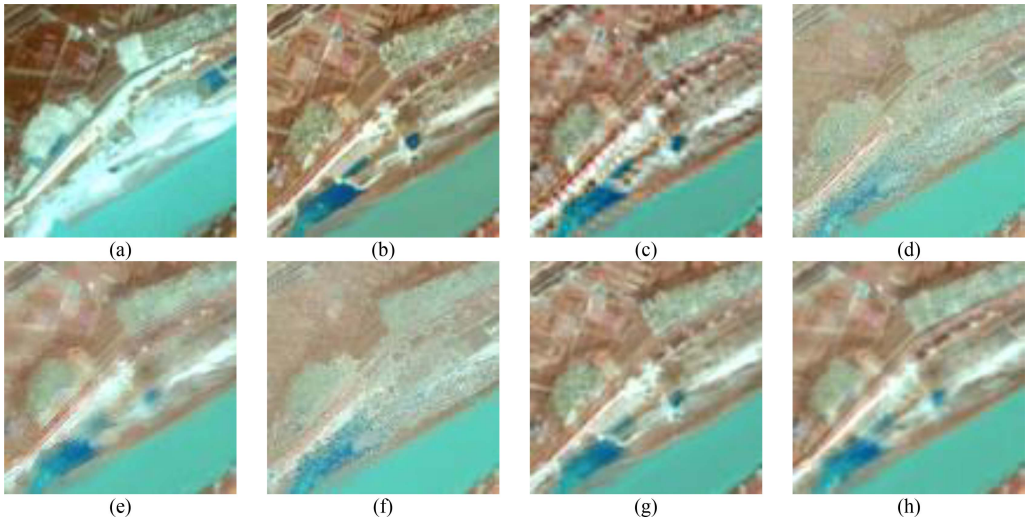



Fig. 5. Zoomed images of the area marked in Fig. 4 extracted from (a) actual Sentinel-2 image observed on December 16, 2018, (b) Sentinel-2 image observed on March 16, 2018, and the fused images generated by (c) ATPRK, (d) STARFM, (e) FSDAF, (f) fit-FC, (g) DSTFN, and (h) MSSF.

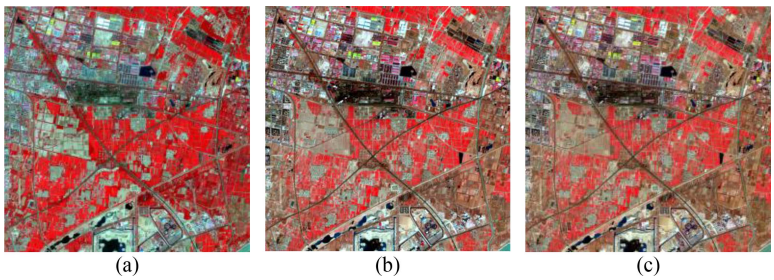



Fig. 6. (a) Sentinel-2 image of Dezhou captured on April 20, 2018, and (b) Landsat-8 image and (c) Sentinel-2 image captured on March 16, 2018.

TABLE I
QUALITY INDICES IN SCENARIO 1

| | SSIM | PSNR | RMSE | CC | SAM |
|---|---|---|---|---|---|
| ATPRK | 0.9111 | 32.2061 | 0.0245 | 0.9444 | 0.1098 |
| STARFM | 0.8950 | 30.8575 | 0.0287 | 0.9285 | 0.1030 |
| FSDAF | 0.9135 | 32.4952 | 0.0237 | 0.9537 | 0.0912 |
| Fit-FC | 0.8962 | 31.1460 | 0.0277 | 0.9347 | 0.1011 |
| DSTFN | 0.9424 | 34.2333 | 0.0194 | 0.9709 | 0.0891 |
| MSSF | 0.9273 | 33.4200 | 0.0213 | 0.9635 | 0.0845 |

missing spatial details. FSDAF produces a fusion result with less reflectance inaccuracy and more structure information compared to STARFM and Fit-FC, but still loses some spatial details. The result of ATPRK, while showing high spectral similarity, suffers from significant over-sharpening. MSSF yields a fusion result that is closer to the actual image at the target date from both spatial and spectral points of view with clear edges and textures. DSTFN also produces an excellent visual result.

The quality indices in Table I are consistent with the visual comparison. The proposed MSSF produces the highest SSIM, PSNR, and CC, and the smallest RMSE and SAM among all machine learning-based methods tested. The result obtained by DSTFN is superior to that of MSSF. However, considering that training is performed on the data set in the same region, and that additional bands were utilized, i.e., the 15-m panchromatic band of Landsat-8 data and the 20-m bands of Sentinel-2 data. We only present this result as a reference. Therefore, the proposed MSSF has outstanding performance in addressing land cover changes from spatial and spectral point of view.

### *B. Experiment 2-Images From a Region Experiencing Land Cover and Phenological Changes*

In this experiment, the Sentinel-2 image on April 20, 2018 [Fig. 6(a)] and the Landsat-8 image on March 16, 2018 [Fig. 6(b)] were collected as inputs to generate a 10-m image on March 16, 2018, and the Sentinel-2 image collected on March 16, 2018 [Fig. 6(c)] was utilized as a comparison image. Fig. 7 presents the two input images, the actual Sentinel-2 image on March 16, 2018 and the fused images obtained by different methods. The zoomed images of the subarea marked in Fig. 7 are shown in Fig. 8.

This heterogeneous region mainly experienced phenological changes, some of which experienced land cover changes. The difficulty of fusion is mainly on capturing land cover changes while recreating the fine structure of the dense construction in the urban area. ATPRK manages to capture time changes in detail, but there is obvious jaggedness and mosaicking in the fused image, which causes low clarity. STARFM, FSDAF, and Fit-FC all generate well-defined edges and structures, but

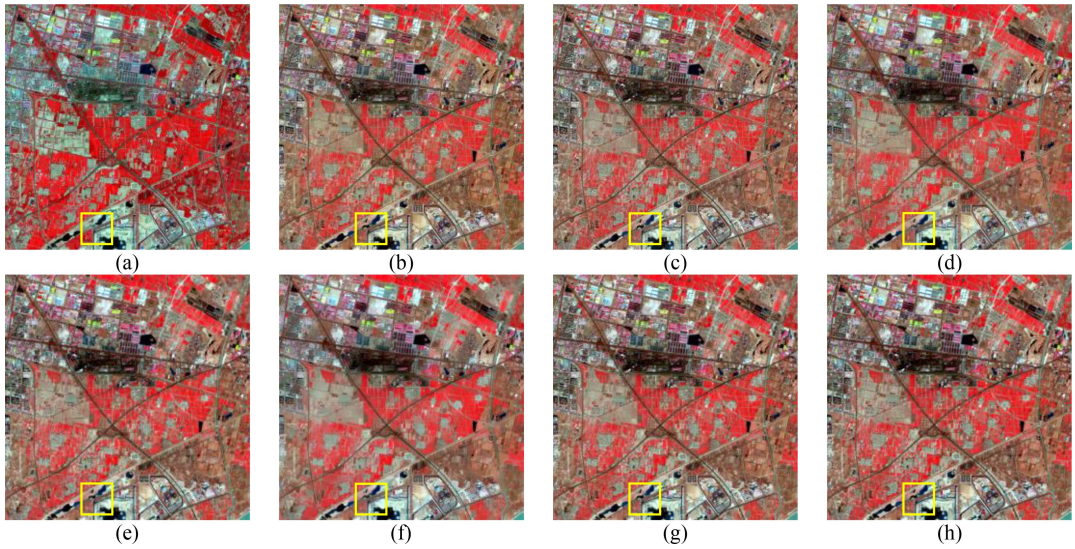



Fig. 7. (a) Actual Sentinel-2 image at the Dezhou study area observed on April 20, 2018. (b) Sentinel-2 image observed on March 16, 2018, and the fused images generated by (c) ATPRK, (d) STARFM, (e) FSDAF, (f) fit-FC, (g) DSTFN, and (h) MSSF.

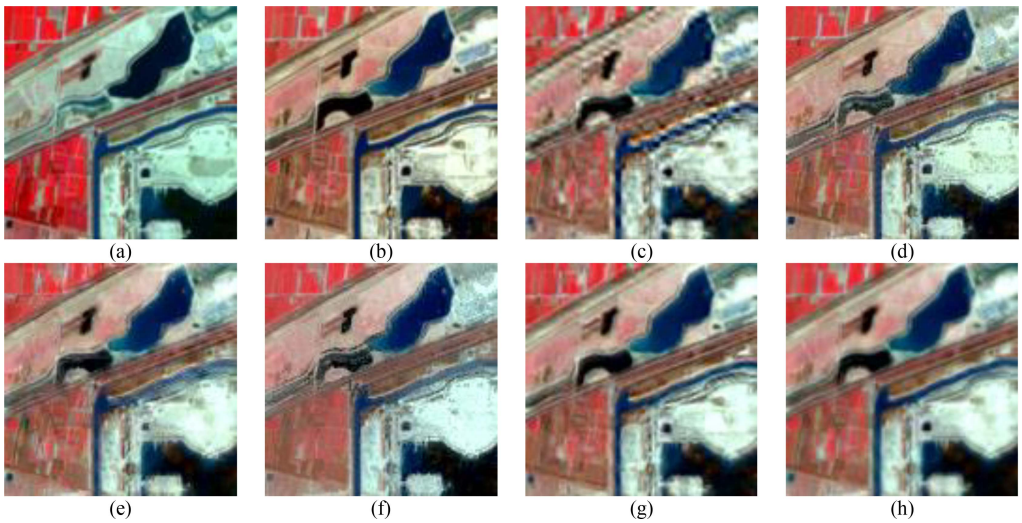



Fig. 8. Zoomed images of the area marked in Fig. 7 extracted from (a) actual Sentinel-2 image observed on April 20, 2018, (b) Sentinel-2 image observed on March 16, 2018, and the fused images generated by (c) ATPRK, (d) STARFM, (e) FSDAF, (f) fit-FC, (g) DSTFN, and (h) MSSF.

TABLE II
QUALITY INDICES IN SCENARIO 2

| | SSIM | PSNR | RMSE | CC | SAM |
|---|---|---|---|---|---|
| ATPRK | 0.9174 | 31.8639 | 0.0255 | 0.9484 | 0.1048 |
| STARFM | 0.9156 | 31.4843 | 0.0267 | 0.9461 | 0.0929 |
| FSDAF | 0.9144 | 31.7719 | 0.0258 | 0.9501 | 0.0985 |
| Fit-FC | 0.9111 | 31.4246 | 0.0268 | 0.9485 | 0.0909 |
| DSTFN | 0.9481 | 33.7874 | 0.0204 | 0.9710 | 0.0787 |
| MSSF | 0.9353 | 33.3073 | 0.0216 | 0.9669 | 0.0777 |

could not capture temporal variations in the waters well. MSSF reproduces land cover changes relatively accurately, and the fusion image appears more uniform and maintains most of the boundary information.

The quality indices of all the compared methods in this experiment are demonstrated in Table II. It can be summarized that the fused result of MSSF has higher SSIM, PSNR, and CC, and smaller RMSE and SAM compared with other machine learning-based methods, indicating that MSSF has the ability to produce accurate fusion results.

### C. Experiment 3-Images From a Region Experiencing Phenological Changes

In this experiment, the Sentinel-2 image on February 26, 2019 [Fig. 9(a)] and the Landsat-8 image on January 7, 2019 [Fig. 9(b)] were utilized as inputs to generate a 10-m image on January 7, 2019. The actual Sentinel-2 image on January 7, 2019 [Fig. 9(c)] was used for validation. Fig. 10 presents the two inputs, the actual Sentinel-2 image on January 7, 2019 and the fused images obtained by different methods. The zoomed images of the subarea marked in Fig. 10 are shown in Fig. 11.

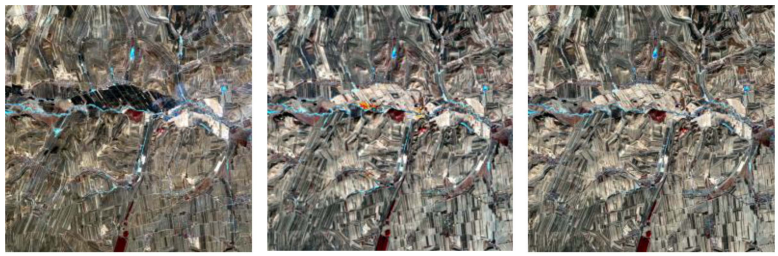



Fig. 9. (a) Sentinel-2 image of Hailar captured on February 26, 2019, (b) Landsat-8 image, and (c) Sentinel-2 image captured on January 7, 2019.

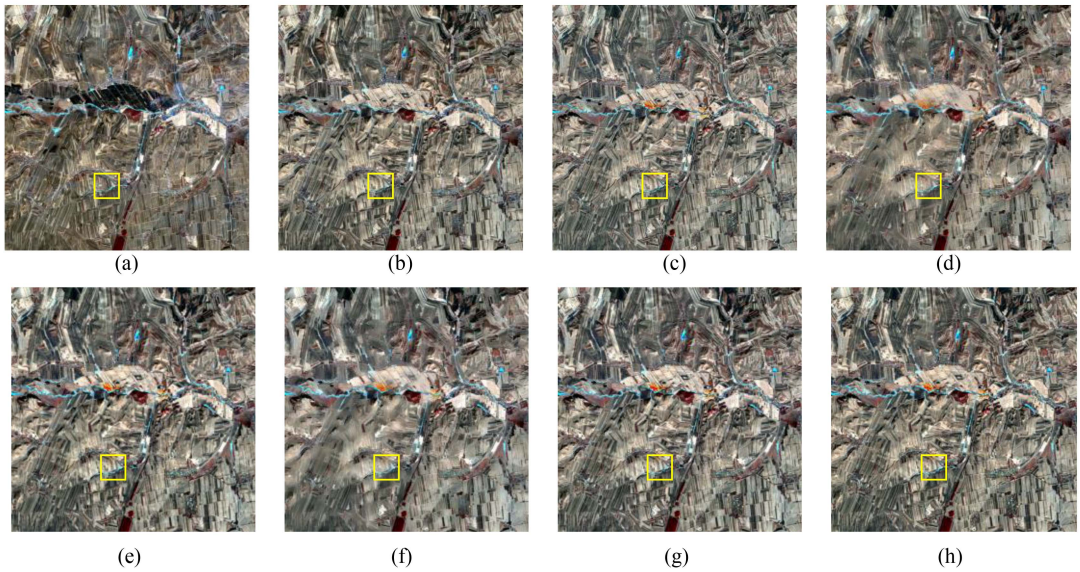



Fig. 10. (a) Actual Sentinel-2 image at the Hailar study area observed on February 26, 2019, (b) Sentinel-2 image observed on January 7, 2019, and the fused images generated by (c) ATPRK, (d) STARFM, (e) FSDAF, (f) fit-FC, (g) DSTFN, and (h) MSSF.

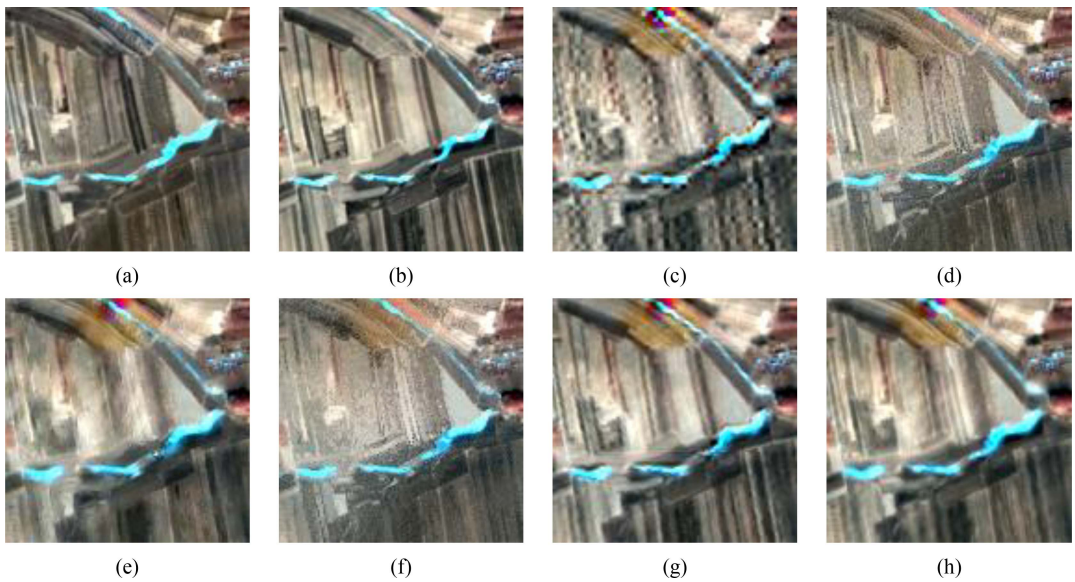



Fig. 11. Zoomed images of the area marked in Fig. 10 extracted from (a) actual Sentinel-2 image observed on February 26, 2019, (b) Sentinel-2 image observed on January 7, 2019, and the fused images generated by (c) ATPRK, (d) STARFM, (e) FSDAF, (f) fit-FC, (g) DSTFN, and (h) MSSF.

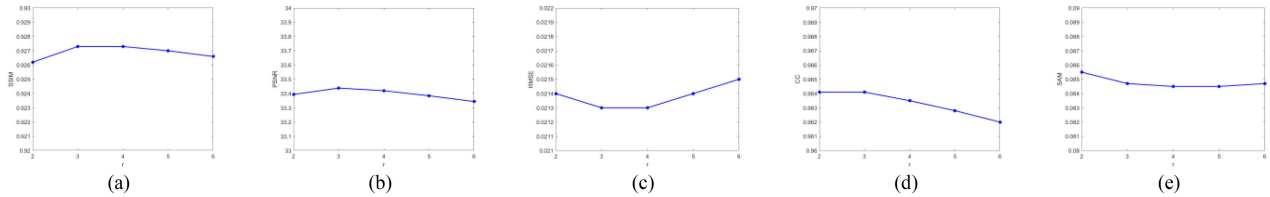



Fig. 12. (a) SSIM, (b) PSNR, (c) RMSE, (d) CC, and (e) SAM evaluating of MSSF fusion results at different $r$ in Scenario 1.

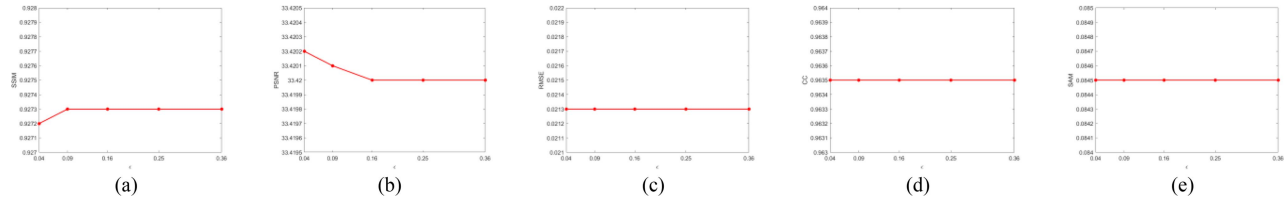



Fig. 13. (a) SSIM, (b) PSNR, (c) RMSE, (d) CC, and (e) SAM evaluating of MSSF fusion results at different $\epsilon$ in Scenario 1.

TABLE III
QUALITY INDICES IN SCENARIO 3

| | SSIM | PSNR | RMSE | CC | SAM |
|---|---|---|---|---|---|
| ATPRK | 0.9239 | 30.5877 | 0.0296 | 0.9385 | 0.1037 |
| STARFM | 0.9359 | 30.7406 | 0.0290 | 0.9439 | 0.0775 |
| FSDAF | 0.9406 | 31.6991 | 0.0260 | 0.9558 | 0.0818 |
| Fit-FC | 0.9352 | 30.9285 | 0.0284 | 0.9474 | 0.0778 |
| DSTFN | 0.9458 | 32.0479 | 0.0250 | 0.9599 | 0.0816 |
| MSSF | 0.9558 | 33.0511 | 0.0223 | 0.9694 | 0.0821 |

The region which consists of extensive farmlands mainly undergoes phenological changes. This data set was selected to determine the ability of MSSF to reproduce the texture of farmlands. STARFM and Fit-FC provide results generally similar to the actual Sentinel-2 image, but fail to accurately calculate the pixel values of each land patch and capture subtle details. In contrast, ATPRK, FSDAF, and MSSF can capture reflectance changes perfectly. However, the results of ATPRK and FSDAF still suffer from fuzzy details and boundaries. The fused image obtained by MSSF is the closer to the actual Sentinel-2 image.

The quality indices for this experiment are given in Table III. Although the proposed MSSF does not perform better in terms of SAM index because the multiscale SSIF possibly affects the spectral similarity when transferring the details of the Sentinel-2 image at the reference date, it performs the best on the remaining indices, i.e., SSIM, PSNR, RMSE, and CC. In addition, the result obtained by DSTFN without training on the data set in the same region is overall inferior to that of MSSF. It can be observed that MSSF is more robust.

## IV. DISCUSSION

### A. Parameter Selection

Selecting appropriate parameters helps to utilize the proposed method adequately. There are four main parameters to be determined in the proposed MSSF, i.e., patch radius ($r$), sharpening-smoothing factor ($\kappa$), sharpening gain ($\epsilon$), and scale ($s$). Through preliminary experiments, we found that $r$, $\epsilon$ and $s$ had a minor effect on the fusion results. The results of experiments in Scenario 1 are shown in line graphs in Figs. 12–14. Similar results were achieved in other scenarios. Therefore, they were fixed to $r = 4$, $\epsilon = 0.4^2$ and $s = 1$ [52]. As shown in (10) and (11), $\kappa$is a key parameter controlling smoothing ($0 \leq \kappa < 1$) or sharpening ($\kappa > 1$), which is set between 0 and 1 in MSSF, since only the smoothing function is employed. Specifically, the closer $\kappa$ gets to 0, the more pronounced the smoothing effect is and the detailed image contains more information. Conversely, the closer $\kappa$ gets to 1, the weaker the smoothing effect is and the sparser the detailed image is. However, simply assuming that a small $\kappa$ leads to a good fusion result is wrong, because the Sentinel-2 image and the Landsat-8 image are captured on different dates, and transferring too much information may cause artifacts. It is therefore necessary to discuss the selection of $\kappa$. Experiments with different values of $\kappa$ are conducted on the same three data sets as those in Section III. The SSIM, PSNR, RMSE, CC, and SAM are computed and demonstrated in line graphs in Figs. 15–17.

In Scenario 1, the PSNR, RMSE, CC, and SAM are optimized at $\kappa = 0.1$. The SSIM reaches its maximum at $\kappa = 0.01$. Considering all quality indices together, $\kappa = 0.1$ is chosen for fusion.

In Scenario 2, the PSNR, RMSE, and CC achieve optimum at $\kappa = 0.3$, the SSIM achieves optimum at $\kappa = 0.2$, while the SAM reaches its minimum at $\kappa = 0.5$. $\kappa$ is determined to be 0.3, where the SSIM and SAM are closer to their optimal values.

In Scenario 3, most quality indices are optimized at $\kappa = 0.1$, i.e., the SSIM, PSNR, RMSE, and CC. The SAM reaches its maximum at $\kappa = 0.5$. Considering the five quality indices together, $\kappa = 0.1$ is chosen for fusion.

We can find some patterns from the abovementioned three experiments. In a single experiment, the values of $\kappa$ for each index to reach optimum are usually close to each other. When $\kappa$ is small, the SSIM is usually better, the visual outcome is clearer and the edges are more defined, while the SAM gets worse. This is probably because when the $\kappa$ value is small, SSIF extracts and transfers considerable details from the Sentinel-2 image, which may synchronously lead to the migration of inaccurate

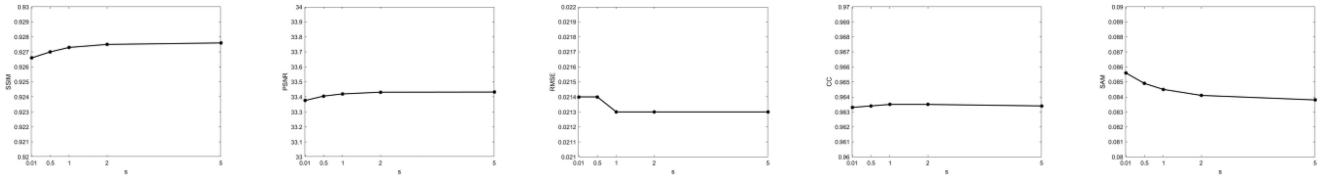



Fig. 14. (a) SSIM, (b) PSNR, (c) RMSE, (d) CC, and (e) SAM evaluating of MSSF fusion results at different $s$ in Scenario 1.

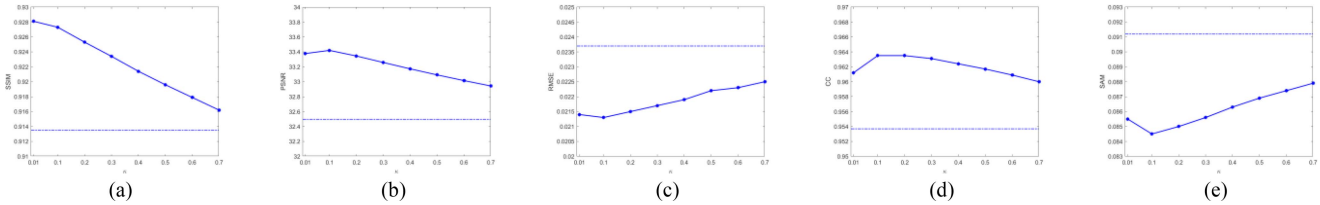



Fig. 15. (a) SSIM, (b) PSNR, (c) RMSE, (d) CC, and (e) SAM evaluating of MSSF fusion results at different $\kappa$ in Scenario 1.

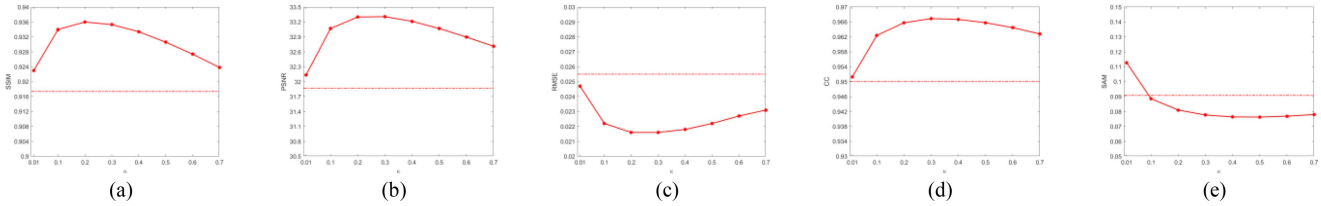



Fig. 16. (a) SSIM, (b) PSNR, (c) RMSE, (d) CC, and (e) SAM evaluating of MSSF fusion results at different $\kappa$ in Scenario 2.

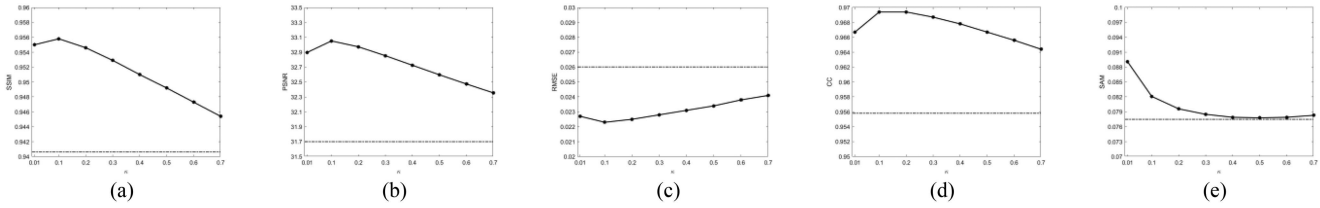



Fig. 17. (a) SSIM, (b) PSNR, (c) RMSE, (d) CC, and (e) SAM evaluating of MSSF fusion results at different $\kappa$ in Scenario 3.

spectral information. It can be reasonably hypothesized that when the temporal variation between $t_0$ and $t_p$ is small, it is recommended that a smaller $\kappa$ value be selected, whereas if the temporal variation is large, a larger $\kappa$ value would be more appropriate.

The optimal values of the quality indices of other compared methods are labeled with dotted lines in the line graphs. Since we mainly study the performance of MSSF among machine learning-based methods, the labeled lines represent the optimal values of machine learning-based methods. It can be noticed that the MSSF still achieves better results than the comparison methods in most of the indices when the $\kappa$ value changes. In addition, when $\kappa$ varies between 0.1 and 0.4, each experiment gives good results and the variation of each index is not drastic. Therefore, it can be prioritized to regulate in this interval.

### B. Ablation Experiments

Ablation experiments are conducted to demonstrate the effectiveness of each proposed module. On the one hand, the multiscale SSIF is kept constant and the preprocessing step is set to bilinear interpolation, TPS interpolation, TPS interpolation, and morphological opening-closing algorithm and the proposed preprocessing, respectively. The results of experiments conducted on the same three data sets as those in Section III are shown in Tables IV–VI. As shown in the table, with the addition of preprocessing modules, the quality indices are progressively improved, which indicates that each preprocessing module has the ability to enhance the performance of the proposed MSSF. Meanwhile, MSSF without preprocessing achieves better performance than all other compared methods in Section III, which proves that multiscale SSIF is an equally valid and important part of the proposed MSSF.

On the other hand, in order to further validate the effectiveness of the proposed multiscale SSIF, the preprocessing is kept constant and the multiscale SSIF is replaced with single-scale SSIF and multiscale guided filter, respectively. The results of experiments in three scenarios are displayed in the last two rows of Tables IV–VI. It can be noticed that the introduction of multiscale processing and SSIF also enhances the fusion effect.

TABLE IV
QUALITY INDICES OF ABLATION EXPERIMENTS IN SCENARIO 1

| | SSIM | PSNR | RMSE | CC | SAM |
|---|---|---|---|---|---|
| Bilinear interpolation | 0.9208 | 32.9108 | 0.0226 | 0.9576 | 0.0886 |
| TPS interpolation | 0.9242 | 33.1573 | 0.0220 | 0.9604 | 0.0860 |
| TPS interpolation and morphological processing | 0.9266 | 33.3228 | 0.0216 | 0.9615 | 0.0848 |
| MSSF | **0.9273** | **33.4200** | **0.0213** | **0.9635** | **0.0845** |
| Single-scale SSIF | 0.9255 | 33.3538 | 0.0215 | 0.9638 | 0.0858 |
| Multi-scale guided filter | 0.9199 | 32.5988 | 0.0234 | 0.9489 | 0.1079 |

TABLE V
QUALITY INDICES OF ABLATION EXPERIMENTS IN SCENARIO 2

| | SSIM | PSNR | RMSE | CC | SAM |
|---|---|---|---|---|---|
| Bilinear interpolation | 0.9308 | 32.9426 | 0.0225 | 0.9632 | 0.0821 |
| TPS interpolation | 0.9328 | 33.0948 | 0.0221 | 0.9644 | 0.0811 |
| TPS interpolation and morphological processing | 0.9343 | 33.2132 | 0.0218 | 0.9649 | 0.0810 |
| MSSF | **0.9353** | **33.3073** | **0.0216** | **0.9669** | **0.0777** |
| Single-scale SSIF | 0.9325 | 33.1092 | 0.0221 | 0.9662 | 0.0777 |
| Multi-scale guided filter | 0.8856 | 29.6829 | 0.0328 | 0.9236 | 0.1868 |

TABLE VI
QUALITY INDICES OF ABLATION EXPERIMENTS IN SCENARIO 3

| | SSIM | PSNR | RMSE | CC | SAM |
|---|---|---|---|---|---|
| Bilinear interpolation | 0.9462 | 32.0303 | 0.0250 | 0.9587 | 0.0867 |
| TPS interpolation | 0.9546 | 32.8683 | 0.0227 | 0.9673 | 0.0846 |
| TPS interpolation and morphological processing | 0.9554 | 32.9616 | 0.0225 | 0.9678 | 0.0865 |
| MSSF | **0.9558** | **33.0511** | **0.0223** | **0.9694** | **0.0821** |
| Single-scale SSIF | 0.9542 | 32.9071 | 0.0226 | 0.9693 | 0.0788 |
| Multi-scale guided filter | 0.9417 | 31.5727 | 0.0264 | 0.9525 | 0.1174 |

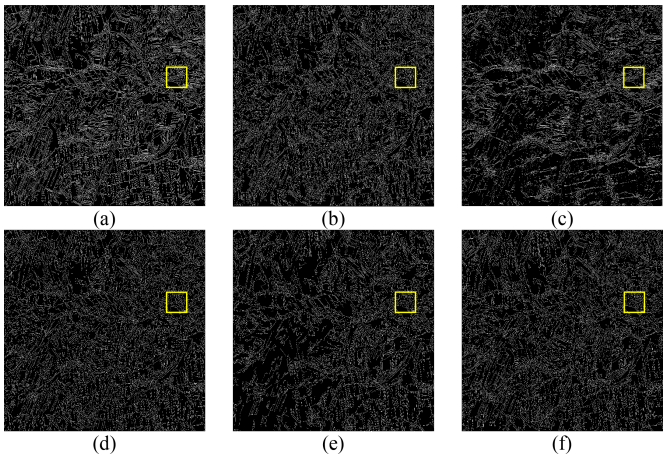



Fig. 18. Boundary maps extracted from the green band of (a) Sentinel-2 image of Hailar captured on January 7, 2019 and the fusion results of (b) ATPRK, (c) STARFM, (d) FSDAF, (e) fit-FC, and (f) MSSF.

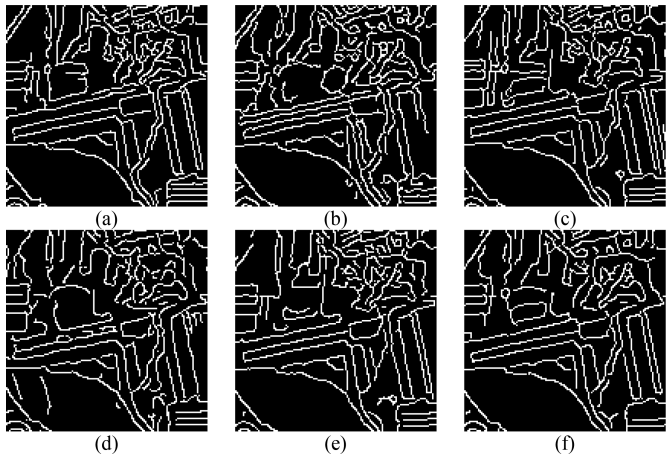



Fig. 19. Zoomed images of the area marked in Fig. 15 extracted from boundary maps of (a) Sentinel-2 image of Hailar captured on January 7, 2019 and the fusion results of (b) ATPRK, (c) STARFM, (d) FSDAF, (e) fit-FC, and (f) MSSF.

### C. Spatial Fidelity Comparison

Since the quality indices evaluate the performance of Landsat-8 and Sentinel-2 image fusion methods mainly based on the perspective of spectral similarity, boundary maps are extracted to test the spatial fidelity of the proposed MSSF. Taking the results of Experiment 3 as an example, the boundary maps obtained by the Canny edge detection operator [53] from green bands of the Sentinel-2 image on January 7, 2019 and the fused images of five methods are shown in Fig. 18. The corresponding zoomed images of the subarea marked in Fig. 18 are shown in Fig. 19.

Compared to the boundary map of the Sentinel-2 image at the target date, the boundaries of the fusion result of ATPRK appear denser, which implies that ATPRK suffers from over-sharpening. The boundary maps of STARFM and Fit-FC are sparser than the actual boundary maps, which indicates a loss of boundary information, especially in the lower left part. Similarly, as can be seen from the zoomed image, the boundaries obtained by all classical STIF-based methods are characterized by omissions and errors to varying degrees. In particular, the boundary maps obtained by FSDAF, although more enriched, are significantly different

TABLE VII
PROCESSING TIME OF SIX METHODS ON THE PREVIOUS DATA SETS

| | Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|---|
| ATPRK | 26.1720s | 26.2843s | 44.3153s |
| STARFM | 1051.1343s | 1023.2061s | 1906.4571s |
| FSDAF | 47.7754s | 35.1190s | 69.1122s |
| Fit-FC | 525.8063s | 522.9659s | 935.5663s |
| DSTFN | 86.4393s | 168.2045s | 211.5476s |
| MSSF | **2.1492s** | **2.0945s** | **3.8112s** |

from the reference boundary map, implying erroneous boundary information that may interfere with the fusion accuracy. Considering the general and local performance, the boundary map of MSSF is the closest to the reference boundary map. It can be concluded that MSSF has the highest spatial fidelity among all compared methods, indicating a strong ability to recover subtle edges and spatial structure.

### D. Processing Time

Processing time is another important measure for assessing the performance of Landsat-8 and Sentinel-2 image fusion. The processing time of all methods in the three experiments in Section III is presented in Table VII. The processing time of DSTFN does not include training time. It can be found that the proposed MSSF achieves the shortest processing time on all three data sets, i.e., 2.1492 s in Experiment 1, 2.0945 s in Experiment 2 and 3.8112 s in Experiment 3. Among other compared methods, STARFM and Fit-FC took longer processing time, since they both apply a weighting function, containing a step of calculating the weights of spectrally similar neighbor pixels for each pixel, which requires large computational effort and processing time. ATPRK and FSDAF operated relatively fast, second only to the proposed MSSF. The data set in Hailar used in Experiment 3 is larger in size than the data sets in Dezhou used in Experiments 1 and 2, and therefore the execution time for all methods increased. In summary, MSSF is an excellent method with low computational complexity and high fusion capability.

## V. CONCLUSION

The study of Landsat-8 and Sentinel-2 image fusion is important as it allows for more frequent acquisition of high-resolution data, thus permitting high-precision monitoring of the rapidly changing land surface. Existing fusion methods cannot satisfy this requirement well because of problems such as spectral distortions, over-sharpening or over-smoothing, and long computation time. MSSF is proposed in order to solve these problems. MSSF implements preprocessing on the input Landsat-8 image at the target date and Sentinel-2 image at the reference date by TPS interpolation with morphological opening-closing algorithm and LOG enhancement algorithm, respectively, then separates the high and low frequency components of the preprocessed image via SSIF. The proposed method extracts and migrates the details from the high frequency components of the preprocessed Sentinel-2 image to the preprocessed Landsat-8 image through multiscale SSIF.

A series of experiments and discussions are conducted to measure the fusion performance of the proposed MSSF in different study areas. The results show that in terms of visual comparison, the proposed MSSF does not only capture phenological and land cover changes accurately, but also reconstructs subtle structures and edges. In terms of quantitative comparison, the proposed MSSF achieves the highest SSIM, PSNR, and CC, and the lowest RMSE and SAM among compared machine learning-based methods on both data sets in Dezhou. On the Hailar data set, except for the SAM value which is in the middle ranking, the proposed MSSF outperforms all compared machine learning-based methods on the rest quality indices. In addition, MSSF is more advantageous in terms of spatial fidelity and computation time. In summary, MSSF is an effective Landsat-8 and Sentinel-2 image fusion method with clear principles, fast computation, and good performance from both spatial and spectral perspectives.

## REFERENCES

[1] W. M. Zhou, Y. J. Liu, S. T. Ata-Ul-Karim, Q. S. Ge, X. Li, and J. F. Xiao, "Integrating climate and satellite remote sensing data for predicting county-level wheat yield in China using machine learning methods," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 111, Jul. 2022, Art. no. 102861.
[2] P. Wang, Z. He, B. Huang, M. D. Mura, H. Leung, and J. Chanussot, "VOGTNet: Variational optimization-guided two-stage network for multispectral and panchromatic image fusion," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2024.3409563.
[3] Y. Wang, L. Gu, X. Li, F. Gao, T. Jiang, and R. Ren, "An improved spatiotemporal fusion algorithm for monitoring daily snow cover changes with high spatial resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5413617.
[4] Y. Chen, Z. Zhang, and F. L. Tao, "Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data," *Eur. J. Agronomy*, vol. 101, pp. 163–173, Nov. 2018.
[5] P. Wang et al., "Low-rank tensor completion pansharpening based on haze correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5405720.
[6] H. Gokon, "Towards an application of remote sensing technology for decision making during natural disaster," in *Proc. Adv. Hum. Side Service. Eng.*, 2020, pp. 271–277.
[7] C. S. Yang and D. W. Shin, "Prototype model of marine pollution monitoring using multiple remote sensing platforms," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 2454–2457.
[8] P. Wang, L. Wang, H. Leung, and G. Zhang, "Super-resolution mapping based on spatial–spectral correlation for spectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2256–2268, Mar. 2021.
[9] G. Q. Zhang and S. Q. Duan, "Lakes as sentinels of climate change on the Tibetan Plateau," *All Earth*, vol. 33, no. 1, pp. 161–165, Jan. 2021.
[10] Z. Yin et al., "Spatiotemporal fusion of land surface temperature based on a convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1808–1822, Feb. 2021.
[11] H. Shirmard, E. Farahbakhsh, R. D. Müller, and R. Chandra, "A review of machine learning in processing remote sensing data for mineral exploration," *Remote Sens. Environ.*, vol. 268, Jan. 2022, Art. no. 112750.
[12] Y. H. Ke, J. Im, J. Lee, H. L. Gong, and Y. Ryu, "Characteristics of Landsat 8 OLI-derived NDVI by comparison with multiple satellite sensors and in-situ observations," *Remote Sens. Environ.*, vol. 164, pp. 298–313, Jul. 2015.
[13] E. C. Rodríguez-Garlito, A. Paz-Gallardo, and A. Plaza, "Mapping invasive aquatic plants in Sentinel-2 images using convolutional neural networks trained with spectral indices," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2889–2899, 2023.
[14] A. Kallel, M. D. Mura, S. Fakhfakh, and N. B. Romdhane, "Physics-based fusion of Sentinel-2 and Sentinel-3 for higher resolution vegetation monitoring," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5403317.
[15] C. Cavallo, M. N. Papa, M. Gargiulo, G. Palau-Salvador, P. Vezza, and G. Ruello, "Continuous monitoring of the flooding dynamics in the Albufera wetland (Spain) by Landsat-8 and Sentinel-2 datasets," *Remote Sens.*, vol. 13, no. 17, Sep. 2021, Art. no. 3525.

[16] T. X. Wu, Y. T. Zhao, S. D. Wang, H. J. Su, Y. Y. Yang, and D. Z. Jia, "Improving the accuracy of fractional evergreen forest cover estimation at subpixel scale in cloudy and rainy areas by harmonizing Landsat-8 and Sentinel-2 time-series data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3373–3385, 2021.
[17] R. Sedona, C. Paris, G. Cavallaro, L. Bruzzone, and M. Riedel, "A high-performance multispectral adaptation GAN for harmonizing dense time series of Landsat-8 and Sentinel-2 images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10134–10146, 2021.
[18] A. Lessio, V. Fissore, and E. Borgogno-Mondino, "Preliminary tests and results concerning integration of Sentinel-2 and Landsat-8 OLI for crop monitoring," *J. Imag.*, vol. 3, no. 4, p. 49, Dec. 2017.
[19] A. Runge and G. Grosse, "Comparing spectral characteristics of Landsat-8 and Sentinel-2 same-day data for Arctic-Boreal regions," *Remote Sens.*, vol. 11, no. 14, Jul. 2019, Art. no. 1730.
[20] E. Mandanici and G. Bitelli, "Preliminary comparison of Sentinel-2 and Landsat 8 imagery for a combined use," *Remote Sens.*, vol. 8, no. 12, Dec. 2016, Art. no. 1014.
[21] Q. Wang et al., "Fusion of Landsat 8 OLI and Sentinel-2 MSI data," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3885–3899, Jul. 2017.
[22] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.
[23] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.*, vol. 172, pp. 165–177, 2016.
[24] H. Guan, Y. Su, T. Hu, J. Chen, and Q. Guo, "An object-based strategy for improving the accuracy of spatiotemporal satellite imagery fusion for vegetation-mapping applications," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 2927.
[25] J. Wu, Q. Cheng, H. Li, S. Li, X. Guan, and H. Shen, "Spatiotemporal fusion with only two remote sensing images as input," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6206–6219, 2020.
[26] Z. Shao, J. Cai, P. Fu, L. Hu, and T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," *Remote Sens. Environ.*, vol. 235, 2019, Art. no. 111425.
[27] S. P. Xiong, S. H. Du, X. Y. Zhang, S. Ouyang, and W. H. Cui, "Fusing Landsat-7, Landsat-8 and Sentinel-2 surface reflectance to generate dense time series images with 10m spatial resolution," *Int. J. Remote Sens.*, vol. 43, no. 5, pp. 1630–1654, Mar. 2022.
[28] J. Wu, L. Lin, T. Li, Q. Cheng, C. Zhang, and H. Shen, "Fusing Landsat 8 and Sentinel-2 data for 10-m dense time-series imagery using a degradation-term constrained deep network," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 108, 2022, Art. no. 102738.
[29] Q. Cheng, R. X. Xie, J. A. Wu, and F. Ye, "Deep learning-based spatiotemporal fusion architecture of Landsat 8 and Sentinel-2 data for 10 m series imagery," *Remote Sens.*, vol. 16, no. 6, Mar. 2024, Art. no. 102738.
[30] D. Vaithiyanathan and K. Sudalaimuthu, "Area-to-point regression Kriging approach fusion of Landsat 8 OLI and Sentinel 2 data for assessment of soil macronutrients at Anaimalai, Coimbatore," *Environ. Monit. Assessment*, vol. 194, no. 12, p. 916, Dec. 2022.
[31] M. S. Dhillon, T. Dahms, C. Kuebert-Flock, I. Steffan-Dewenter, J. Zhang, and T. Ullmann, "Spatiotemporal fusion modelling using STARFM: Examples of Landsat 8 and Sentinel-2 NDVI in Bavaria," *Remote Sens.*, vol. 14, no. 3, p. 677, Feb. 2022.
[32] Y. Xu, B. Huang, Y. Xu, K. Cao, C. Guo, and D. Meng, "Spatial and temporal image fusion via regularized spatial unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1362–1366, Jun. 2015.
[33] G. Deng, F. Galetto, M. Alnasrawi, and W. Waheed, "A guided edge-aware smoothing-sharpening filter based on patch interpolation model and generalized gamma distribution," *IEEE Open J. Signal Process.*, vol. 2, pp. 119–135, 2021.
[34] O. Dubrule, "Comparing splines and kriging," *Comput. Geosci.*, vol. 10, no. 2-3, pp. 327–338, 1984.
[35] Y. L. Zhu, C. Huang, and Z. H. Xu, "Image denoising algorithm based on the median morphological filter," in *Proc. 7th World Congr. Intell. Control Automat.*, 2008, pp. 3985–3989.
[36] M. Iwanowski, "Image contrast enhancement based on Laplacian-of-Gaussian filter combined with morphological reconstruction," in *Proc. Prog. Comput. Recognit. Syst.*, 2020, pp. 305–315.
[37] J. Chen, R. Feng, L. Wang, W. Han, and J. Huang, "Multi-level strategy-based spatial information prediction for spatiotemporal remote sensing imagery fusion," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 637–640.
[38] B. Rasti, B. Koirala, P. Scheunders, P. Ghamisi, and R. Gloaguen, "Boosting hyperspectral image unmixing using denoising: Four scenarios," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 3821–3824.
[39] S. Y. Jung, Y. J. Chyung, and P. W. Kim, "Kernel design for real-time denoising implementation in low-resolution images," *J. Real-Time Image Process.*, vol. 16, no. 1, pp. 31–47, 2017.
[40] W. Dong, S. Xiao, X. Xue, and J. Qu, "An improved hyperspectral pan-sharpening algorithm based on optimized injection model," *IEEE Access*, vol. 7, pp. 16718–16729, 2019.
[41] C. Xu, X. P. Du, Z. Z. Yan, J. J. Zhu, S. Xu, and X. T. Fan, "VSDF: A variation-based spatiotemporal data fusion method," *Remote Sens. Environ.*, vol. 283, Dec. 2022, Art. no. 113309.
[42] K. M. He, J. Sun, and X. O. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[43] K. Kobayashi, "On generalized gamma-functions occurring in diffraction theory," *J. Phys. Soc. Jpn.*, vol. 60, no. 5, pp. 1501–1512, May 1991.
[44] L. Jian, X. Yang, W. Wu, A. Ahmad, A. K. Sangaiah, and G. Jeon, "Pan-sharpening using a guided image filter based on dual-scale detail extraction," *J. Ambient Intell. Humanized Comput.*, vol. 15, no. 3, pp. 1849–1863, 2018.
[45] Y. Demir and N. H. Kaplan, "Low-light image enhancement based on sharpening-smoothing image filter," *Digit. Signal Process.*, vol. 138, 2023, Art. no. 104054.
[46] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.
[47] 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.
[48] J. Li, Y. Li, L. He, J. Chen, and A. Plaza, "Spatio-temporal fusion for remote sensing data: An overview and new benchmark," *Sci. China Inf. Sci.*, vol. 63, no. 4, pp. 1–7, 2020.
[49] Q. Cheng, H. Liu, H. Shen, P. Wu, and L. Zhang, "A spatial and temporal nonlocal filter-based data fusion method," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4476–4488, Aug. 2017.
[50] P. Jagalingam and A. V. Hegde, "A review of quality metrics for fused image," in *Proc. Int. Conf. Water Resour., Coastal Ocean Eng.*, 2015, vol. 4, pp. 133–142.
[51] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.
[52] N. H. Kaplan, "Real-world image dehazing with improved joint enhancement and exposure fusion," *J. Vis. Commun. Image Representation*, vol. 90, 2023, Art. no. 103720.
[53] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

**Peng Wang** (Senior Member, IEEE) received the B.E. degree in microelectronics, and the Ph.D. degree in information and communications engineering from the College of Information and Communications Engineering, Harbin Engineering University, Harbin, China, in 2012 and 2018, respectively.

He is currently an Associate Professor and Doctoral Supervisor with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China. He is also a Hong Kong Scholar with the Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong. In 2016, he was a Visiting Ph.D. Student with the Grenoble Images Parole Signals Automatics Laboratory, Grenoble Institute of Technology, Saint Martin d'Hères, France. He has authored 2 books and more than 50 articles. His research interests include remote-sensing imagery processing and machine learning.

Dr. Wang is an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and a Reviewer of more than 20 international journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Remote Sensing of Environment*, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

**Mingxuan Huang** received the B.E. degree in electronic information engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2022. She is currently working toward the M.S. degree with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing.

Her research interests include remote sensing image processing.

**Shupeng Shi** received the B.E. degree in bioengineering from the Central South University of Forestry and Technology, Changsha, China, in 2012.

Since 2012, he has worked successively with the Hunan Geological Environment Monitoring Station and the Hunan Natural Resources Affairs Center. His research interests include engaged in land surveying, remote sensing image processing, and other work.

**Bo Huang** (Member, IEEE) received the Ph.D. degree in remote sensing and mapping from the Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing, China, in 1997.

He is currently a Professor with the Department of Geography, The University of Hong Kong, Hong Kong. His research interests include most aspects of GIScience, specifically the design and development of models and algorithms for unified satellite image fusion, spatiotemporal statistics and multiobjective spatial optimization, and their applications in environmental monitoring and sustainable land use and transportation planning.

Dr. Huang is currently the Asia-Pacific Editor for the *International Journal of Geographical Information Science* (Taylor & Francis) and the Editor-in-Chief for *Comprehensive GIS* (Elsevier), a three-volume GIS sourcebook.

**Bilian Zhou** received the B.E. degree in surveying and mapping engineering from Chongqing University, Chongqing, China, in 2008.

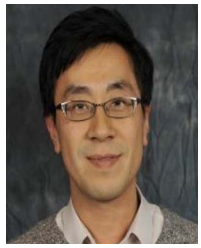
She is currently engaged in aerial photography and remote sensing related work with the Technology Innovation Center for Collaborative Applications of Natural Resources Data in GBA, Ministry of Natural Resources, Guangdong, China.

**Gang Xu** received the B.E. degree in geographic information systems from the Donghua University of Technology, Shanghai, China, in 2003, and the M.S. degree in software engineering from Xiamen University, Xiamen, China, in 2013.

He is currently a Professor with Wenzhou Future City Research Institute, Wenzhou, China, and also with the Zhejiang College of Security Technology, Wenzhou, China. His research interest includes artificial intelligence.

**Liguo Wang** received the M.S. and Ph.D. degrees in signal and information processing from the Harbin Institute of Technology, Harbin, China, in 2002 and 2005, respectively.

From 2006 to 2008, he held a Postdoctoral Research position with the College of Information and Communications Engineering, Harbin Engineering University, Harbin. He is currently a Distinguished Professor with Dalian Minzu University, Dalian, China. He has authored 3 books about hyperspectral image processing and more than 180 papers. His current research interests include remote sensing and machine learning.



**Henry Leung** (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from McMaster University, Hamilton, ON, Canada, in 1991.

Before joining the University of Calgary, he worked with the Department of National Defence (DND) of Canada as a Defense Scientist. His research interests include information fusion, machine learning, IoT, nonlinear dynamics, robotics, and signal and image processing.

Dr. Leung is an Associate Editor for the *IEEE Circuits and Systems Magazine*. He is the topic Editor on "Robotic Sensors" of the *International Journal of Advanced Robotic Systems*. He is also an Editor for the Springer book series on "Information Fusion and Data Science." In addition, he is a Fellow Member of SPIE.