



## Article

# Comparative Analysis of Pixel-Level Fusion Algorithms and a New High-Resolution Dataset for SAR and Optical Image Fusion

Jinjin Li <sup>1</sup>, Jiacheng Zhang <sup>1</sup>, Chao Yang <sup>1</sup>, Huiyu Liu <sup>1</sup>, Yangang Zhao <sup>2</sup> and Yuanxin Ye <sup>1,\*</sup>

<sup>1</sup> Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China; swjtuljj@163.com (J.L.); swjtu\_zjc@163.com (J.Z.); yc18483685462@my.swjtu.edu.cn (C.Y.); huiyu20200131@163.com (H.L.)

<sup>2</sup> The Second Topographic Surveying Brigade of Ministry of Natural Resources, Xian 710054, China; zyg15926@163.com

\* Correspondence: yeyuanxin110@163.com

**Abstract:** Synthetic aperture radar (SAR) and optical images often present different geometric structures and texture features for the same ground object. Through the fusion of SAR and optical images, it can effectively integrate their complementary information, thus better meeting the requirements of remote sensing applications, such as target recognition, classification, and change detection, so as to realize the collaborative utilization of multi-modal images. In order to select appropriate methods to achieve high-quality fusion of SAR and optical images, this paper conducts a systematic review of current pixel-level fusion algorithms for SAR and optical image fusion. Subsequently, eleven representative fusion methods, including component substitution methods (CS), multiscale decomposition methods (MSD), and model-based methods, are chosen for a comparative analysis. In the experiment, we produce a high-resolution SAR and optical image fusion dataset (named YYX-OPT-SAR) covering three different types of scenes, including urban, suburban, and mountain. This dataset and a publicly available medium-resolution dataset are used to evaluate these fusion methods based on three different kinds of evaluation criteria: visual evaluation, objective image quality metrics, and classification accuracy. In terms of the evaluation using image quality metrics, the experimental results show that MSD methods can effectively avoid the negative effects of SAR image shadows on the corresponding area of the fusion result compared with CS methods, while model-based methods exhibit relatively poor performance. Among all of the fusion methods involved in the comparison, the non-subsampled contourlet transform method (NSCT) presents the best fusion results. In the evaluation using image classification, most experimental results show that the overall classification accuracy after fusion is better than that before fusion. This indicates that optical-SAR fusion can improve land classification, with the gradient transfer fusion method (GTF) yielding the best classification results among all of these fusion methods.

**Keywords:** synthetic aperture radar (SAR); optical image; image fusion; image classification



**Citation:** Li, J.; Zhang, J.; Yang, C.; Liu, H.; Zhao, Y.; Ye, Y. Comparative Analysis of Pixel-Level Fusion Algorithms and a New High-Resolution Dataset for SAR and Optical Image Fusion. *Remote Sens.* **2023**, *15*, 5514. <https://doi.org/10.3390/rs15235514>

Academic Editors: Piotr Kaniewski, Stefano Mattoccia and Fabio Tosi

Received: 22 October 2023

Revised: 17 November 2023

Accepted: 24 November 2023

Published: 27 November 2023

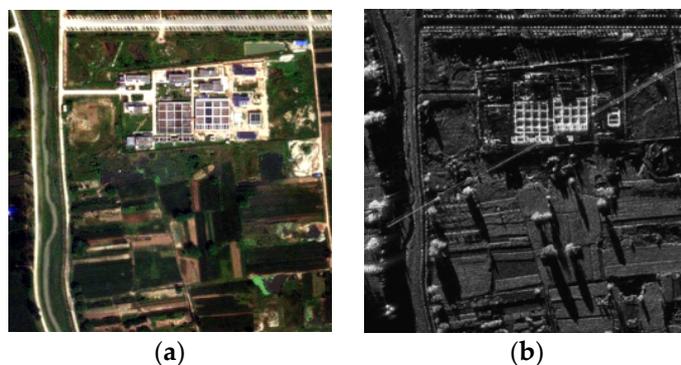


**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the rapid development of different types of sensors that obtain information from the Earth, various remote sensing images have become available for users. Among them, optical images and synthetic aperture radar (SAR) images are two of the most commonly used data in remote sensing applications. SAR images have unique characteristic structure and texture information, making them adaptable for collection at any time without being affected by weather conditions. However, due to the special measurement method of SAR systems (i.e., side-looking imaging), the gray values of SAR images are different from the spectral reflectance of the Earth's surface, which brings difficulties for the interpretation of SAR images in certain scenarios. As is shown in Figure 1, considering that optical images

contain rich spectral information, they can directly reflect the colors and textural details of ground objects. Therefore, optical and SAR images are fused to obtain fusion results containing complementary information, thus enhancing the performance of subsequent remote sensing applications [1].



**Figure 1.** Optical image and SAR image of the same scene: (a) Optical image. (b) SAR image.

According to the stage of data integration, the fusion technology can be divided into three categories: pixel-level, feature-level, and decision-level [2]. Compared with feature-level and decision-level methods, pixel-level fusion methods involve higher computational complexity. Pixel-level fusion methods, despite their higher computational complexity compared to feature-level and decision-level approaches, are widely employed in remote sensing image fusion due to their superior accuracy. These methods have the properties of effective retention of original data, limited information loss, and abundant and accurate image information [3]. As more and more algorithms and their improved versions have been used to fuse optical and SAR images, researchers have compared the performance of these methods for improving ground object interpretation. For instance, Battsengel et al. compared the performance of intensity–hue–saturation (IHS) transform, Brovey transformation, and principal components analysis (PCA) in urban feature enhancement [4]. The analysis revealed that the images transformed through IHS have better characteristics in spectral and spatial separation of different urban levels. However, in a comparative experiment conducted by Sanli et al., IHS showed the worst results [5].

As mentioned above, the performance of the same fusion method can exhibit significant variations across different scenes owing to the special imaging mechanism of SAR and its distinct image content. The fusion quality is affected not only by the quality of the input image, but also by the performance of the fusion method. Accordingly, it is worth considering the selection of a suitable method among many fusion methods and the choice of appropriate metrics for evaluation. In order to compare the performance of various fusion methods objectively, some researchers quantitatively evaluate the effect of fusion methods through objective fusion quality evaluation metrics [6–9], but there are few fusion methods and evaluation metrics involved in experiments, which fail to cover all of the categories of pixel-level fusion methods.

In addition to evaluating fusion quality based on traditional image quality evaluation metrics, it is worth exploring how to use classification accuracy to evaluate fusion quality, particularly in the context of improving image interpretation and land classification. The quality of the input images will affect the accuracy of the classification results [10–13]. Radar can penetrate clouds, rain, snow, haze, and other weather conditions, thus obtaining the reflection information from the target surface. As a result, SAR data can be collected at nearly any time and under any environmental conditions. However, these data are susceptible to speckle noise, thus resulting in poor interpretability, and they lack spectral information. In contrast, optical images contain rich spectral information. In the application of land cover classification, the fusion of optical and SAR data is beneficial to distinguish ground object types that might be indistinguishable due to their similar spectral charac-

teristics. Thus, in order to improve the image classification results, numerous researchers have used SAR and optical image fusion for land cover classification [14–18].

Gaetano et al. deal with the fusion of optical and SAR data for land cover monitoring. Experiments show that the fusion of optical and SAR data can greatly improve the classification accuracy compared with raw data or even multitemporal filtering data [15]. Hu et al. propose a fusion approach for the joint use of SAR and hyperspectral data, which is used for land use classification. The classification results show that the fusion method can improve the classification performance of hyperspectral and SAR data, and it can collect the complementary information of the two datasets well [17]. Kulkarni et al. present a hybrid fusion approach to integrate information from SAR and MS imagery to improve land cover classification [18]. Dabbiru et al. investigate the impact of an oil spill in an ocean area. The main purpose of that study was to apply fusion technology to SAR and optical images and explore the application value of fusion technology in the classification of oil-covered vegetation in coastal zones [19]. However, few researchers take classification accuracy as an evaluation metric to evaluate the performance of different fusion methods.

Optical and SAR image fusion has garnered significant attention owing to the special complementary advantages. However, many existing methods borrow migrations of fusion models from other fields (e.g., optical and infrared images, multi-focused images), with a lack of algorithmic exploration for the study of optical and SAR specificity. In recent years, deep learning has greatly driven the applied research on image fusion, but the studies are mostly focused on specific application scenarios, such as target extraction, cloud removal, land classification, etc. [20–22], in which the algorithms mainly deal with local feature information rather than global pixel information. In most of the latest research articles on pixel-level fusion of optical and SAR images based on deep learning, no specific code files have been published to objectively verify the advantages and disadvantages of the algorithms. Therefore, in this paper, in order to better experimentally verify the algorithms within the field of pixel-level image fusion, several types of traditional algorithms that are well-established and publicly available are selected for comparative analysis.

This paper makes the following three contributions:

1. We systematically review the current pixel-level fusion algorithms for optical and SAR image fusion, and then we select eleven representative fusion methods, including CS methods, MSD methods, and model-based methods for comparison analysis.
2. Based on the evaluation indicators of low-level visual tasks, we combine these with the evaluation indicators of subsequent high-level visual tasks to analyze the advantages and disadvantages of existing pixel-level fusion algorithms.
3. We produce a high-resolution SAR and optical image fusion dataset, including 150 pairs of images of urban, suburban, and mountain settings, which can provide data support for relevant research. The download link for the dataset is <https://github.com/yeyuanxin110/YYX-OPT-SAR> (accessed on 21 January 2023).

This paper extends our early work [23] by adding two datasets, including a self-built high-resolution dataset named YYX-OPT-SAR and a publicly medium-resolution dataset named WHU-OPT-SAR, to evaluate the fusion methods. In order to evaluate the performance of different fusion methods in subsequent classification applications, we also employ classification accuracy as an evaluation criterion to assess the quality of different fusion methods.

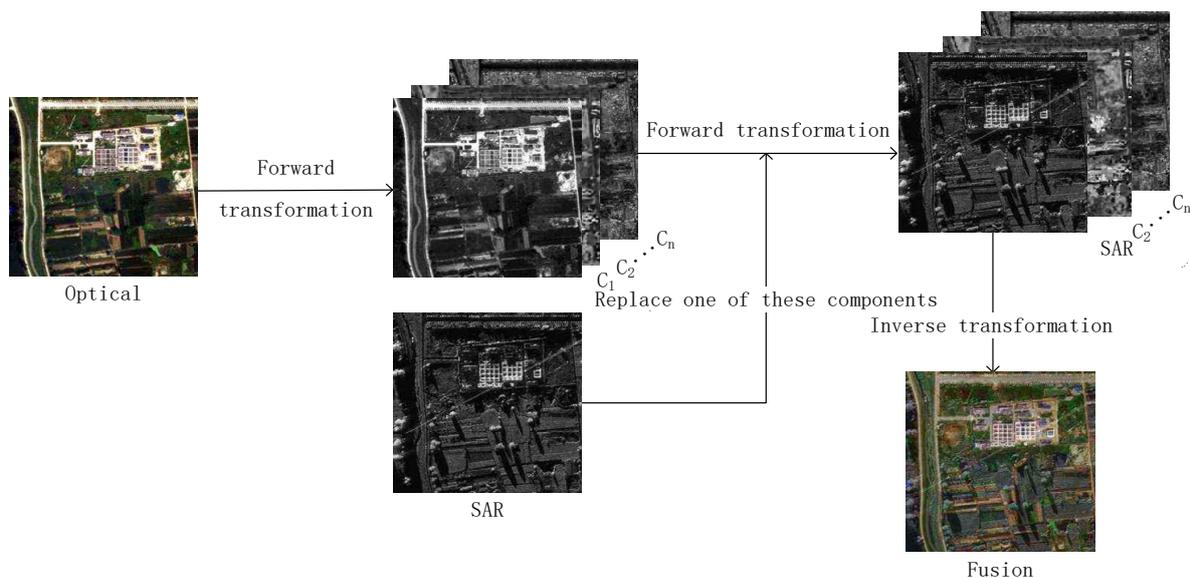
## 2. Pixel-Level Methods of Optical–SAR Fusion

As an important branch of information fusion technology, the pixel-level fusion of images can be traced back to the 1980s. With the increasing maturity of SAR technology, researchers have explored the fusion of optical and SAR images to enhance the performance of remote sensing data across various applications. In the multi-source remote sensing data fusion competition held by the IEEE Geoscience and Remote Sensing Society (IEEE GRSS) in 2020 and 2021, the theme of SAR and multispectral image fusion has been consistently included, which underscores the growing significance of optical and SAR

image fusion in recent years. At present, research on pixel-level fusion algorithms of optical and SAR images based on deep learning remains relatively limited in depth, so the pixel-level fusion algorithms selected in this paper are relatively mature, traditional algorithms. Generally speaking, traditional pixel-level fusion methods can be divided into CS methods, MSD methods, and model-based methods [2]. Because of their different data processing strategies, these three methods have their own advantages and disadvantages in optical and SAR image fusion.

### 2.1. CS Methods

The fusion process of CS methods is shown in Figure 2. CS methods aim to obtain the final image fusion result by replacing a certain component of the positive transformation of the optical image with the SAR image and then applying the corresponding inverse transformation. In this way, the obtained image fusion result incorporates the spectral information from the optical image and the texture information from the SAR image. For instance, Chen et al. utilize the IHS transform to fuse hyperspectral and SAR images. The fusion results not only have a high spectral resolution but also contain the surface texture features of SAR images, which enhances the interpretation of urban surface features [24]. The conventional PCA method is improved by Yin and Jiang, and the fusion result demonstrates better performance in preserving both spatial and spectral contents [25]. Yang et al. use the Gram–Schmidt algorithm to fuse GF-1 images with SAR images and successfully improve the classification accuracy of coastal wetlands by injecting SAR image information into the fusion results [26].



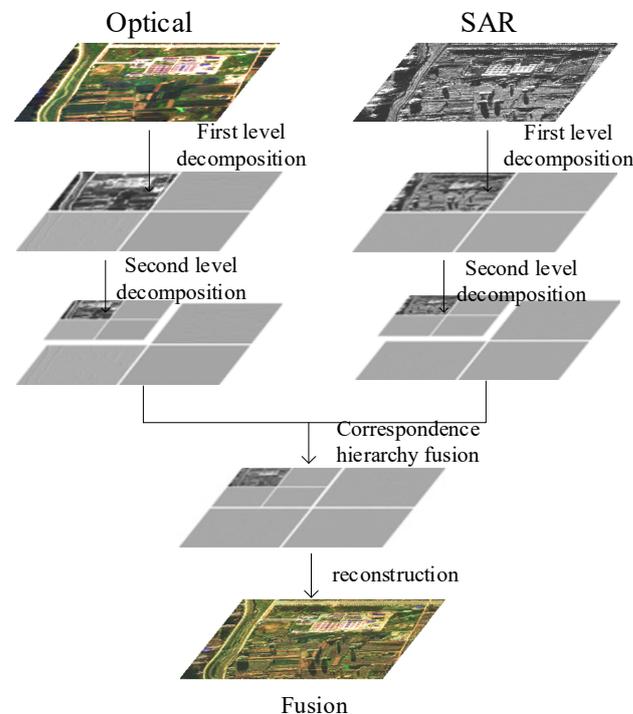
**Figure 2.** Fusion process of the CS method.

With the characteristics of simplicity and low computational complexity, CS methods can obtain fusion results with abundant spatial information in pan-sharpening and other fusion tasks. However, in multi-sensor and multi-modal image fusion, such as SAR–optical image fusion, serious spectral distortions occur frequently in partial areas because of low correlation. Recently, the research on pixel-level fusion algorithms of optical and SAR images has developed toward the multi-scale decomposition method.

### 2.2. MSD Methods

MSD methods divide the original image into the main image and the multilayer detail image according to the decomposition strategy, and each image encapsulates distinct potential information from the original image [27]. While the number of subbands decomposed by different methods varies, these methods share a similar process framework,

which is shown in Figure 3. According to the decomposition strategies, MSD methods can be divided into three categories: wavelet-based methods, pyramid-based methods, and multi-scale geometric analysis (MGA)-based methods [27].



**Figure 3.** Schematic diagram of fusion process of multi-scale decomposition methods.

Kulkarni and Rege use the wavelet transform to fuse SAR and multispectral images, and they apply the activity level measurement method based on local energy to merge detail subbands, which not only enhances spatial information but also avoids spectral distortions [1]. Eltaweel and Helmy apply the Non-subsampled shearlet transform (NSST) for multispectral and SAR image fusion. The fusion rules based on local energy and the dispersion index are used to integrate the low-frequency coefficients decomposed through NSST, and the multi-channel pulse coupled neural network (m-PCNN) is utilized to guide the fusion process of bandpass subbands. The fusion results show good object contour definition and structural details [28].

The primary goal of MSD methods is to extract multiplex features of the input image into different scales of subbands, and thus to realize the optimal selection and integration of diverse pieces of salient information through specifically designed fusion rules. Activity-level measurement and coefficient combination are essential steps in MSD methods. As a critical factor affecting the quality of the fused image, activity-level measurement is used to express the salience of each coefficient and then provide the evaluation criterion and calculation basis for the weight assignment in the coefficient combination process. And the activity-level measurement methods can be divided into three categories: the coefficient-based, window-based, and region-based measures. Equally important are coefficient combination rules, which involve various operations, such as weighted average, maximum value, and consistency verification, that help to control the contribution of different frequency bands to the merging results with predefined or adaptive rules [29].

### 2.3. Model-Based Methods

Model-based fusion methods relate to the fusion of optical and SAR images as an image generation problem. The final fusion result is derived by establishing a mathematical model that describes the mapping relationship from the source image to the fusion result, or by establishing a constraint relationship between the fusion result and the source image.

In addition, in order to enhance the fusion effect, a probability model and a priori constraint can be introduced into the model, albeit at the expense of increased solution complexity. Representative methods within this category include variational model methods and sparse representation (SR) methods. Variational model methods establish an energy functional consisting of different terms based on prior constraint information. The fusion result is obtained by minimizing the energy functional under the premise that the existence of a minimum for the energy functional is proved. On the other hand, the methods based on SR select different linear combinations from overcomplete dictionaries to describe image signals. Yang and Li are pioneers in employing SR for the image fusion task, and they propose an SR-based image fusion method using the sliding window technique [30]. The schematic diagram of the method is shown in Figure 4.

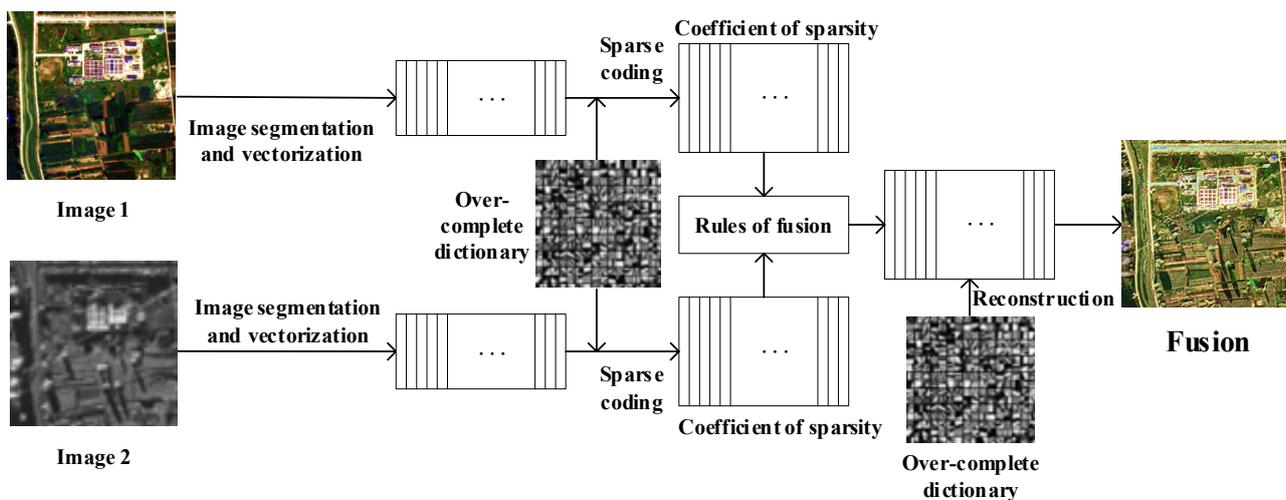


Figure 4. Flow chart of image fusion based on sparse representation.

Wei Zhang and Le Yu introduce the variational model for pan-sharpening into the fusion process of SAR and multispectral images, which obtains the final fusion result by minimizing the energy functional composed of linear combination constraints, color constraints, and geometric constraints. The experiment demonstrates that a variational model-based fusion method is acceptable for SAR and multispectral image fusion in terms of spectral preservation [31]. Additionally, Huang proposes a cloud removal method for optical images based on sparse representation fusion, which uses SAR and low-resolution optical images to provide high-frequency and low-frequency information for reconstructing the cloud occlusion area and achieves good visual effect and radiation consistency [20].

#### 2.4. Method Selection

Generally speaking, the traditional pixel-level fusion methods can be divided into CS methods, MSD methods, and model-based methods. According to the decomposition strategy, MSD methods can be divided into three categories: wavelet-based methods, pyramid-based methods, and multi-scale geometric analysis (MGA) methods. In order to compare the differences between fusion methods in different categories, we choose some classical methods for the following two points in each category. First, there are publicly available algorithms with dependable performance to conduct comparative experiments. Second, they have been used in optical and SAR image fusion fields. Table 1 shows a list of investigated methods.

For MSD methods, the “averaging” rule is selected to merge low-pass bands, while the “max-absolute” rule is employed to merge high-pass MSD bands. Two instances of the “max-absolute” rule are applied, one being the conventional rule and the other incorporating a local window-based consistency verification scheme [32]. These are denoted by the numbers “1” and “2” appended to the corresponding abbreviation, as shown in Table 2, to explore their respective impacts on the final fusion results. The decomposition levels and

decomposition filters presented in Table 3 are chosen according to the research conclusion of the literature [33]. In the sparse representation based on the sliding window, the step size and the window size are fixed to one and eight, respectively [30], the K-means generalized singular value decomposition (K-SVD) algorithm [34] is used to build an overcomplete dictionary, and the orthogonal matching pursuit (OMP) algorithm [35] is utilized for sparse coding. The parameter selection of other methods adopts the recommended values from the corresponding literature.

**Table 1.** Pixel-level fusion methods participating in comparison.

Category		Method
CS		Intensity–Hue–Saturation (IHS) transform [36] Principal Component Analysis (PCA) [37] Gram–Schmidt (GS) transform [38]
	Pyramid-based	Laplacian pyramid (LP) [39] Gradient pyramid (GP) [40]
MSD	Wavelet-based	Discrete wavelet transform (DWT) [41] Dual tree complex wavelet transform (DTCWT) [42]
	MGA	Curvelet transform (CVT) [43] Non-subsampled contourlet transform (NSCT) [44]
Model-based		SR [30] Gradient Transfer Fusion (GTF) [45]

**Table 2.** The fusion rule of the high-frequency component, represented by different serial numbers.

Method	Rule
XX_1	max-absolute
XX_2	“max-absolute” rule with a local window-based consistency verification scheme

(XX denotes a fusion method).

**Table 3.** Filters and number of decomposition layers in MSD methods.

Category	Method	Filters	Levels
Pyramid-based	LP	/	4
	RP	/	4
Wavelet-based	DWT	Daubechies (db1)	4
	DTCWT	First: LeGall 5-3 Other: Q-shift_06	4
MGA	CVT	/	4
	NSCT	Pyramid: pyrexc Orientation: 7–9	{4, 8, 8, 16}

### 3. Evaluation Criteria for Image Fusion Methods

#### 3.1. Visual Evaluation

The visual evaluation is conducted to assess the quality of the fused image based on human observation. Observers judge the spectral fidelity, the visual clarity, and the amount of information in the image according to their subjective feelings. Although visual evaluation has no technical obstacles in implementation and directly reflects the visual quality of images, its reliability is influenced by various factors, such as the observer’s self-experience, display variations in hardware, and ambient lighting conditions, leading to lower reproducibility and stability. Generally, the visual evaluation serves as a supplement in combination with statistical evaluation methods.

### 3.2. Statistical Evaluation

The statistical evaluation of image quality is a fundamental aspect of digital image processing encompassing various fields, such as image enhancement, restoration, and compression. Numerous conventional image quality evaluation metrics, like standard deviation, information entropy, mutual information, and structural similarity, have been widely applied. These metrics can objectively evaluate the quality of fusion results and provide quantitative numerical references for the comparative analysis of fusion methods. In addition to these conventional metrics, researchers have proposed some quality metrics specially designed for image fusion, such as the weighted fusion quality index  $Q_W$  and the edge-dependent fusion quality index  $Q_E$  [46], as well as the objective quality metric based on structural similarity  $Q_Y$  [47].

The objective quality evaluation of image fusion can be carried out in two ways [48]. The first way is to compare the fusion results with a reference image, which is commonly used in pan-sharpening and multi-focus image fusion. However, in multimodal image fusion tasks, such as SAR–optical image fusion, obtaining an ideal reference image is challenging. Therefore, this paper uses the non-reference metrics to objectively evaluate the quality of the fusion image. The fusion results are comprehensively compared from different aspects through nine representative fusion evaluation metrics: information entropy (EN), peak signal-to-noise ratio (PSNR), mutual information (MI), standard deviation (SD), the metric  $Q^{AB/F}$  based on edge information preservation [49], the universal image quality index  $Q_o$  [50], the weighted fusion quality index  $Q_W$ , the edge-dependent fusion quality index  $Q_E$ , the similarity-based image fusion quality index  $Q_Y$ , and the human visual system (HVS)-model-based quality index  $Q_{CB}$  [51]. Based on the different emphases of these evaluation indexes, they can be divided into four categories [52,53]. Table 4 presents the definitions and characteristics of the selected nine quality metrics.

### 3.3. Fusion Evaluation According to Classification

Most of the subsequent applications of remote sensing images focus on image classification and object detection. At present, there have been researches on object detection of remote sensing images [54], but there are few traditional methods. Therefore, this paper chooses image classification as an index to evaluate the performance of image fusion in subsequent applications. In the evaluation of image classification, three classic methods, including Support Vector Machine (SVM) [55], Random Forests (RF) [56], and Convolutional Neural Network (CNN) [57], are used to perform image classification. It is crucial to evaluate the accuracy of classification results. According to the results of the accuracy evaluation, we can judge whether the classification method is accurate and whether the classification degree meets the needs of the subsequent analysis. This information enables us to identify which fusion method yields the best classification result. The commonly used method to evaluate the accuracy of classification results is the confusion matrix, also known as the error matrix. It reflects the correct and incorrect classification of the corresponding classification results of each category in the validation data. The confusion matrix is a square matrix with a side length of  $c$ , where  $c$  is the total number of categories and the values on the diagonal are the number of correctly classified pixels in each category.

Overall accuracy (OA) refers to the ratio of the total number of pixels correctly classified to the total number of pixels in the verified sample. It provides the overall evaluation of the quality of the classification results. User accuracy (UA) represents the degree to which a class is correctly classified in the classification results. It is calculated as the ratio of the number of correctly classified pixels in each class to the total number of pixels sorted into that class by the classifier (the sum of row elements corresponding to that class).

**Table 4.** Definition and significance of nine quality indices.

Category	Definition	Range	Characteristic
Information-theory-based	$EN = -\sum_{i=0}^{L-1} p_i \log_2 p_i,$ L is the number of the gray level and $p_i$ is the normalized histogram of an image	$[0, \log_2 L]$	Reflects the amount of information contained in the image
	$PSNR = 10 \log_{10} \frac{L^2}{MSE}$ where $MSE = \frac{MSE_{AF} + MSE_{BF}}{2}$ , $MSE_{XF} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (X(i, j) - F(i, j))^2$	$[0, +\infty]$	Reflects the distortion
Image-feature-based	$SD = \sqrt{\sum_{i=1}^M \sum_{j=1}^N (F(i, j) - \mu)^2}$	$[0, +\infty]$	Reflects the distribution and contrast of the image
	$Q_{AB/F} = \frac{\sum_{i=1}^N \sum_{j=1}^M (Q^{AF}(i, j)w^A(i, j) + Q^{BF}(i, j)w^B(i, j))}{\sum_{i=1}^N \sum_{j=1}^M (w^A(i, j) + w^B(i, j))}$ $Q^{XF}(i, j) = Q_s^{XF}(i, j)Q_\alpha^{XF}(i, j), Q_g^{XF}(i, j) \text{ and } Q_\alpha^{XF}(i, j)$ denote the edge strength and orientation preservation values at pixel $(i, j)$ ; $w$ is the weighting factor	$[0, 1]$	Evaluates the edge information preserved in the fused image
	$Q_Y = \begin{cases} \lambda(w)SSIM(A, B w) + (1 - \lambda(w))SSIM(B, F w), & SSIM(A, B w) \geq 0.75 \\ \max\{SSIM(A, B w), SSIM(A, B w)\}, & SSIM(A, B w) < 0.75 \end{cases}$ $SSIM(X, Y w)$ is local structural similarity, $\lambda(w)$ is the weighting factor	$[0, 1]$	Reflects the structural similarity between two images
Structural-similarity-based	$Q_0 = (Q_0(A, F) + Q_0(B, F))/2,$ where $Q_0(X, Y) = \frac{2\sigma_{XY}}{\sigma_X^2 + \sigma_Y^2} \cdot \frac{2\mu_X\mu_Y}{\mu_X^2 + \mu_Y^2}$	$[-1, 1]$	Reflects the loss of correlation, luminance distortion, and contrast distortion of the fused image
	$Q_W = \sum_{w \in W} c(w)(\lambda(w)Q_0(A, F w) + (1 - \lambda(w))Q_0(B, F w))$ $c(w)$ is normalized saliency, $\lambda(w)$ is saliency weight, and $Q_0(X, Y w)$ is Wang–Bovik image quality index	$[-1, 1]$	Indicates the amount of salient information transferred into the fused image
	$Q_E = Q_W(A, B, F) \cdot Q_W(\hat{A}, \hat{B}, \hat{F})^\alpha$ $\hat{A}, \hat{B}, \hat{F}$ are edge images of A, B, and F; $\alpha$ is the adjustable parameter	$[-1, 1]$	Evaluates the edge information preserved in the fused image
Human-perception-inspired	$Q_{CB} = \lambda_A(x, y)Q_{AF}(x, y) + \lambda_B(x, y)Q_{BF}(x, y)$ $Q_{AF}, Q_{BF}$ is the information preservation value	$[0, 1]$	Assesses the image quality of the fused image

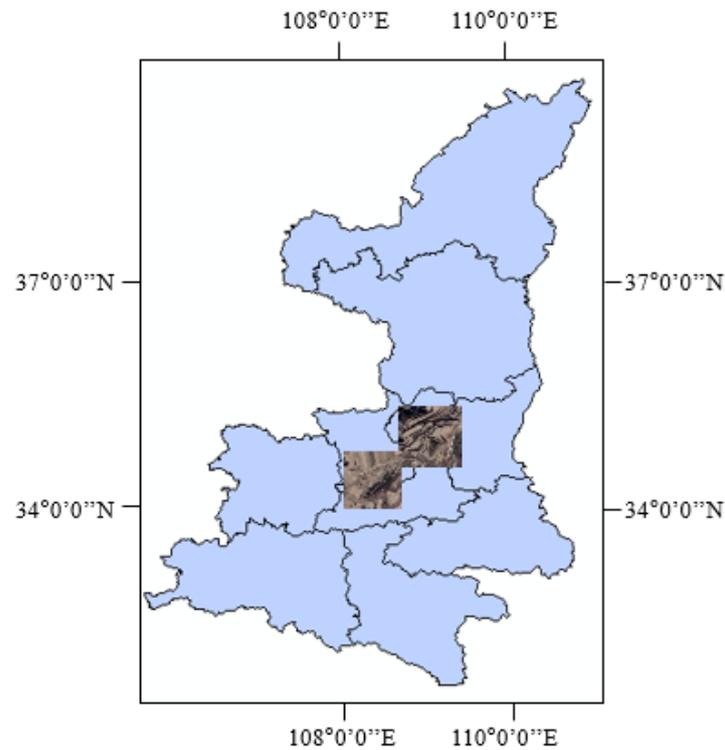
(M, N are the width and height of the image; X and Y represent any image; A and B represent the source image; F represents the fusion result).

#### 4. Datasets

To promote the development of optical–SAR data fusion methods, access to a substantial volume of high-quality optical and SAR image data is essential. SAR and optical images with a sub-meter resolution provide abundant shape structure and texture information of landscape objects. Accordingly, their fusion results are beneficial for accurate image interpretation, and they reflect the specific performance of the used algorithm, thus enabling a persuasive assessment of the fusion methods.

To facilitate research in optical and SAR image fusion technology, we constructed a dataset named YYX-OPT-SAR. This dataset comprises 150 pairs of optical and SAR images covering urban, suburban, and mountain settings, and it is characterized by scene diversity with sub-meter resolution. This dataset can also provide data support for the study of optical and SAR image fusion technology.

The SAR images were collected around Weinan City, Shaanxi Province, China. In order to form high-resolution SAR and optical image pairs, we downloaded optical images of the corresponding areas from Google Earth. The exact location of data collection is shown in Figure 5.

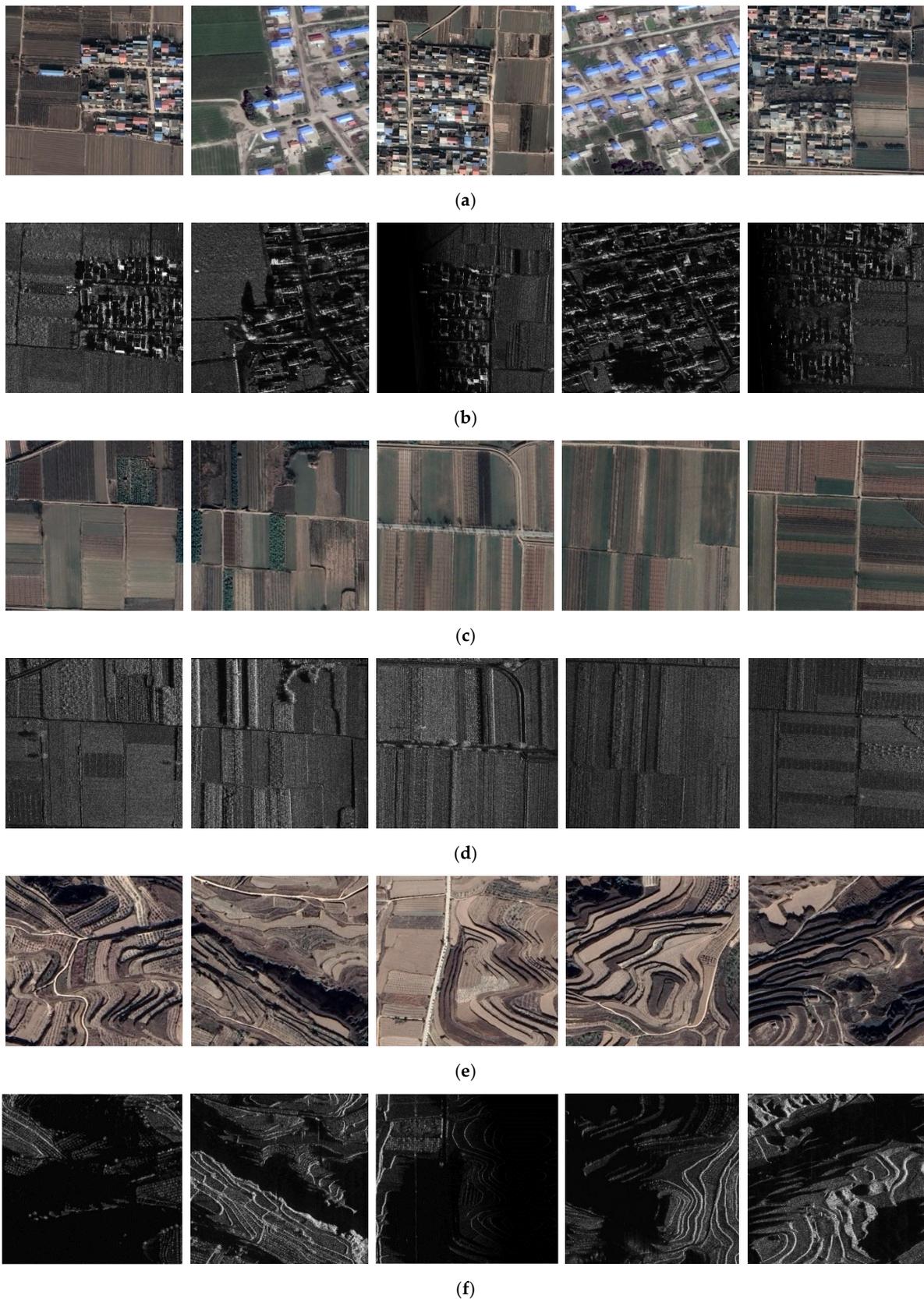


**Figure 5.** YYX-OPT-SAR dataset: Geographic location of the dataset in Shaanxi Province, China.

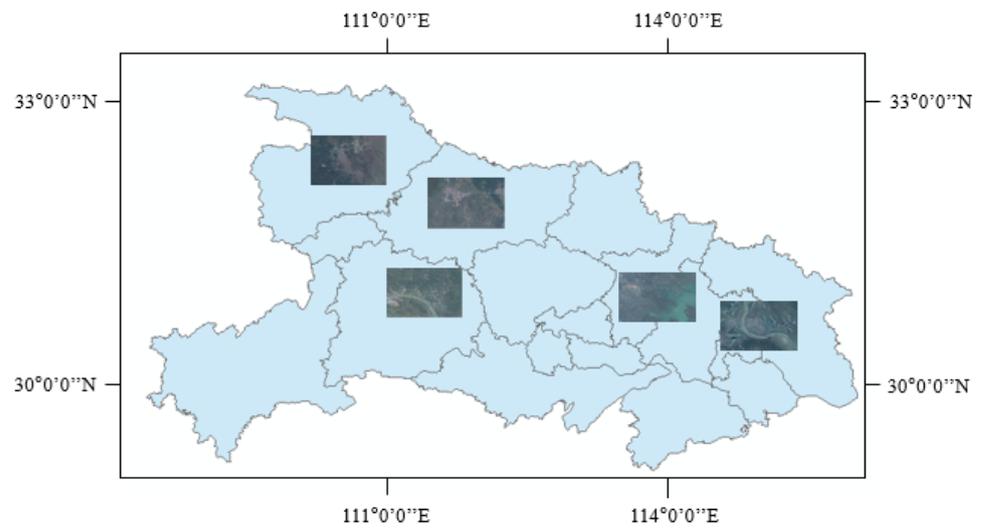
After the acquisition of heterogeneous image data, image registration is required to carry out the subsequent phase of fusion. At present, there are many excellent heterologous image registration methods [58,59]. In this paper, an efficient matching algorithm named channel features of orientated gradients (CFOG) [60] is utilized to achieve high accuracy registration with a match error of less than one pixel. In order to maximize the use of available scenes and ensure that each pair of cropped images can fully express the features of optical and SAR images, so as to facilitate the visual evaluation and the subsequent fusion result analysis, we crop the registered optical and SAR image pairs into non-overlapping image blocks with a size of  $512 \times 512$  pixels. Then, according to different image coverage scenes, we categorize the obtained image pairs into three types: urban, suburban, and mountain. Each type comprises 50 pairs of images, resulting in a total of 150 pairs of images. Some samples are shown in Figure 6.

Another large ground object fusion dataset used in this paper named WHU-OPT-SAR [61] contains medium-resolution optical and SAR images. This dataset, with a resolution of 5 m, covers 51,448.56 square kilometers in Hubei Province, including 100 pairs of  $5556 \times 3704$  (pixel) images. The exact location and coverage of these images on the map are shown in Figure 7. The optical images in the dataset were obtained from the GF-1 satellite (2 m resolution), while the SAR images were obtained from the GF-3 satellite (5 m resolution), and a unified resolution of 5 m was achieved through bilinear interpolation. Some samples of this dataset are shown in Figure 8.

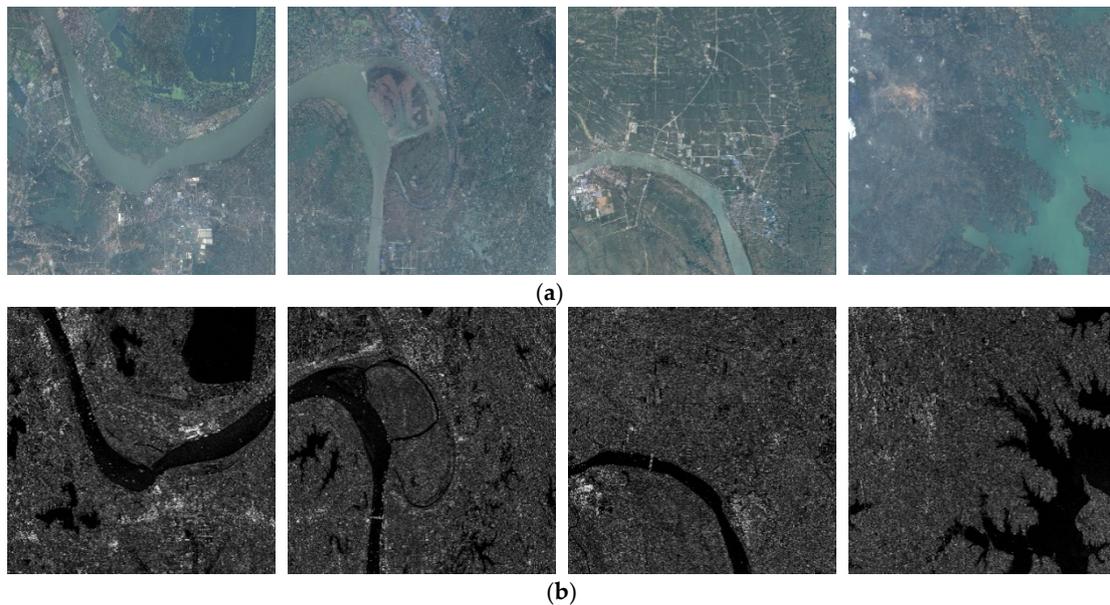
In the experiment, we produce a high-resolution SAR and optical image dataset covering three different types of scenes: urban, suburban, and mountain. Such a dataset and a publicly available medium-resolution dataset named WHU-OPT-SAR are used collectively to evaluate these fusion methods using three different kinds of evaluation criteria. Detailed specifications of the two datasets are given in Table 5.



**Figure 6.** Three types of images for the experiment: (a) Optical images covering the urban setting. (b) SAR images covering the urban setting. (c) Optical images covering the suburban setting. (d) SAR images covering the suburban setting. (e) Optical images covering the mountains. (f) SAR images covering the mountains.



**Figure 7.** WHU-OPT-SAR dataset: Geographic location of the dataset in Hubei Province, China.



**Figure 8.** Some samples from WHU-OPT-SAR: (a) Optical images. (b) SAR images.

**Table 5.** Specifications of the datasets.

	YYX-OPT-SAR	WHU-OPT-SAR
Number of images (pairs)	150	100
Image pixel size	512 × 512	5556 × 3704
Ground resolution (m)	0.5	5
The surrounding areas	Weinan City, Shaanxi Province in China	Wuhan City, Hubei Province in China

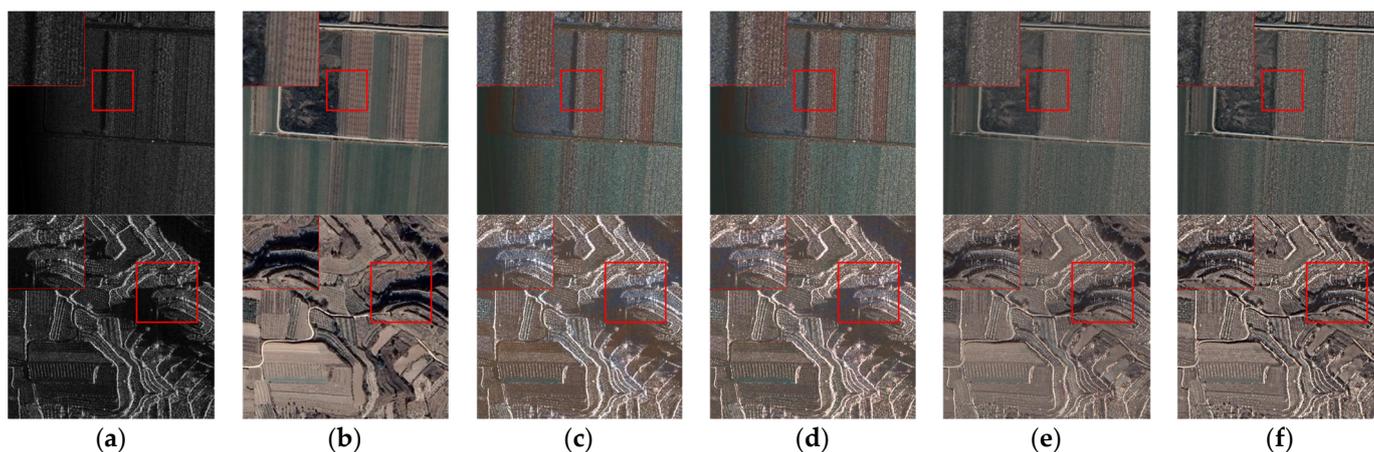
## 5. Experimental Analysis

### 5.1. Visual Evaluation

#### 5.1.1. Visual Evaluation of High-Resolution Images

The datasets proposed in the previous section are fused using the 11 fusion methods given in the second section (Table 1) to generate the corresponding fusion results. CS methods select a specific component from the forward transform and replace it with the SAR image for inverse transformation, which makes full use of SAR image information.

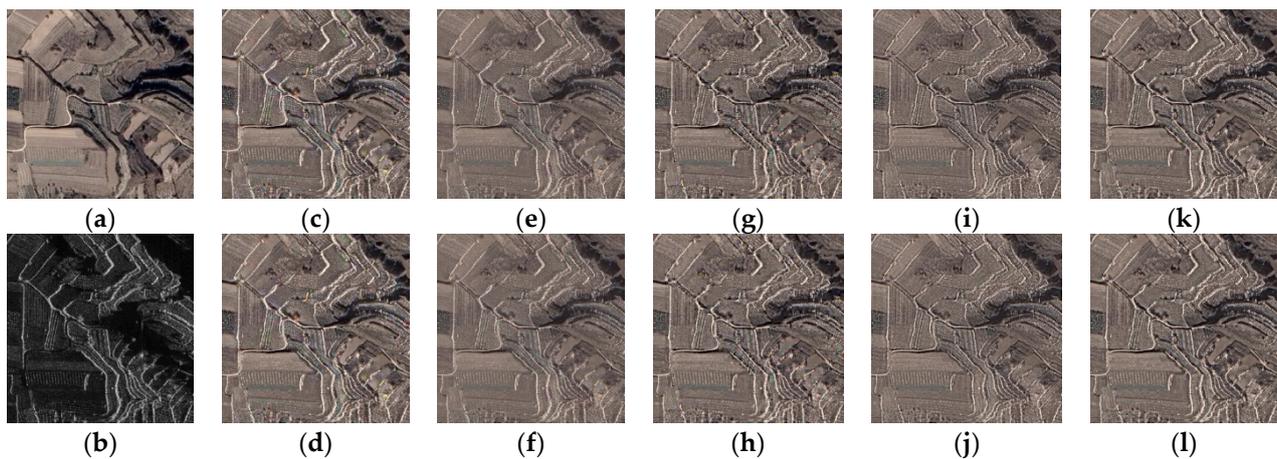
Compared with other types of fusion methods, this strategy makes the fusion results include the texture feature of SAR images and introduce shadows in SAR images. Figure 9 illustrates the fusion results of CS methods (including IHS and PCA) and MSD methods (including GP and NSCT). It can be clearly seen that the fusion results of CS methods introduce shadows in the SAR images, which makes image interpretation challenging and fails to achieve the purpose of fusing complementary information. Compared with the results of MSD methods, those of CS methods present worse global spectral quality, often manifesting as color distortion in the areas of roads and vegetation.



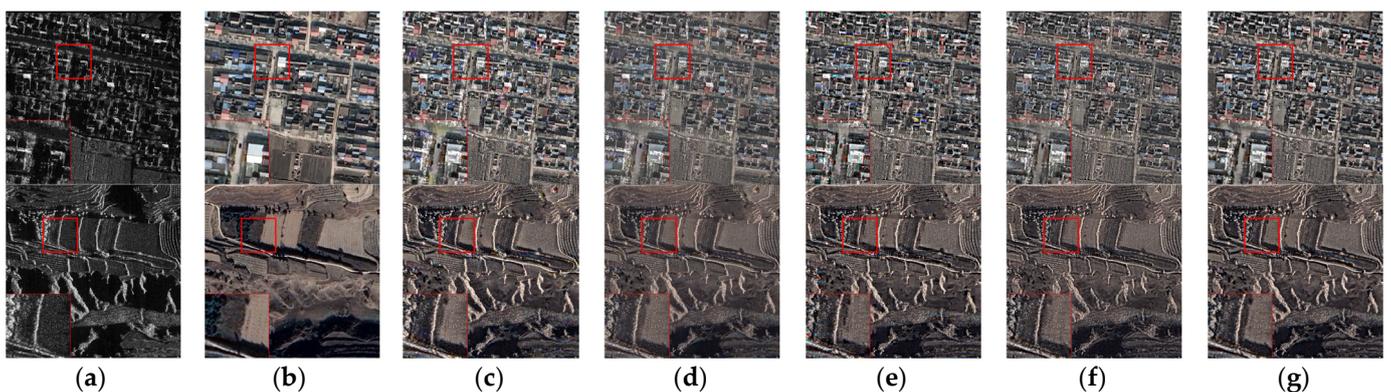
**Figure 9.** Fusion results of component substitution methods and partial multiscale decomposition methods. (a) SAR. (b) Optical image. (c) IHS. (d) PCA. (e) GP\_1. (f) NSCT\_1. A larger version of the red square is shown in the upper left corner.

MSD methods exhibit lower overall color distortion compared to CS methods. Visual observation of the results obtained by applying the two different high-frequency component fusion rules is basically consistent, as shown in Figure 10. Therefore, in the rest of the qualitative evaluation, we only select the fusion results obtained based on one high-frequency fusion rule in each MSD method. As a result, we select the “max-absolute” rule for experimental analysis. On the one hand, by merging the separated low-frequency components, MSD methods effectively retain the spectral information of optical images; on the other hand, by using specific fusion rules for the integration of high-frequency components, the bright textures and edge features of the SAR images are combined into the fusion results to effectively filter out the shadows (as seen in the last two columns of Figure 9). Figure 11 shows the fusion results of different MSD methods. It is apparent that the fusion results of LP and DWT combine more SAR image information, thus introducing the brighter edge features and noise information from SAR images. However, these two methods have color distortion in some areas, such as the edges of houses (the first row of Figure 11) and trees (the second row of Figure 11).

The model-based fusion methods, including SR and GTF, showcase their advantages and disadvantages due to their different fusion strategies. SR tends to make an either–or choice between optical and SAR images, which is consistent with the sparse coefficient selection rule (specifically, the “choose-max” fusion rule with the L1-norm activity level measure). Consequently, the fusion result of SR resembles that of the SAR image in the non-shaded part and that of the optical image in the shaded part, with a rough transition between the two regions. In comparison, GTF integrates the optical image information effectively, but the texture details of SAR are not well introduced, resulting in fuzzy object edges. Figure 12 shows the selected fusion results of these two model-based methods.

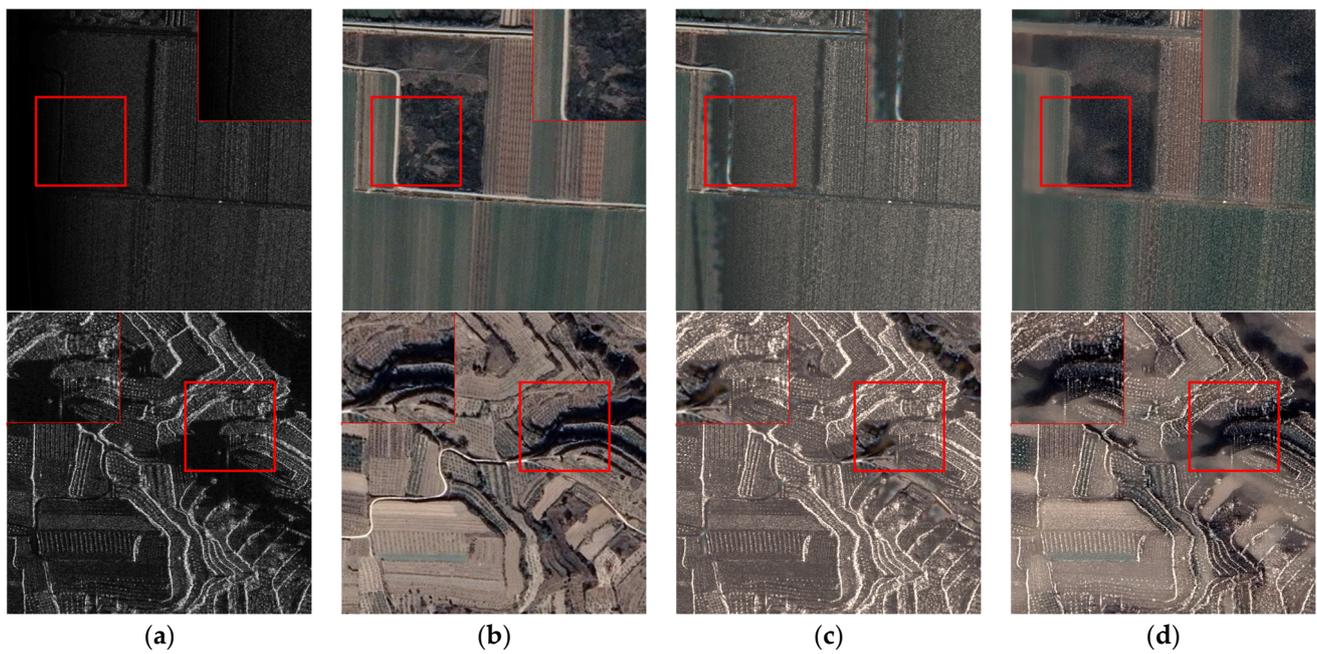


**Figure 10.** The fusion results obtained using two high-frequency component fusion rules of multi-scale decomposition. (a) Optical. (b) SAR. (c) LP\_1. (d) LP\_2. (e) GP\_1. (f) GP\_2. (g) DWT\_1. (h) DWT\_2. (i) CVT\_1. (j) CVT\_2. (k) NSCT\_1. (l) NSCT\_2. XX\_1 denotes “max-absolute” rule; XX\_2 denotes “max-absolute” rule with a local window-based consistency verification scheme.

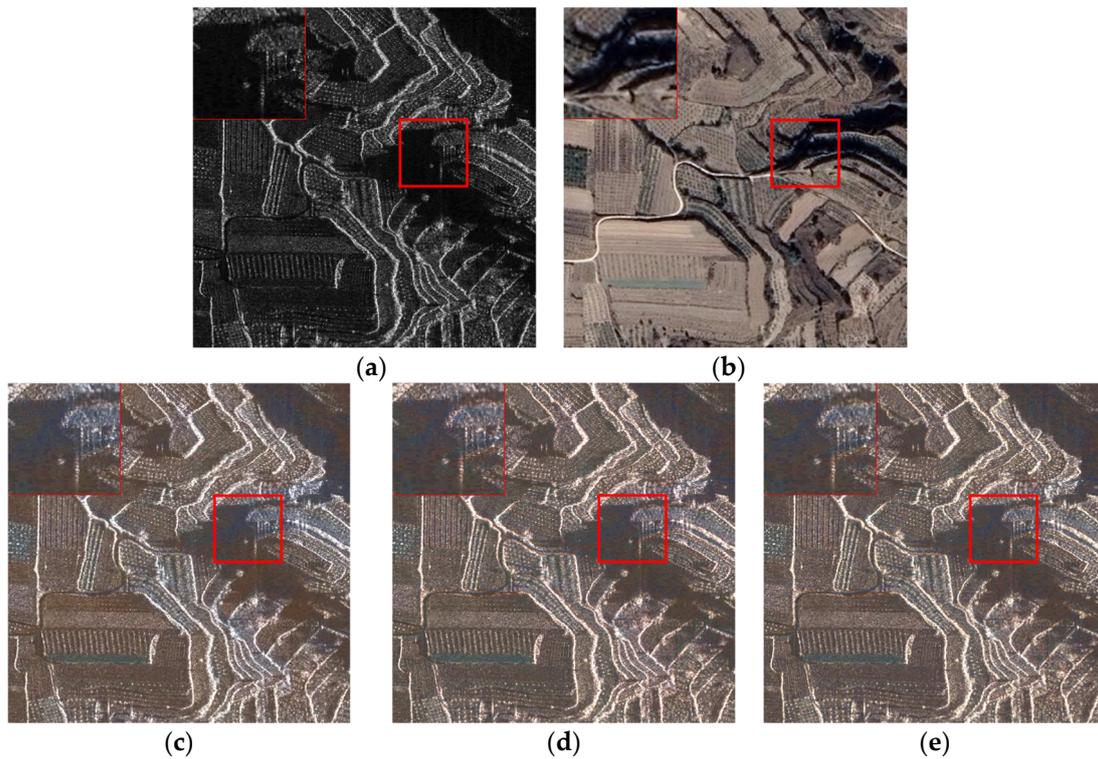


**Figure 11.** Fusion results of different multiscale decomposition methods. (a) SAR. (b) Optical. (c) LP\_1. (d) GP\_1. (e) DWT\_1. (f) CVT\_1. (g) NSCT\_1. A larger version of the red square is shown in the lower left corner.

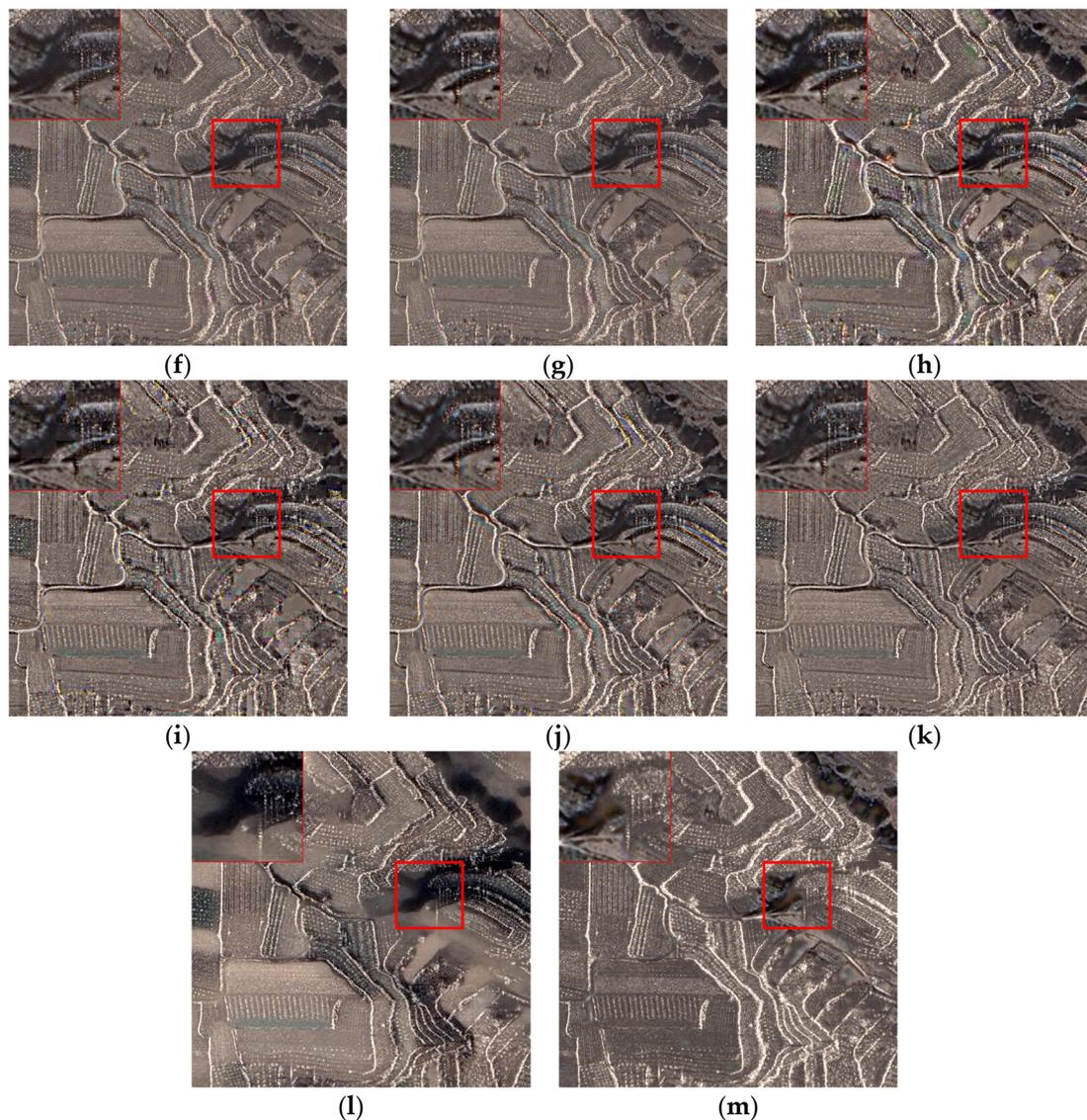
The fusion results of all fusion methods under the same image are shown in Figure 13, and the enlarged image of the selected area is displayed in the upper left corner. From the perspective of visual effect, the fusion results of different types of fusion methods are obviously different. CS methods take the SAR image as a component to participate in inverse transformation and effectively use the pixel intensity information of the SAR image. Compared with other fusion methods, CS methods combine more SAR image information, thus introducing more SAR image texture features and shadows. From the box selection area, it can also be seen that the image texture features and shadows are more similar to the SAR image. MSD methods have advantages in preserving the spectral information of optical images by combining the separated low-frequency components. At the same time, the high-frequency component is selected through specific fusion rules, and the bright and edge features of the SAR image are fused into the fusion result effectively, while the shadows are filtered. Among them, LP, DWT, and DTCWT combine more SAR image information and introduce brighter edge features and noise information in the SAR image, while the color transition is not natural, such as the roads in the figure. GTF can retain spectral information better, but the boundary of ground objects is fuzzy.



**Figure 12.** Fusion results of different model-based methods: (a) SAR. (b) Optical. (c) SR. (d) GTF. A larger version of the red square is shown in the corner.



**Figure 13.** Cont.



**Figure 13.** Fusion results of different methods for the high-resolution images: (a) SAR. (b) Optical. (c) IHS. (d) PCA. (e) GS. (f) GP. (g) NSCT. (h) LP. (i) DWT. (j) DTCWT. (k) CVT. (l) SR. (m) GTF. A larger version of the red square is shown in the upper left corner.

### 5.1.2. Visual Evaluation of Medium-Resolution Images

From the perspective of visual effect, the fusion results of medium-resolution images exhibit obvious differences among different types of fusion methods. As is shown in Figure 14, CS methods take the SAR image as a component and participate in the inverse transformation, utilizing the pixel intensity information of the SAR image effectively. Compared with other fusion methods, CS methods combine more SAR image information, which introduces more texture features and shadows from SAR images.

By combining the separated low-frequency components, MSD methods excel in preserving the spectral information of optical images. At the same time, by using specific fusion rules to select high-frequency components, they effectively incorporate the bright point features and edge features of SAR images into the fusion results while filtering out shadows.

GTF can retain the spectral information better. The same kind of ground objects share the same color in the optical images, but the boundaries of ground objects appear blurred. The fusion results of GTF contain less texture information from SAR images, and they only contain the brighter edge information.

## 5.2. Statistical Evaluation

### 5.2.1. Statistical Evaluation of High-Resolution Images

The nine quality assessment metrics shown in Table 4 are used for quantitative analysis of the fusion methods. Because each fused image has three bands, we calculate the average value of the metrics of these three bands and use it as the final assessment metric. In addition, we analyze the fusion results of different scenes, including urban, suburban, and mountain scenes, separately. Considering that each scene contains 50 fused images, the average value of their metrics is taken as the result of each method in such a scene. Figure 15 depicts the assessment metric values of the fusion results of each compared method. The higher the metric value, the better the fusion quality.

The EN index reflects the amount of image information, and the PSNR index can measure the ratio of signal to noise and then reflect the degree of image distortion. Figure 15a shows that the fusion result of the mountain scene contains more information than the other two types of ground objects, indicating that more SAR image information is combined. However, when SAR image information is introduced, the noise information of the SAR image is also introduced. Therefore, as shown in Figure 15c, the PSNR value corresponding to the mountain scene is lower than that of the other two types of ground objects.

From the perspective of different types of fusion methods, the fusion quality of CS methods (such as IHS, PCA, and GS) is generally at the same level. For the images in suburban and mountain areas, this achieves the maximum values on most metrics (such as EN, SD,  $Q^{AB/F}$ , and  $Q_y$ ). In the images covering the urban scene, most metrics of the fusion results obtained through PCA achieve the maximum values (such as PSNR,  $Q_w$ ,  $Q_e$ , and  $Q_{cb}$ ).

In the MSD methods, LP has the highest EN and SD values in all images of the three different scenes, which proves that the fusion results of LP contain higher contrast and richer information content than those of other MSD methods. In the three types of images, NSCT achieves the highest values in most metrics (such as  $Q_w$ ,  $Q_e$ , and  $Q_o$ ), indicating a better fusion effect for NSCT.

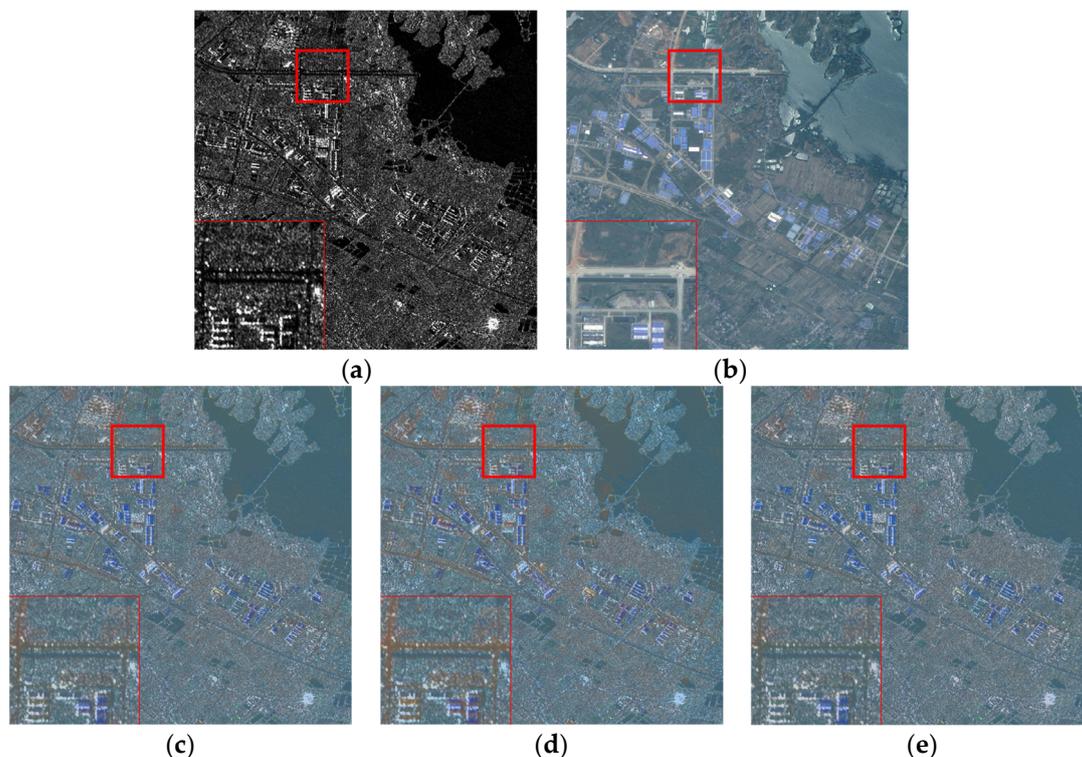
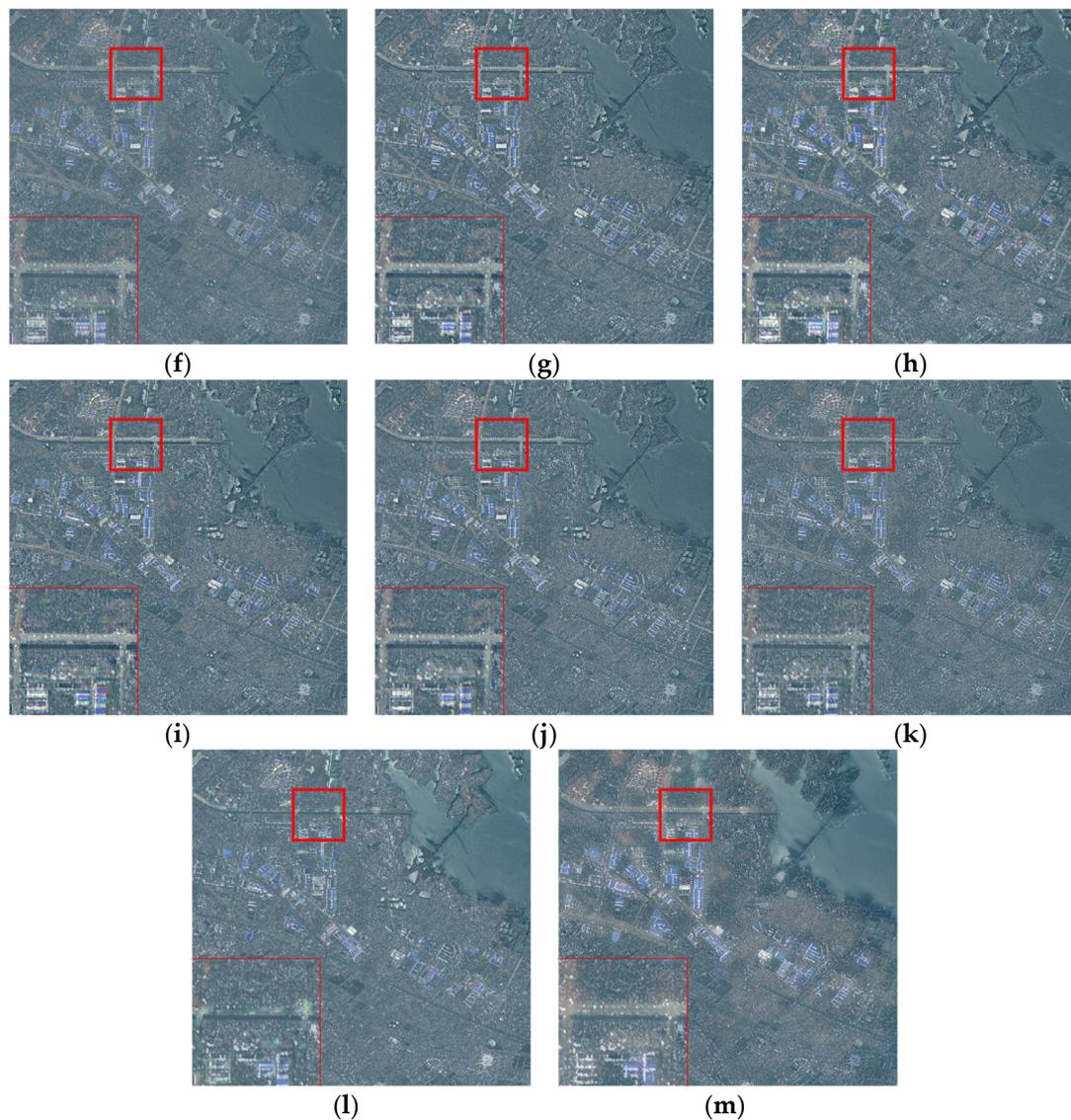


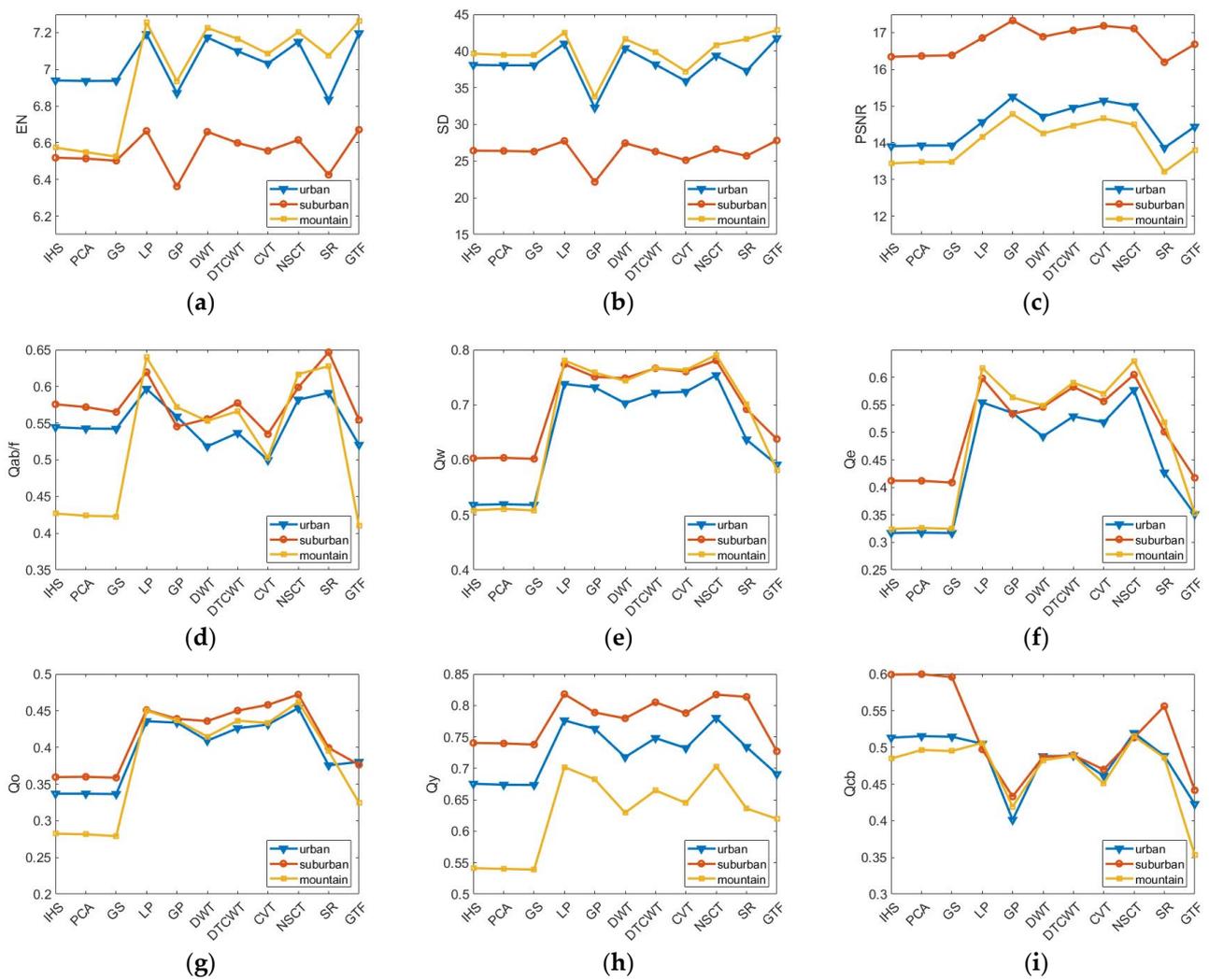
Figure 14. Cont.



**Figure 14.** Fusion results of different methods for the medium-resolution images: (a) SAR. (b) Optical. (c) IHS. (d) PCA. (e) GS. (f) GP. (g) NSCT. (h) LP. (i) DWT. (j) DTCWT. (k) CVT. (l) SR. (m) GTF. A larger version of the red square is shown in the lower left corner.

In the model-based methods, GTF obtains higher EN, SD, and PSNR for the three types of images, revealing that the fusion results of GTF contain more information. From the image quality assessment metrics, the fusion results of SR are worse than those of GTF and MSD methods on the whole.

Considering that the visual interpretation of SR is poor and the spectral distortion is serious, the fusion methods that obtain the highest values on each metric except SR are listed in Table 6. GTF has the highest EN and SD values, indicating that the fusion result contains more information. MSD methods obtain the highest values in most of the image quality assessment metrics in the three types of images; in particular, NSCT has the highest  $Q_w$ ,  $Q_e$ , and  $Q_o$  values, which indicates that NSCT presents the best fusion result in all of these fusion methods.



**Figure 15.** Nine indexes obtained using each method for the high-resolution images of the three types of features: (a) EN. (b) SD. (c) PSNR. (d)  $Q^{AB/F}$ . (e)  $Q_w$ . (f)  $Q_e$ . (g)  $Q_o$ . (h)  $Q_y$ . (i)  $Q_{cb}$ .

**Table 6.** Fusion methods for obtaining the highest index value for various types of high-resolution images (excluding SR).

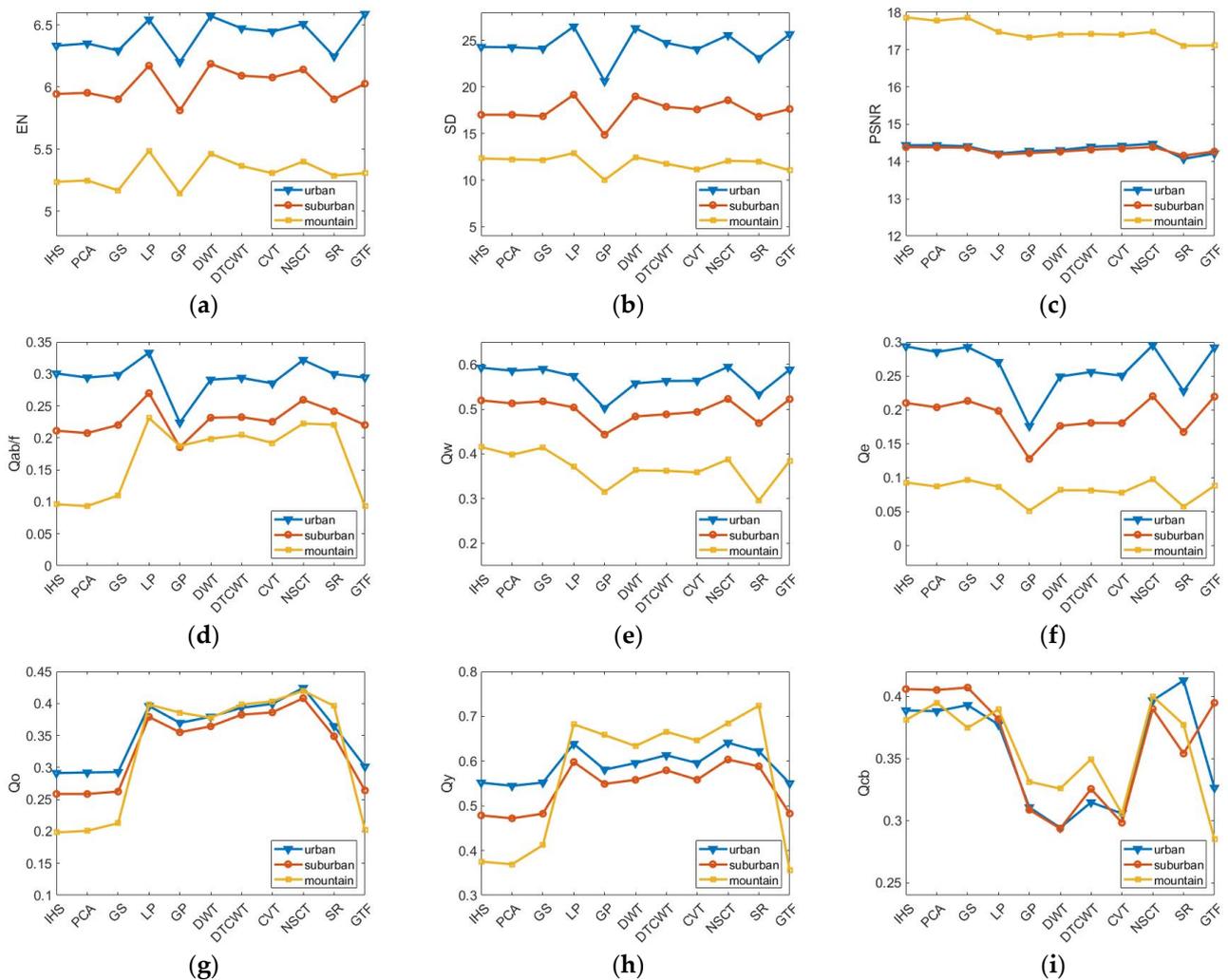
	EN	SD	PSNR	$Q^{AB/F}$	$Q_o$	$Q_w$	$Q_e$	$Q_y$	$Q_{cb}$
Urban	GTF	GTF	GP	LP	NSCT	NSCT	NSCT	NSCT	NSCT
Suburban	GTF	GTF	GP	LP	NSCT	NSCT	NSCT	LP	PCA
Mountain	GTF	GTF	GP	LP	NSCT	NSCT	NSCT	NSCT	NSCT

Based on the above subjective comparison and objective analysis, NSCT performs best when dealing with the fusion of optical and SAR images, mainly including urban and mountain scenes. In terms of statistical evaluation metrics, NSCT and LP have their own advantages in optical and SAR image fusion of suburban areas. However, from the perspective of visual effect, the fusion images obtained through LP have color distortion. Therefore, combining visual effect and statistical evaluation metrics, NSCT can obtain the best fusion effect in the image fusion of these three types of ground objects.

### 5.2.2. Statistical Evaluation of Medium-Resolution Images

In addition to the quantitative analysis of the fusion results of the high-resolution images, we also select the dataset of the medium-resolution images for image fusion

and quantitatively analyze the performance of different fusion methods on this dataset. Figure 16 shows the assessment metric values of the fusion results of all of the compared methods.



**Figure 16.** Nine indexes obtained using each method for the medium-resolution images of the three types of features: (a) EN, (b) SD, (c) PSNR, (d)  $Q^{AB/F}$ , (e)  $Q_w$ , (f)  $Q_e$ , (g)  $Q_o$ , (h)  $Q_y$ , (i)  $Q_{cb}$ .

From the perspective of different types of fusion methods, the fusion results of the medium-resolution images present a similar law to those of the high-resolution images. The fusion quality of CS methods (such as IHS, PCA, and GS) is almost at the same level, among which IHS achieves the highest PSNR and  $Q_w$  values in mountain scenes.

Most of the quality assessment metrics of MSD methods are higher than those of CS methods. Among them, LP has the highest SD and  $Q^{AB/F}$  values among the three types of images. NSCT obtains the highest  $Q_o$ ,  $Q_y$ , and  $Q_e$  values for the three types of images, indicating that NSCT can obtain better fusion results.

The fusion result obtained through GTF for the medium-resolution images is worse than that for the high-resolution images. This discrepancy is because the quality of the fusion results of GTF depends on the information richness of the original optical and SAR images, and the information of the medium-resolution images is less than that of the high-resolution images.

The fusion methods (excluding SR) that obtain the highest value on each metric are listed in Table 7. It can be observed that LP has the highest SD and  $Q^{AB/F}$  and NSCT has

the highest  $Q_y$ ,  $Q_e$ , and  $Q_o$  for the three types of images. Therefore, NSCT demonstrates the best performance in image fusion across various metrics.

**Table 7.** Fusion methods for obtaining the highest index value for various types of medium-resolution images (excluding SR).

	EN	SD	PSNR	$Q^{ABIF}$	$Q_o$	$Q_w$	$Q_e$	$Q_y$	$Q_{cb}$
Urban	GTF	LP	NSCT	LP	NSCT	NSCT	NSCT	NSCT	NSCT
Suburban	DWT	LP	NSCT	LP	NSCT	NSCT	NSCT	NSCT	GS
Mountain	LP	LP	IHS	LP	NSCT	IHS	NSCT	NSCT	NSCT

Based on the subjective comparison and objective metric analysis of the two groups of data, the conclusions can be drawn as follows. The fusion results of CS methods combine more SAR image information, like texture features and shadows. Nevertheless, the visual effect is worse than that of MSD methods. The surface boundary of GTF is fuzzy, and the visual effect is not as good as that of MSD methods. In the MSD methods, LP and NSCT are at the forefront of most metrics, indicating that these two methods obtain better fusion results. However, considering the color distortion of LP, NSCT performs best among all of the compared methods.

### 5.3. Fusion Evaluation According to Classification

#### 5.3.1. Fusion Evaluation of High-Resolution Images According to Classification

The datasets mentioned earlier are fused using the 11 fusion methods outlined in the second section (Table 1) to generate the corresponding fusion results. CS methods select the specific component of the forward transform and replace it with the SAR image for inverse transformation, thus maximizing the utilization of SAR image information, like texture feature. Figure 9 represents the fusion results of CS methods (including IHS and PCA) and MSD methods (including GP and NSCT). It is evident that the fusion results of the CS methods introduce shadows in the SAR images, thus complicating image interpretation and failing to achieve the purpose of fusing complementary information. In comparison to MSD methods, CS methods present worse global spectral quality, with noticeable color distortion in road and vegetation areas. In this section, we evaluate the 11 fusion methods through image classification for high-resolution images. In the experiment, 50 pairs of fused images, including some typical ground objects, such as bare ground, low vegetation, trees, houses, and roads, are classified by SVM, RF, and CNN, respectively. Given that the dataset contains multiple fused images, the average of their measurements is taken as the result of each method. Tables 8–10 show the classification accuracy results. For instance, the classification results of a pair of optical and SAR images are shown in Figure 17. From the classification accuracy table, it is obvious that CNN achieves higher overall accuracy compared to SVM and RF. Simultaneously, the bare ground is more prone to be misclassified, while the houses and roads exhibit lower misclassification rates. This is because the spectral characteristics of the bare ground are highly uncertain, and the spectral characteristics of the houses and roads are obviously different from those of other categories. From Figure 17, it can be seen that CNN produces classification results more similar to the labels, indicating superior performance compared to SVM and RF, which show more instances of misclassification.

From Figure 17 and the classification accuracy table, it can be concluded that the fused images obtain better classification accuracy compared with single optical or SAR image, which demonstrates that image fusion effectively integrates complementary information of multimodal images and improves classification accuracy. Compared with the single SAR image, the optical image yields better classification accuracy. The special imaging mechanism of SAR leads to the inherent multiplicative speckle noise in SAR images, seriously affecting the interpretation of SAR images. As a result, the classification accuracy of SAR images is poor. During the data fusion process, these effects are transmitted to the fusion image, resulting in the same confusion in the image classification. However, the

classification accuracy of the fused image surpasses that of the single optical and single SAR images. This shows the feasibility of using optical and SAR image fusion to improve classification results.

**Table 8.** SVM classification accuracy table of the high-resolution images. The bolded item is the highest value of classification accuracy for each feature category.

SVM	Bare Ground	Low Vegetation	Trees	Houses	Roads	OA
RGB	49.75%	60.92%	60.37%	62.36%	62.91%	59.69%
SAR	34.08%	27.88%	45.68%	22.14%	42.44%	39.29%
CVT	39.45%	53.32%	50.03%	<b>73.97%</b>	63.96%	59.85%
DTCWT	37.39%	56.28%	46.60%	62.99%	61.40%	56.68%
DWT	33.71%	57.02%	48.33%	48.98%	58.98%	52.37%
GP	40.08%	59.70%	57.16%	73.61%	65.13%	61.98%
GS	50.26%	58.50%	59.19%	68.31%	70.23%	64.08%
GTF	<b>50.96%</b>	60.45%	58.01%	70.96%	<b>72.25%</b>	<b>65.59%</b>
IHS	49.48%	<b>62.39%</b>	<b>61.15%</b>	67.61%	65.35%	63.11%
LP	37.68%	57.15%	50.77%	58.96%	66.57%	57.46%
NSCT	44.38%	55.32%	52.13%	68.98%	65.79%	60.28%
PCA	49.74%	61.59%	60.25%	68.82%	67.78%	63.96%
SR	36.08%	29.89%	45.41%	60.20%	46.13%	42.24%

**Table 9.** RF classification accuracy table of the high-resolution images. The bolded item is the highest value of classification accuracy for each feature category.

RF	Bare Ground	Low Vegetation	Trees	Houses	Roads	OA
RGB	46.78%	57.77%	60.74%	70.10%	61.67%	57.79%
SAR	32.41%	27.85%	44.89%	22.47%	42.44%	39.20%
CVT	32.58%	48.85%	47.21%	77.40%	62.76%	56.75%
DTCWT	31.59%	51.37%	43.11%	70.37%	60.72%	54.04%
DWT	31.37%	51.00%	45.63%	59.17%	59.72%	52.11%
GP	33.88%	55.27%	53.09%	74.74%	64.45%	58.57%
GS	45.11%	56.05%	59.23%	80.65%	69.97%	64.13%
GTF	45.87%	55.00%	55.78%	78.43%	<b>72.27%</b>	<b>64.45%</b>
IHS	<b>47.33%</b>	57.97%	<b>60.77%</b>	81.14%	64.74%	63.69%
LP	34.29%	52.22%	47.32%	70.71%	66.73%	56.68%
NSCT	37.20%	50.37%	49.08%	73.99%	64.31%	57.96%
PCA	45.95%	<b>58.59%</b>	60.22%	<b>81.26%</b>	68.12%	64.41%
SR	34.49%	29.48%	44.42%	58.35%	48.54%	42.77%

From the perspective of different types of fusion methods, the overall classification accuracy of the three CS methods (such as IHS, PCA, and GS) is nearly the same. The overall classification accuracy of PCA and GS is higher than that of others, suggesting that the two methods have better effect in classification applications though the performance of image fusion is poorer than that of MSD methods. Among the MSD methods, the overall classification accuracy of GP and NSCT is higher, indicating their superior classification effectiveness. Nonetheless, the overall classification accuracy of DTCWT, DWT and LP

is lower, with the classification accuracy of fused images obtained by the three methods even lower than that of the optical image. This illustrates that not all optical–SAR fusion methods can improve land classification.

**Table 10.** CNN classification accuracy table of the high-resolution images. The bolded item is the highest value of classification accuracy for each feature category.

CNN	Bare Ground	Low Vegetation	Trees	Houses	Roads	OA
RGB	56.39%	71.32%	72.54%	81.01%	73.61%	70.07%
SAR	35.37%	62.39%	70.44%	57.78%	55.14%	57.86%
CVT	50.42%	66.82%	67.55%	82.74%	75.67%	71.41%
DTCWT	47.07%	67.81%	66.73%	78.46%	74.64%	69.79%
DWT	44.96%	65.50%	62.32%	73.21%	73.00%	66.76%
GP	51.06%	70.78%	68.67%	81.67%	76.21%	71.94%
GS	53.85%	72.02%	79.72%	<b>86.69%</b>	<b>78.24%</b>	75.74%
GTF	<b>58.43%</b>	72.05%	74.97%	83.75%	77.87%	<b>75.83%</b>
IHS	52.88%	74.11%	<b>80.35%</b>	85.32%	72.60%	73.89%
LP	48.83%	65.07%	65.67%	74.62%	76.44%	68.87%
NSCT	53.05%	65.89%	65.33%	82.72%	77.40%	71.18%
PCA	53.76%	<b>74.74%</b>	80.20%	85.68%	75.40%	75.12%
SR	36.17%	62.11%	70.18%	71.32%	56.79%	61.01%

Overall, GTF obtains the highest overall classification accuracy among the eleven fusion methods. Compared with the single optical image and single SAR image, the fused image has a better classification effect, with up to about 5% improvement.

### 5.3.2. Fusion Evaluation of Medium-Resolution Images According to Classification

In this section, similarly to the previous section, we evaluate the 11 fusion methods according to image classification for medium-resolution images. Some typical ground objects, such farmland, city, village, water, forest, and roads, are classified by SVM, RF, and CNN, respectively. Tables 11–13 show the classification accuracy of the 11 fusion methods. The overall classification accuracy for medium-resolution images is observed to be lower than that of the high-resolution images. Like the high-resolution images, the city and water have a smaller chance of being misclassified due to their distinct spectral characteristics. Among the 11 fusion methods, GTF obtains the highest overall classification accuracy. The results in Figure 18 and the classification accuracy tables indicate that, for most cases, the overall classification accuracy after fusion is better than before fusion across all of the three classification methods. This indicates that optical–SAR fusion has the potential to improve land classification. But, the visual effect is not as good as that of the high-resolution images, possibly due to the lower image resolution and a larger number of categories. Figure 18 also reveals that the overall classification result of CNN is more similar to the ground truth, indicating that this method has better classification results, whereas SVM and RF have more misclassification.

Building on the above groups of classification experiments, we can obtain the following results: (1) The classification effect of CNN is better than that of RF and SVM. (2) Features with relatively different spectral characteristics from other features have a lower probability of misclassification, while features with relatively uncertain spectral characteristics have a higher probability of misclassification. (3) Fused images obtained using fusion methods exhibit better a classification effect compared to single SAR or optical images. (4) Among

the 11 fusion methods selected, GTF consistently achieves the highest overall classification accuracy for all three classification methods.

**Table 11.** SVM classification accuracy table of the medium-resolution images. The bolded item is the highest value of classification accuracy for each feature category.

SVM	Farmland	City	Village	Water	Forest	Road	Others	OA
RGB	39.97%	66.02%	44.46%	58.82%	61.56%	42.61%	43.75%	51.97%
SAR	21.37%	46.66%	20.39%	46.92%	27.26%	24.61%	28.45%	33.65%
CVT	34.48%	54.90%	26.92%	48.13%	45.93%	24.74%	28.60%	43.89%
DTCWT	37.15%	59.48%	23.55%	47.51%	41.30%	25.24%	25.58%	43.54%
DWT	37.41%	50.44%	25.49%	44.33%	42.41%	26.15%	26.27%	44.12%
GP	27.03%	54.97%	28.11%	45.74%	33.75%	20.73%	28.53%	45.19%
GS	<b>39.98%</b>	61.50%	<b>44.53%</b>	66.78%	51.12%	36.39%	40.98%	53.40%
GTF	35.90%	<b>79.87%</b>	40.55%	<b>79.14%</b>	<b>62.20%</b>	<b>45.96%</b>	<b>60.42%</b>	<b>56.95%</b>
IHS	30.10%	61.86%	34.67%	68.20%	47.75%	39.19%	36.38%	52.77%
LP	29.48%	45.80%	26.94%	53.76%	48.20%	30.86%	30.62%	48.53%
NSCT	31.18%	54.98%	27.11%	61.42%	45.80%	33.19%	34.65%	52.01%
PCA	30.13%	69.53%	31.01%	66.86%	49.01%	38.20%	36.23%	52.81%
SR	27.29%	34.26%	25.64%	47.45%	27.95%	29.12%	29.19%	32.61%

**Table 12.** RF classification accuracy table of the medium-resolution images. The bolded item is the highest value of classification accuracy for each feature category.

RF	Farmland	City	Village	Water	Forest	Road	Others	OA
RGB	37.41%	66.55%	37.54%	61.87%	58.16%	40.54%	38.30%	48.50%
SAR	20.92%	44.62%	20.14%	46.59%	27.35%	24.69%	28.44%	33.63%
CVT	31.10%	55.19%	24.59%	41.19%	41.96%	22.30%	23.44%	40.16%
DTCWT	32.37%	52.41%	22.82%	42.09%	37.09%	22.68%	22.49%	40.71%
DWT	36.78%	54.63%	25.54%	41.15%	40.63%	24.93%	24.08%	42.96%
GP	26.51%	55.10%	26.83%	45.92%	41.06%	31.84%	27.61%	43.62%
GS	<b>37.61%</b>	60.41%	<b>39.71%</b>	56.83%	56.69%	33.09%	33.79%	49.51%
GTF	34.18%	<b>80.71%</b>	38.39%	<b>72.33%</b>	<b>58.85%</b>	<b>44.69%</b>	<b>56.69%</b>	<b>54.82%</b>
IHS	28.54%	61.01%	30.42%	58.33%	43.75%	35.98%	31.34%	49.46%
LP	27.45%	48.89%	25.83%	49.26%	44.14%	28.16%	27.92%	46.36%
NSCT	29.15%	52.90%	27.16%	54.44%	40.18%	33.56%	29.81%	48.80%
PCA	29.80%	69.39%	37.81%	66.62%	49.57%	35.18%	28.63%	49.37%
SR	25.93%	44.14%	26.21%	49.20%	27.69%	27.81%	28.96%	34.47%

**Table 13.** CNN classification accuracy table of the medium-resolution images. The bolded item is the highest value of classification accuracy for each feature category.

CNN	Farmland	City	Village	Water	Forest	Road	Others	OA
RGB	52.83%	81.49%	61.08%	72.85%	70.89%	70.25%	67.85%	65.23%
SAR	30.49%	53.45%	35.20%	62.07%	31.84%	41.26%	29.70%	39.00%
CVT	44.48%	75.19%	56.32%	69.73%	64.91%	51.44%	57.65%	53.91%

Table 13. Cont.

CNN	Farmland	City	Village	Water	Forest	Road	Others	OA
DTCWT	48.35%	79.58%	54.52%	67.41%	61.56%	56.86%	55.43%	53.76%
DWT	47.81%	71.64%	56.71%	65.38%	63.62%	56.61%	57.37%	54.26%
GP	49.33%	74.68%	58.56%	66.84%	63.85%	61.64%	59.56%	55.39%
GS	51.38%	80.91%	<b>66.70%</b>	<b>83.02%</b>	66.07%	69.20%	66.87%	68.11%
GTF	<b>53.14%</b>	85.05%	65.94%	78.39%	<b>73.33%</b>	<b>72.45%</b>	<b>67.97%</b>	<b>69.49%</b>
IHS	40.62%	71.46%	54.82%	68.12%	67.78%	59.29%	56.43%	66.56%
LP	49.91%	76.82%	57.64%	63.98%	69.25%	60.55%	50.76%	60.73%
NSCT	42.73%	75.78%	58.17%	72.56%	67.18%	63.73%	55.69%	66.98%
PCA	50.28%	<b>86.13%</b>	64.26%	80.53%	68.96%	70.11%	66.43%	68.19%
SR	31.25%	61.46%	40.71%	66.54%	34.59%	45.33%	31.29%	42.91%

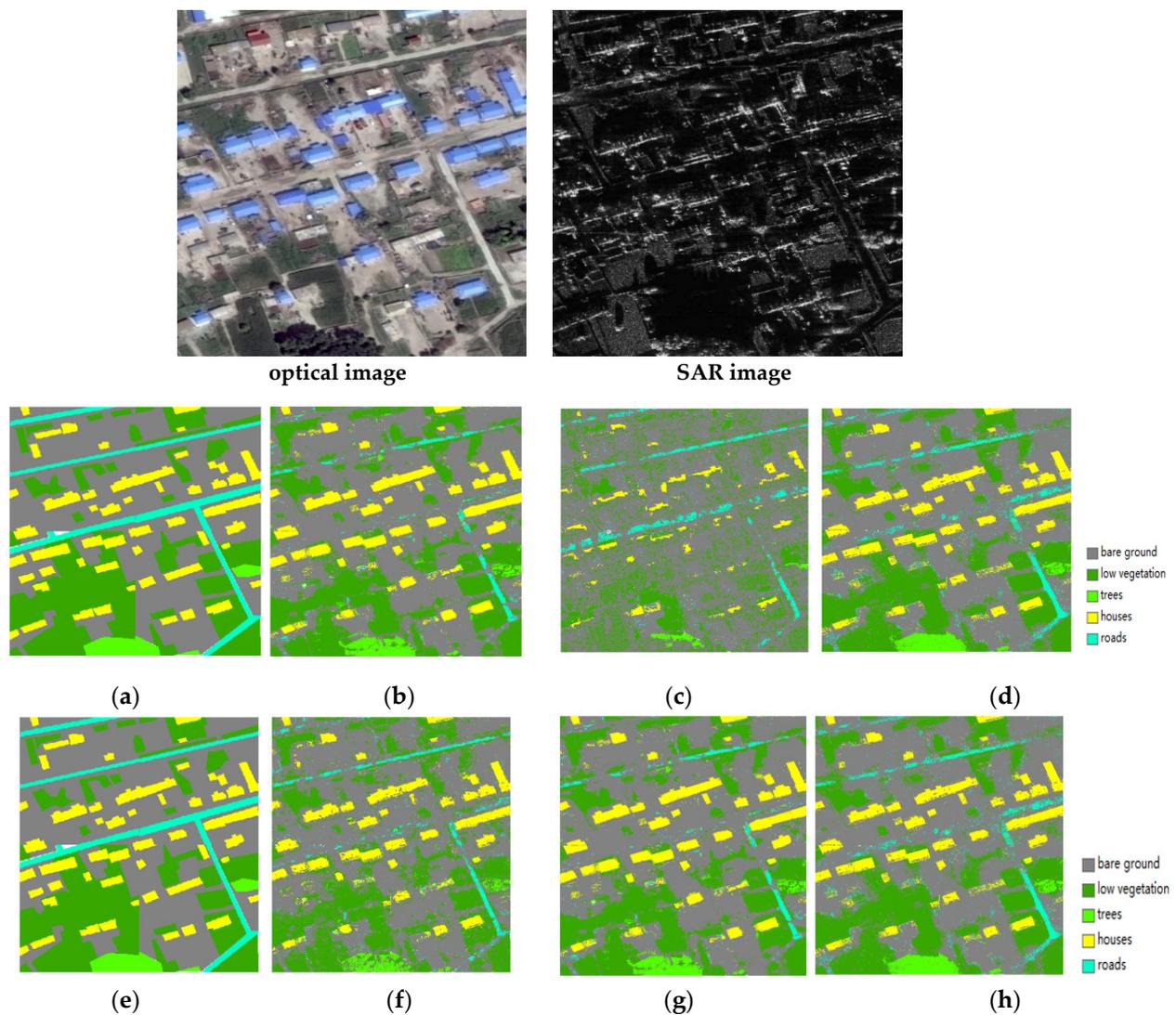
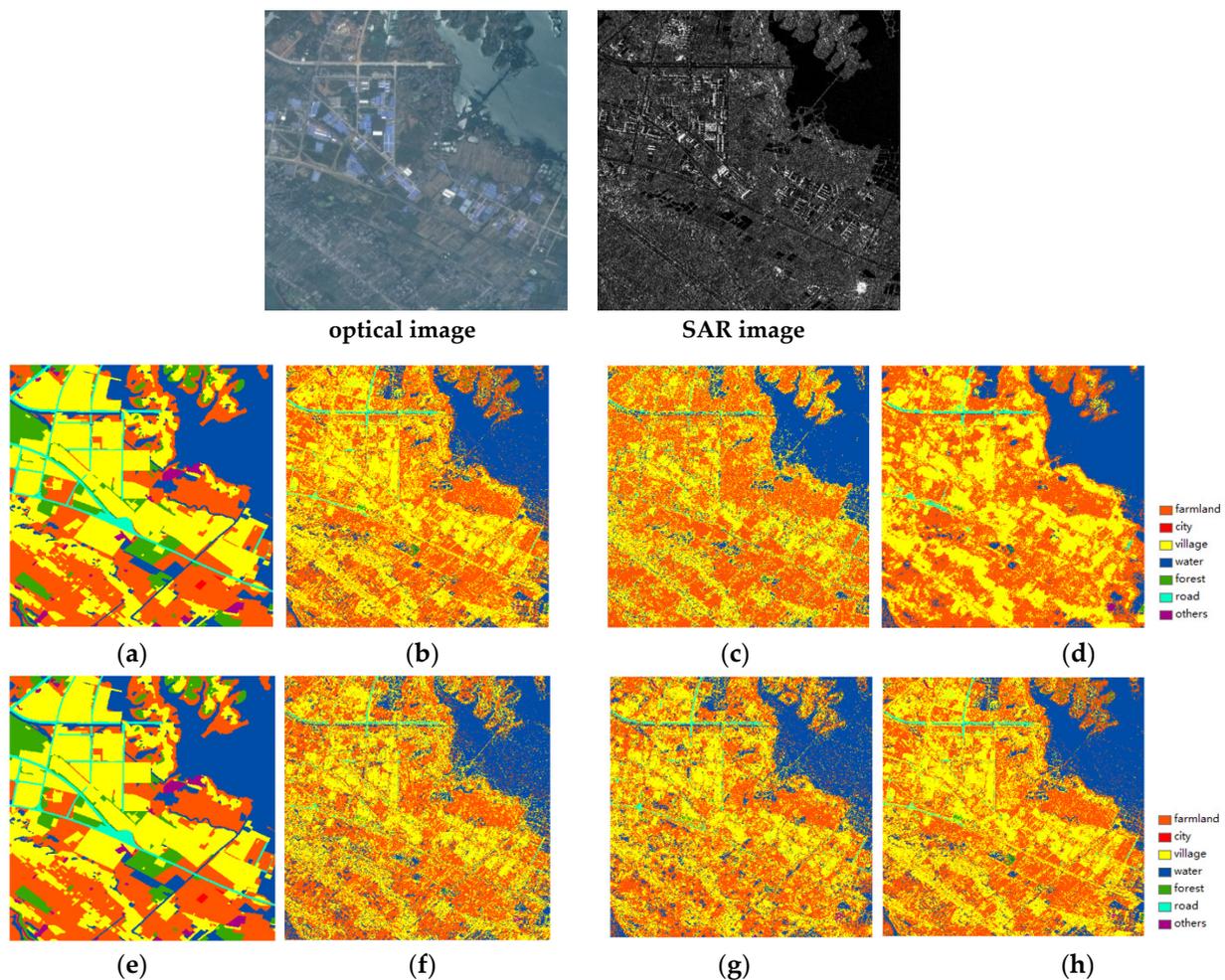


Figure 17. Classified images of (a) label, (b) RGB, (c) SAR, (d) GTF, (e) label, (f) SVM, (g) RF, (h) CNN. (YYX-OPT-SAR).



**Figure 18.** Classified images of (a) label, (b) RGB, (c) SAR, (d) GTF, (e) label, (f) SVM, (g) RF, (h) CNN (WHU-OPT-SAR).

For image quality metrics, among all of the fusion methods involved in comparison, NSCT has a superior visual effect in image fusion, and its quantitative metrics of fusion results are at the forefront. In the evaluation using image classification, three classic methods, including SVM, RF, and CNN, are used to perform image classification. Most experimental results show that the overall classification accuracy after fusion is better than that before fusion for all three classification methods. This demonstrates that optical–SAR fusion can improve land classification. In all of these fusion methods, GTF obtains the best classification results. Therefore, we recommend NSCT for image fusion and GTF for classification applications.

## 6. Conclusions

The fusion of optical and SAR images is an important research direction in remote sensing. This fusion allows for the effective integration of complementary information from SAR and optical sources, thus better meeting the requirements of remote sensing applications, such as target recognition, classification, and change detection, so as to realize the collaborative utilization of multi-modal images.

In order to select appropriate methods to achieve high-quality fusion of SAR and optical images, this paper systematically reviews the current pixel-level fusion algorithms for SAR and optical image fusion and then selects eleven representative fusion methods, including CS methods, MSD methods, and model-based methods for comparison analysis. In the experiment, we produce a high-resolution SAR and optical image dataset (named

YYX-OPT-SAR) covering three different types of scenes, including urban, suburban, and mountain scenes. Additionally, a publicly available medium-resolution dataset named WHU-OPT-SAR is utilized to evaluate these fusion methods according to three different kinds of evaluation criteria, including the visual evaluation, the objective image quality metrics, and the classification accuracy.

The evaluation based on image quality metrics reveals that MSD methods can effectively avoid the negative effects of SAR image shadows on the corresponding area of the fusion result compared with the CS methods, while the model-based methods show comparatively poorer performance. Notably, among all of the evaluated fusion methods, NSCT presents the most effective fusion result.

It is suggested that image quality metrics should not be the only option for the interpretation of fused images. Therefore, image classification should also be used as an additional metric to evaluate the quality of fused images, because some fused images with poor image quality metrics can obtain the highest classification accuracy. The experiment utilizes three classic classification methods (SVM, RF, and CNN) to perform image classification. Most experimental results show that the overall classification accuracy after fusion is better than that before fusion for all three classification methods, indicating that optical–SAR fusion can improve land classification. Notably, in all of these fusion methods, GTF obtains the best classification results. Consequently, the suggestion is to employ NSCT for image fusion and GTF for image classification based on the experimental findings.

The differences between this paper and the previous conference paper are mainly related to the following four aspects: First, we extend the original self-built dataset from the original 60 image pairs to 150 image pairs, and we add classification labels to provide data support for subsequent advanced visual tasks. While previous contributions did not expose the dataset, this paper exposes the produced dataset. Second, because the self-built dataset is a high-resolution image, in order to better evaluate the fusion effect of the fusion method at different resolutions, we added the experiment under the published medium-resolution images as a comparison, so as to prove that the excellent fusion method can obtain better results in images with different resolutions. Third, the previous contribution is a short paper, and there is no detailed introduction to optical and SAR pixel-level image fusion algorithms. This paper systematically reviews the current pixel-level fusion algorithms of optical and SAR image fusion. Fourth, we evaluate the fusion quality between different fusion methods by combining subsequent advanced visual tasks, and we verify the effectiveness of image fusion in image classification, proving that the fused image can obtain better results than the original image in image classification.

At present, most pixel-level fusion methods of optical and SAR images rely on traditional algorithms, which may lack comprehensive analysis and interpretation of these highly heterogeneous data. Consequently, these methods inevitably encounter performance bottlenecks. Therefore, this non-negligible limitation further creates a strong demand for alternative tools with powerful processing capabilities. As a cutting-edge technology, deep learning has made remarkable breakthroughs in many computer vision tasks due to its impressive capabilities in data representation and reconstruction. Naturally, it has been successfully applied to other types of multimodal image fusion, such as optical-infrared fusion [62,63]. Accordingly, we will also explore the application of deep learning methods for optical and SAR image fusion in the future.

**Author Contributions:** Conceptualization and methodology, J.L. and Y.Y.; writing—original draft preparation, J.L.; writing—review and editing, J.L., J.Z. and Y.Y.; supervision, C.Y., H.L. and Y.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported in part by the National Natural Science Foundation of China under Grant 41971281, Grant 42271446, and in part by the Natural Science Foundation of Sichuan Province under Grant 2022NSFSC0537. We are also very grateful for the constructive comments of the anonymous reviewers and members of the editorial team.

**Data Availability Statement:** We produced a high-resolution SAR and optical image fusion dataset named YYX-OPT-SAR. The download link for the dataset is <https://github.com/yeyuanxin110/YYX-OPT-SAR> (accessed on 21 January 2023). The publicly available medium-resolution SAR and optical dataset named WHU-OPT-SAR can be downloaded from this link: <https://github.com/AmberHen/WHU-OPT-SAR-dataset> (accessed on 3 June 2021).

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

## References

1. Kulkarni, S.C.; Rege, P.P. Fusion of RISAT-1 SAR Image and Resourcesat-2 Multispectral Images Using Wavelet Transform. In Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 7–8 March 2019.
2. Ghassemian, H. A review of remote sensing image fusion methods. *Inf. Fusion* **2016**, *32*, 75–89. [[CrossRef](#)]
3. Pohl, C.; Van Genderen, J.L. Review article Multisensor image fusion in remote sensing: Concepts, methods and applications. *Int. J. Remote Sens.* **1998**, *19*, 823–854. [[CrossRef](#)]
4. Battsengel, V.; Amarsaikhan, D.; Bat-Erdene, T.; Egshiglen, E.; Munkh-Erdene, A.; Ganzorig, M. Advanced Classification of Lands at TM and Envisat Images of Mongolia. *Adv. Remote Sens.* **2013**, *2*, 102–110. [[CrossRef](#)]
5. Sanli, F.B.; Abdikan, S.; Esetlili, M.T.; Sunar, F. Evaluation of image fusion methods using PALSAR, RADARSAT-1 and SPOT images for land use/land cover classification. *J. Indian Soc. Remote Sens.* **2016**, *45*, 591–601. [[CrossRef](#)]
6. Abdikana, S.; Sanlia, F.B.; Balcikb, F.B.; Gokselb, C. Fusion of Sar Images (Palsar and Radarsat-1) with Multispectral Spot Image: A Comparative Analysis of Resulting Images. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Beijing, China, 3–11 July 2008.
7. Klonus, S.; Ehlers, M. Pansharpening with TerraSAR-X and optical data, 3rd TerraSAR-X Science Team Meeting. In Proceedings of the 3rd TerraSAR-X Science Team Meeting, Darmstadt, Germany, 25–26 November 2008; pp. 25–26.
8. Abdikan, S.; Sanli, F.B. Comparison of different fusion algorithms in urban and agricultural areas using sar (palsar and radarsat) and optical (spot) images. *Bol. Ciênc. Geod.* **2012**, *18*, 509–531. [[CrossRef](#)]
9. Sanli, F.B.; Abdikan, S.; Esetlili, M.T.; Ustuner, M.; Sunar, F. Fusion of terrasarsar-x and rapideye data: A quality analysis. *ISPRS—Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *XL-7/W2*, 27–30. [[CrossRef](#)]
10. Clerici, N.; Calderón, C.A.V.; Posada, J.M. Fusion of Sentinel-1A and Sentinel-2A Data for Land Cover Mapping: A Case Study in the Lower Magdalena Region, Colombia. *J. Maps* **2017**, *13*, 718–726. [[CrossRef](#)]
11. He, W.; Yokoya, N. Multi-Temporal Sentinel-1 and -2 Data Fusion for Optical Image Simulation. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 389. [[CrossRef](#)]
12. Benedetti, P.D.; Ienco, R.; Gaetano, K.; Ose, R.G.; Pensa, S.D. M<sup>3</sup>Fusion: A Deep Learning Architecture for Multiscale Multimodal Multitemporal Satellite Data Fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4939–4949. [[CrossRef](#)]
13. Hughes, L.H.; Merkle, N.; Bürgmann, T.; Auer, S.; Schmitt, M. Deep Learning for SAR-Optical Image Matching. In Proceedings of the IEEE International Geoscience and Remote Sensing, Yokohama, Japan, 28 July–2 August 2019; pp. 4877–4880.
14. Abdikan, S.; Bilgin, G.; Sanli, F.B.; Uslu, E.; Ustuner, M. Enhancing land use classification with fusing dual-polarized TerraSAR-X and multispectral RapidEye data. *J. Appl. Remote Sens.* **2015**, *9*, 096054. [[CrossRef](#)]
15. Gaetano, R.; Cozzolino, D.; D’Amiano, L.; Verdoliva, L.; Poggi, G. Fusion of sar-optical data for land cover monitoring. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 23–28 July 2017; pp. 5470–5473.
16. Gibril, M.B.A.; Bakar, S.A.; Yao, K.; Idrees, M.O.; Pradhan, B. Fusion of RADARSAT-2 and multispectral optical remote sensing data for LULC extraction in a tropical agricultural area. *Geocarto Int.* **2017**, *32*, 735–748. [[CrossRef](#)]
17. Hu, J.; Ghamisi, P.; Schmitt, A.; Zhu, X.X. Object based fusion of polarimetric SAR and hyperspectral imaging for land use classification. In Proceedings of the 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA, 21–24 August 2016; pp. 1–5.
18. Kulkarni, S.C.; Rege, P.P.; Parishwad, O. Hybrid fusion approach for synthetic aperture radar and multispectral imagery for improvement in land use land cover classification. *J. Appl. Remote Sens.* **2019**, *13*, 034516. [[CrossRef](#)]
19. Dabbiru, L.; Samiappan, S.; Nobrega, R.A.A.; Aanstoos, J.A.; Younan, N.H.; Moorhead, R.J. Fusion of synthetic aperture radar and hyperspectral imagery to detect impacts of oil spill in Gulf of Mexico. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 1901–1904.
20. Gao, J.; Yuan, Q.; Li, J.; Zhang, H.; Su, X. Cloud Removal with Fusion of High Resolution Optical and SAR Images Using Generative Adversarial Networks. *Remote Sens.* **2020**, *12*, 191. [[CrossRef](#)]
21. Kang, W.; Xiang, Y.; Wang, F.; You, H. CFNet: A Cross Fusion Network for Joint Land Cover Classification Using Optical and SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 1562–1574. [[CrossRef](#)]
22. Ye, Y.; Liu, W.; Zhou, L.; Peng, T.; Xu, Q. An Unsupervised SAR and Optical Image Fusion Network Based on Structure-Texture Decomposition. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [[CrossRef](#)]

23. Liu, H.; Ye, Y.; Zhang, J.; Yang, C.; Zhao, Y. Comparative Analysis of Pixel Level Fusion Algorithms in High Resolution SAR and Optical Image Fusion. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 2829–2832.
24. Chen, C.-M.; Hepner, G.; Forster, R. Fusion of hyperspectral and radar data using the IHS transformation to enhance urban surface features. *ISPRS J. Photogramm. Remote Sens.* **2003**, *58*, 19–30. [[CrossRef](#)]
25. Yin, N.; Jiang, Q.-G. Feasibility of multispectral and synthetic aperture radar image fusion. In Proceedings of the 6th International Congress on Image and Signal Processing (CISP), Hangzhou, China, 16–18 December 2013; pp. 835–839.
26. Yang, J.; Ren, G.; Ma, Y.; Fan, Y. Coastal wetland classification based on high resolution SAR and optical image fusion. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016.
27. Liu, Y.; Liu, S.; Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **2015**, *24*, 147–164. [[CrossRef](#)]
28. Eltaweel, G.S.; Helmy, A.K. Fusion of Multispectral and Full Polarimetric SAR Images in NSST Domain. *Comput. Sci. J.* **2014**, *8*, 497–513.
29. Pajares, G.; de la Cruz, J.M. A wavelet-based image fusion tutorial. *Pattern Recognit.* **2004**, *37*, 1855–1872. [[CrossRef](#)]
30. Yang, B.; Li, S. Multifocus image fusion and restoration with sparse representation. *IEEE Trans. Instrum. Meas.* **2009**, *59*, 884–892. [[CrossRef](#)]
31. Zhang, W.; Yu, L. SAR and Landsat ETM+ image fusion using variational model. In Proceedings of the 2010 International Conference on Computer and Communication Technologies in Agriculture Engineering (CCTAE), Chengdu, China, 12–13 June 2010; pp. 205–207.
32. Li, H.; Manjunath, B.S.; Mitra, S.K. Multisensor Image Fusion Using the Wavelet Transform. *Graph. Models Image Process.* **1995**, *57*, 235–245. [[CrossRef](#)]
33. Li, S.; Yang, B.; Hu, J. Performance comparison of different multi-resolution transforms for image fusion. *Inf. Fusion* **2011**, *12*, 74–84. [[CrossRef](#)]
34. Aharon, M.; Elad, M.; Bruckstein, A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322. [[CrossRef](#)]
35. Bruckstein, A.M.; Donoho, D.L.; Elad, M. From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images. *SIAM Rev.* **2009**, *51*, 34–81. [[CrossRef](#)]
36. Tu, T.-M.; Huang, P.S.; Hung, C.-L.; Chang, C.-P. A Fast Intensity—Hue—Saturation Fusion Technique with Spectral Adjustment for IKONOS Imagery. *IEEE Geosci. Remote Sens. Lett.* **2004**, *1*, 309–312. [[CrossRef](#)]
37. Wang, Z.; Ziou, D.; Armenakis, C.; Li, D.; Li, Q. A comparative analysis of image fusion methods. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1391–1402. [[CrossRef](#)]
38. Aiuzzi, B.; Baronti, S.; Selva, M. Improving Component Substitution Pansharpening through Multivariate Regression of MS +Pan Data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3230–3239. [[CrossRef](#)]
39. Burt, P.J.; Adelson, E.H. *The Laplacian Pyramid as a Compact Image Code*; Fischler, M.A., Firschein, O., Eds.; Readings in Computer Vision; Morgan Kaufmann: San Francisco, CA, USA, 1987; pp. 671–679.
40. Burt, P.J. A Gradient Pyramid Basis for Pattern-Selective Image Fusion. *Proc. SID* **1992**, 467–470.
41. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [[CrossRef](#)]
42. Selesnick, I.; Baraniuk, R.; Kingsbury, N. The dual-tree complex wavelet transform. *IEEE Signal Process. Mag.* **2005**, *22*, 123–151. [[CrossRef](#)]
43. Candès, E.; Demanet, L.; Donoho, D.; Ying, L. Fast Discrete Curvelet Transforms. *Multiscale Model. Simul.* **2006**, *5*, 861–899. [[CrossRef](#)]
44. Da Cunha, A.; Zhou, J.; Do, M. The Nonsubsampled Contourlet Transform: Theory, Design, and Applications. *IEEE Trans. Image Process.* **2006**, *15*, 3089–3101. [[CrossRef](#)] [[PubMed](#)]
45. Ma, J.; Chen, C.; Li, C.; Huang, J. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **2016**, *31*, 100–109. [[CrossRef](#)]
46. Piella, G.; Heijmans, H. A new quality metric for image fusion. In Proceedings of the 2003 International Conference on Image Processing (Cat. No.03CH37429), Barcelona, Spain, 14–17 September 2003; p. III-173-6.
47. Yang, C.; Zhang, J.-Q.; Wang, X.-R.; Liu, X. A novel similarity based quality metric for image fusion. *Inf. Fusion* **2008**, *9*, 156–160. [[CrossRef](#)]
48. Jagalingam, P.; Hegde, A.V. A Review of Quality Metrics for Fused Image. *Aquat. Procedia* **2015**, *4*, 133–142. [[CrossRef](#)]
49. Xydeas, C.; Petrović, V. Objective image fusion performance measure. *Electron. Lett.* **2000**, *36*, 308–309. [[CrossRef](#)]
50. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [[CrossRef](#)]
51. Chen, Y.; Blum, R.S. A new automated quality assessment algorithm for image fusion. *Image Vis. Comput.* **2007**, *27*, 1421–1432. [[CrossRef](#)]
52. Liu, Z.; Blasch, E.; Xue, Z.; Zhao, J.; Laganieri, R.; Wu, W. Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 94–109. [[CrossRef](#)]
53. Zhang, X.; Ye, P.; Xiao, G. VIFB: A Visible and Infrared Image Fusion Benchmark. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020.

54. Ye, Y.; Ren, X.; Zhu, B.; Tang, T.; Tan, X.; Gui, Y.; Yao, Q. An adaptive attention fusion mechanism convolutional network for object detection in remote sensing images. *Remote Sens.* **2022**, *14*, 516. [[CrossRef](#)]
55. Melgani, F.; Bruzzone, L. Support vector machines for classification of hyperspectral remote-sensing images. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Toronto, ON, Canada, 4–28 June 2002; Volume 1, pp. 506–508.
56. Feng, T.; Ma, H.; Cheng, X. Greenhouse Extraction from High-Resolution Remote Sensing Imagery with Improved Random Forest. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 553–556.
57. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. In Proceedings of the Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
58. Ye, Y.; Zhu, B.; Tang, T.; Yang, C.; Xu, Q.; Zhang, G. A robust multimodal remote sensing image registration method and system using steerable filters with first-and second-order gradients. *ISPRS J. Photogramm. Remote Sens.* **2022**, *188*, 331–350. [[CrossRef](#)]
59. Ye, Y.; Tang, T.; Zhu, B.; Yang, C.; Li, B.; Hao, S. A multiscale framework with unsupervised learning for remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [[CrossRef](#)]
60. Ye, Y.; Bruzzone, L.; Shan, J.; Bovolo, F.; Zhu, Q. Fast and Robust Matching for Multimodal Remote Sensing Image Registration. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9059–9070. [[CrossRef](#)]
61. Li, X.; Zhang, G.; Cui, H.; Hou, S.; Wang, S.; Li, X.; Chen, Y.; Li, Z.; Zhang, L. MCANet: A joint semantic segmentation framework of optical and SAR images for land use classification. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *106*, 102638. [[CrossRef](#)]
62. Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2018**, *48*, 11–26. [[CrossRef](#)]
63. Tang, W.; He, F.; Liu, Y. TCCFusion: An infrared and visible image fusion method based on transformer and cross correlation. *Pattern Recognit.* **2023**, *137*, 109295. [[CrossRef](#)]

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