Registration of Multiresolution Remote Sensing Images Based on L2-Siamese Model

Rongbo Fan, Bochuan Hou, Jinbao Liu, Jianhua Yang, and Zenglin Hong

***Abstract*—The registration of multiresolution optical remote sensing images has been widely used in image fusion, change detection, and image stitching. However, traditional registration methods achieve poor accuracy in the registration of multiresolution remote sensing images. In this study, we propose a framework for generating deep features via a deep residual encoder (DRE) fused with shallow features for multiresolution remote sensing image registration. Through an L2 normalization Siamese network (L2-Siamese) based on the DRE, the multiscale loss function is used to learn the attribute characteristics and distance characteristics of two key points and obtain the trained feature extractor. Finally, the DRE is used to extract the deep features of the key points and their neighbors, which are concatenated with the shallow features into a fusion feature vector to complete the image registration. We performed comprehensive experiments on four sets of multiresolution optical remote sensing images and two sets of synthetic aperture radar images. The results demonstrate that the proposed registration model can achieve subpixel registration. The relative registration accuracy improved by 1.6%–7.5%, whereas the overall performance improved by 4.5%–14.1%.**

***Index Terms*—Deep descriptors, L2-Siamese, multiresolution image registration, residual encoder, satellite remote sensing, Siamese network.**

I. Introduction

With rapid development of remote sensing technology in recent years, remote sensing images are advancing toward multiresolution and multispectrum. Improved ground observation requires the integration of heterogeneous remote sensing data, and multiresolution remote sensing image registration is fundamental in the field of remote sensing image processing. The goal is to make any pair of pixels in a multiresolution remote sensing image at the same location represent the same geographic location [1], [2].

Manuscript received June 28, 2020; revised August 21, 2020 and October 19, 2020; accepted November 3, 2020. Date of publication November 19, 2020; date of current version January 6, 2021. This work was supported by the Key Research and Development Plan of Shaanxi Province under Grant D5140200023. *(Corresponding author: Jianhua Yang.)*

Rongbo Fan is with the School of Automation, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: fanke@mail.nwpu.edu.cn).

Bochuan Hou is with the School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: houbochuan@bupt.edu.cn).

Jinbao Liu is with the School of Water Resources and Hydro-Electric Engineering, Xi'an University of Technology, Xi'an 710082, China (e-mail: jinbaoliu@xaut.edu.cn).

Jianhua Yang is with the School of Automation, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: yangjianhua@nwpu.edu.cn).

Zenglin Hong is with the Shaanxi Institute of Geological Survey, Xi'an 710054, China (e-mail: lhqhzl@163.com).

Digital Object Identifier 10.1109/JSTARS.2020.3038922

Multiresolution satellite image registration technology is used extensively in remote sensing image fusion [3], ground feature change detection [4], and disaster monitoring and assessment [5], and its accuracy determines the performance of these algorithms. However, it is challenging to obtain remote sensing images with the same sensor type and spatial resolution to complete the above application by following the actual process. Moreover, because of the technical limitations of current satellite sensors, typically only high-spectral low-spatial resolution multispectral images (MS) and high-spatial low-spectral resolution panchromatic images (PAN) can be obtained. To exploit full use of limited remote sensing image resources and improve the accuracy of multiresolution remote sensing image registration, multiresolution remote sensing image registration technology is very important.

Traditionally, there have been three main types of image registration: region-based methods, feature-based methods, and combination of both methods [6]–[8]. Feature-based image registration is the approach most commonly used for remote sensing. This approach attempts to estimate the geometric transformation between images by identifying and matching features, which must be significant and stable. These features include points, edges, and contours [9], [10], and the features (descriptors) with the most similar invariance are matched (the features are matched between the remote sensing and reference images). There are currently several types of feature descriptors, including the neighbor image intensity value, momentum-based descriptor, and scale-invariant feature transform (SIFT). After feature extraction, the texture and gray information of feature points can be used for local deformation registration of remote sensing images through the optical flow estimation method [11], [12]. In addition, the large-scale geometric transformation of remote sensing images can be performed through a transformation matrix [13].

The most commonly used descriptor is SIFT [14], which filters out unstable matches through a predefined distance ratio threshold [15]. Numerous improved algorithms SIFT have been developed. For example, the SIFT variant speeded-up robust features (SURFs) offer a faster rate of operation and better robustness [16]. Ma *et al.* [17] proposed the position scale orientation-SIFT (PSO-SIFT) using the new gradient definition and feature matching method. They combined the position, scale, and direction of each key point to improve the matching point pairs. Paul and Pati [18] proposed an improved uniform R-SIFT that can effectively generate sufficiently robust, reliable, and evenly distributed matched key points.

The synthetic aperture radar-SIFT (SAR-SIFT) [19] is a new gradient calculation method to generate directional and robust descriptors to speckle noise. SAR-SIFT has better performance, especially in registration tasks using SAR images with different incidence angles. Wu *et al.* [20] proposed a novel point matching algorithm named fast sample consensus (FSC), which can improve matches in fewer iterations.

Registration methods based on feature have achieved good results. However, the disadvantage of this approach is that feature extraction requires manual design and often ignores the neighborhood information of key points, resulting in a low matching rate. In the remote sensing image registration task, the reference and remote sensing images may originate from different sensors and possess different spatial resolutions. Taking the QuickBird satellite image as an example, the spatial resolution of the PAN image is 0.61–0.72 m, but the spatial resolution of its MS image is only 2.44–2.88 m [21], which leads to a clear difference in the feature of the same target.

Taking the SIFT registration algorithm as an example, SIFT determines a large number of key points, but there exists a large proportion of mismatched points. This is due to the following reasons. First, remote sensing images contain several features with high similarity. However, SIFT-based descriptors only express the shallow information of key points, making the descriptors less significant. Second, the SIFT descriptor ignores the neighborhood information of key points. Hence, its image feature is not used effectively.

Deep neural networks (DNNs) have been applied to achieve remote sensing image registration via four main approaches at present. The first approach involves using the feature map generated by the intermediate layer of the DNN, instead of the original image, for image registration [22], [23]. The idea is to put image pairs in a pretrained DNN, and then merge high- and low-order feature maps (such as SIFT feature and SURF maps). Moreover, the deep features extracted by the existing methods are less robust and more sensitive to image deformation.

The second approach involves directly using a DNN to match key points [24]–[26], which translates the image matching problem into two types of classification problems. This approach has a high matching accuracy, but the number of calculations required increases with the square of the number of key points, significantly increasing the registration time. These limitations make the approach impractical. Moreover, models cannot perform the nearest neighbor search (NNS) [15].

The third approach is to use the deformation field to correct the remote sensing image. Zhang *et al.* [27] proposed a method that uses a deep fully convolutional neural network to perform multiscale deformation correction on remote sensing images. Ma *et al.* [28] proposed using an effective coarse-to-fine strategy and to develop a new two-step registration method based on deep features and local features for completing multi-modal remote sensing image registration.

The fourth approach is based on deep network image matching (registration) that not only can extract the deep features of key points, but also can automatically extract key points. Methods that apply this approach include D2-Net [29], Super-Point [30], and RF-Net models [31]. These methods are commonly used for natural image matching, and have good robustness to angle and illumination.

The Siamese network [32] is a DNN often used for image matching. Unlike in the general neural network structure, the input of the Siamese network comprises two images, and the output is the similarity between the two images [26], [33], [34]. Siamese networks can also be used in image registration. The dual channel of the Siamese network is used to extract the deep characteristic information of patched cropped images, and then this deep characteristic information is normalized as deep descriptors [35]–[37].

In particular, the models used for feature extraction of two images share weights to ensure that the deep features of the two images can be obtained under the same metric. Considering this feature, in this study, we applied the L2 normalization Siamese network (L2-Siamese) to train the depth feature extractor and used a deep residual encoder (DRE) network to extract the depth feature of key points and their neighborhood to improve the robustness of image transformation and noise reduction.

In summary, the traditional shallow-feature matching and DNN-based image registration methods cannot always meet the requirements of multiresolution remote sensing image registration tasks. Therefore, in this study, we propose a multiresolution remote sensing image registration framework involving the extraction of deep features of key points and their neighbor, followed by the merging of deep and shallow features. The main contributions of this work are as follows.

1) We propose the use of the L2-Siamese model to train the deep feature extractor of key points and introduce a model training method and a loss function.

2) The DRE model extracts significantly robust deep features. The fusion feature vector formed by the fusion of deep and shallow features achieves improved performance in multiresolution remote sensing image registration tasks.

3) To solve the problem of scarcity of fully registered multiresolution remote sensing images, we propose a method to create high-generality datasets.

The remainder of this article is organized as follows. Section II details the registration algorithm of multiresolution remote sensing images, including the algorithm framework and model, data set production, and registration. Section III outlines the experiments conducted using the proposed algorithm on six datasets. Section IV discusses the experimental results of the study. The conclusion of the study is presented in Section V.

II. Methodology

A. Algorithm Framework

The algorithm framework presented in this article comprises three main sections: an L2-Siamese model used to train the DRE network; a DRE network that extracts the deep features of key points and their neighborhoods; and a combination of deep features and SIFT descriptors to generate feature vectors and the multiresolution remote sensing image registration. Fig. 1 shows the block diagram of the proposed algorithm framework.

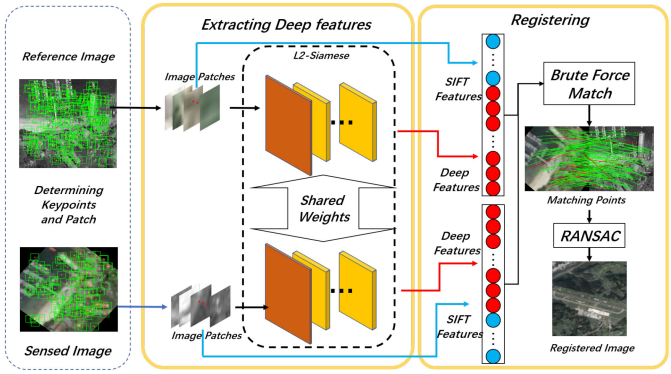


Fig. 1. Framework of the proposed algorithm.

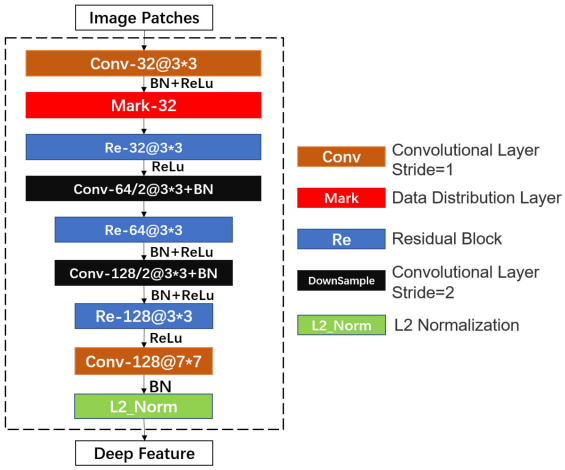


Fig. 2. Structure of the deep residual encoder.

B. Deep Residual Encoder (DRE)

We found a 28×28 patch size to be optimal by analyzing the registration results of different sizes by Wang *et al.* [25], and conducting the experiment in this study (in Section III). Fig. 2 shows the structure of the residual coding network designed in this study. The basic structure of the encoder is the convolutional [38]. The network down-samples the feature map twice through the convolution layer with a stride of two. The convolution layer with a step size of two can retain the characteristic information more effectively than maximum pooling. The feature map is divided into three parts according to the size, and each part includes a convolution layer (convolution kernel size: 3×3 and 7×7; activation function: ReLU) and a residual block. The network transforms the size of the feature map from $28 \times 28 \times 32$ to $14 \times 14 \times 64$ and $7 \times 7 \times 128$ through two down-samplings. Then, through the 7×7 convolution kernel, the feature map is converted into 128 vectors with the same size as SIFT descriptors. The size of the network output is 128.

L2 normalization layer: L2 normalization has been used mainly to perform dimensionless quantization on features. Homogenization of feature can solve the problem that the descriptors of two patches are quite different in nature. Dimensionless quantization can aid the comparison of features by changing each index value to the same order of magnitude, which is conducive to subsequent descriptor matching tasks. Moreover, experiments show that the L2 normalization layer can improve the convergence speed and accuracy of the model. Here, x represents the element of input X. The relationship between the input X and the output $Norm_{L2}$ is expressed as follows:

$$Norm_{L2} = \frac{\mathrm{X}}{\sqrt{\max(\sum_{i=0}^{128} x_i^2)}}. \quad (1)$$

Residual block: The residual block [39] is used to improve the ability of the network to extract information and make the network easier to optimize.

Data distribution layer: Because there is often a distortion between the images to be registered in remote sensing image registration, the neighborhood information of key points will also undergo a rotation transformation centering on the key points. At this time, the importance of each pixel of the matched patches is inversely proportional to the distance between the key points. Therefore, to improve the feature weights at the key points, we propose the use of templates to change the feature map weights of certain layers in the model. Fig. 3 shows the visualization of this template.

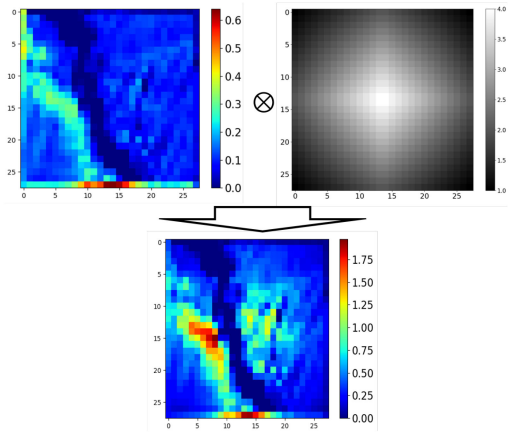


Fig. 3. Data distribution layer. Left: Feature maps of the output of the convolution operation of the first layer. Right: Proposed data distribution layer. Bottom: feature map after the data distribution operation.

C. L2-Siamese Model

In general, the feature map elements output by the middle layer of a deep network are difficult to understand, and the deep descriptors required for image registration must be suitable for the NNS algorithm. Furthermore, the distances between the descriptors of the matched key points are much smaller than those between other mismatched descriptors.

In the multiresolution remote sensing image registration task, the descriptors of key points must have high robustness to the rotation, scaling, and affine transformation of the image but need higher saliency for different key points. However, the deep features generated by the general DNN cannot be applied to the matching algorithm based on the NNS, so they do not have the basic characteristics of descriptors.

Therefore, it is impossible to extract the deep features of image patches solely through the DRE, and the optimization of the DRE network is a key problem. Using ideas from L2-Net [36] and Autoencoder, we optimize the DRE model structure and loss function in this study. We propose a Siamese network for deep descriptor extraction to ensure the robustness of the deep features extracted by the encoder through multiple transformations for the same image.

To improve the rotation invariance of deep features, we use the L2-Siamese network to train the residual encoder and to optimize the DRE. Fig. 4 shows the Siamese network model based on the residual encoder. In Section III, we discuss why the deep features obtained by the multiscale loss function is more suitable for image registration.

The L2-Siamese model first extracts the deep features of the image patches with the same weight through the DRE and then calculates their distance matrix. To avoid the gradient explosion of the model, the concatenation features are input to the three-layer fully connected layer, and the number of nodes is 64–32–2. The activation function of the first two layers is the ReLU; the last layer obtains the matching condition of the two image blocks. The activation function is SoftMax.

The learning of the local feature descriptor is a more specific problem than general image classification in ImageNet. This is because, compared with different objects of the same visual category, local patches can experience limited transformation [35]. Therefore, we propose to use the distance matrix of two deep feature descriptors to determine the optimization parameters of the model. In this study, the deep features are obtained after the L2 normalization layer of the DRE. The distance matrix $D_{N\times N}$ is expressed as follows:

$$D_{N\times N} = \sum_{axis=-1}\left[\left(\begin{bmatrix} f_1' & \cdots & f_N' \\ \vdots & \ddots & \vdots \\ f_1' & \cdots & f_N' \end{bmatrix}^T - \begin{bmatrix} f_1'' & \cdots & f_N'' \\ \vdots & \ddots & \vdots \\ f_1'' & \cdots & f_N'' \end{bmatrix}\right)^2\right] = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1N} \\ d_{21} & d_{22} & \cdots & d_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & \cdots & d_{NN} \end{bmatrix}_{N\times N} \tag{2}$$

where N is the value of a batch ($N = 48$ in this model). f_k' and f_k'' are 128-D descriptors by two image patches with $k = 1, 2, \ldots, N$. d_{ij} is the Euclidean distance between f_i'' and f_j''.

According to the distance matrix $D_{N\times N}$, to simulate the descriptor in the matching algorithm based on Euclidean distance, we determine the loss function L_{same} as the loss function of the matched image patches. When there is a matched label ($y^t = 0$), the minimum value is selected from each row and each column of the distance matrix $D_{N\times N}$ and is compared with the value of the matched feature. Ideally, the Euclidean distance between the matched descriptors is the smallest. The difference between the minimum value of the row and column and the value of the matched feature vector is calculated through L_{same} to improve

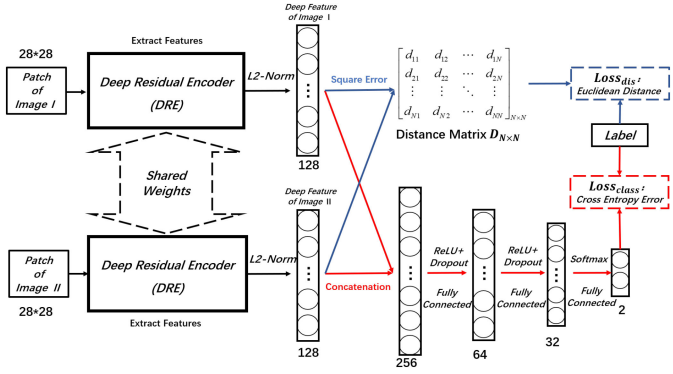


Fig. 4. L2-Siamese model based on the DRE. The model is used to train the residual encoder and generate the DRE model.

the significance of the descriptor:

$$L_{\text{same}} = -\sum_N \log(\exp(\min(d_{row}) - d_{ii})) - \sum_N \log(\exp(\min(d_{col}) - d_{ii})) \tag{3}$$

where d_{ii} is the Euclidean distance between the matched features, $\min(d_{\text{row}})$ is the minimum value of the row where d_{ii} is located, and $\min(d_{\text{col}})$ is the minimum value of the column where d_{col} is located. Here,$d_{ii} \geq \{\min(d_{row}), \min(d_{col})\}$ is always satisfied.

The Euclidean distance between matched and unmatched descriptors should be small and large, respectively. In addition, paired patches representing counter examples usually vary greatly, simplifying the learning process to optimize the distance. Because there is a large number of mismatched samples in the actual registration task, we select all nondiagonal elements in the distance matrix $D_{N\times N}$ as the sample design loss function L_{diff}, which is written as follows:

$$\text{diff}_{ij}^{\text{col}} = \frac{\exp(\max(0, 2 - d_{ij}))}{\sum_N \exp(\max(0, 2 - d_{cj}))}$$
$$\text{diff}_{ij}^{\text{row}} = \frac{\exp(\max(0, 2 - d_{ij}))}{\sum_N \exp(\max(0, 2 - d_{jn}))} \tag{4}$$

$$L_{\text{diff}} = -\frac{1}{2}\cdot\left(\sum_N \log(\text{diff}_{ij}^{col}) + \sum_N \log(\text{diff}_{ij}^{\text{row}})\right) \tag{5}$$

where d_{ij} are the off-diagonal elements in the distance matrix $D_{N\times N}$, and diff_{ij}^{col} and diff_{ij}^{row} are operations on the rows and columns in the distance matrix $D_{N\times N}$, respectively. The expected minimum Euclidean distance is 2.

After calculating the loss functions of the matching and non-matching image patches separately, we add them to generate the loss function $Loss_{dis}$ based on the distance matrix:

$$Loss_{\text{dis}} = L_{\text{same}} + L_{\text{diff}}. \tag{6}$$

The third part of the loss function ($Loss_{\text{class}}$) evaluates the difference between the probability distribution of the current training and the real distribution. The objective function is determined as the categorical cross-entropy loss function, which

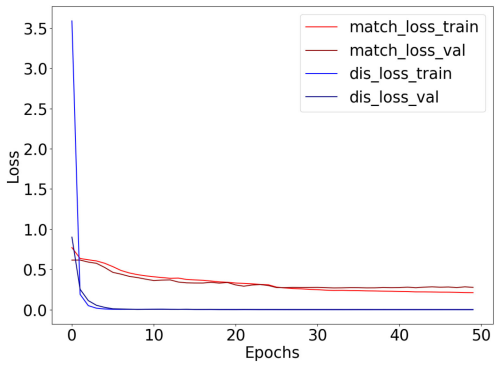


Fig. 5. Training variation curve based on the Siamese network model: change in the loss value of each part; $Loss_{\text{diss}}$ and $Loss_{\text{class}}$ are denoted as "dis-loss" and "match-loss," respectively.

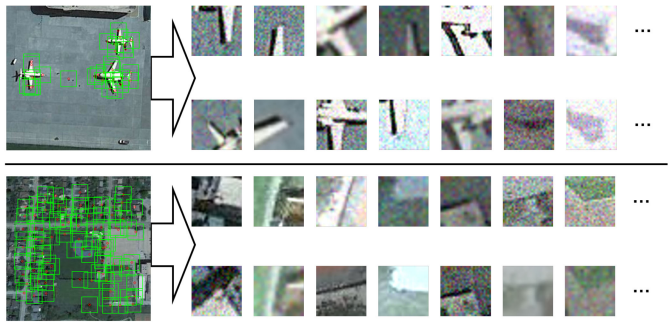


Fig. 6. Training images. These are two remote sensing images randomly selected from the NWPU-RESISC45 dataset. The images on the right and left sides of the arrow are patch pairs with the same key points and the images generated by transformation, respectively.

describes the distance between the actual output probability and expected probability. Therefore, a smaller cross-entropy value corresponds to closer probability distributions. $Loss_{\text{class}}$ enables the entire model to converge quickly during training, and it is given by

$$Loss_{\text{class}}(y^t, y^p) = -[y^p \cdot \log(y^t) + (1 - y^p) \cdot \log(1 - y^t)] \tag{7}$$

where $Loss_{\text{class}}$ represents the loss function between the actual output of the network y^p and the true value y^t ($y^t = 0$ when matching; $y^t = 1$ when not matching).

Finally, the total loss function of the model is as follows:

$$Loss(y^t, y^p, f', f'') = \lambda_1 \cdot Loss_{\text{diss}} + \lambda_2 \cdot Loss_{\text{class}} \tag{8}$$

where f' and f'' are the deep features output by two encoders with shared weights, and $\lambda_1 = 10^{-2}$ and $\lambda_2 = 5 \times 10^{-1}$ are weights in the model.

Adam optimizer is used in the model. The learning rate was found to fall from 0.1 to 0.001, and 50 training iterations were performed. The method of creating the training set required for the Siamese network training is introduced in the next section. Fig. 5 shows the loss change curves during the network training process.

D. Datasets Preparation

A large number of training datasets are required to train a DNN, and the performance of the model correlates with the scale and generality of the training datasets. Thus, a large number of image transformations is required to generate and expand datasets for strong generalization when the number of remote sensing images is small. The problem of insufficient images can be solved by generating low-resolution remote sensing images from each high-resolution remote sensing image by down-sampling ($\phi : R^{3\times t\cdot W\times t\cdot H} \rightarrow R^{3\times W\times H}$, where t is a random number between one and three). The high-resolution remote sensing images were selected from the NWPU-RESISC45 dataset [40].

A high-resolution remote sensing image is used as the reference image I_1, and the low-resolution image is the sensed image I_2; I_1 and I_2 are fully registered. The SIFT algorithm is used to determine the key points for I_1 and I_2, and the key point (x, y) is used as the center to cut 28×28 patches for I_1 and I_2, which are called $p^1_{x,y}$ and $p^2_{x,y}$, respectively. To ensure that the position of the center point does not change, $p^2_{x,y}$ is subjected to random rotational transformation, affine transformation, brightness transformation, distortion transformation, and noise addition to obtain $p^{2'}_{x,y}$. The rotation, zoom, affine, and brightness coefficients are controlled within $[0, 360]$, $[0.5, 2]$, $[-10, 10]$, and $[0.5, 2]$, respectively. The training samples are randomly divided into training and test sets in a ratio of $8:2$. Fig. 6 shows some of the training data.

We labeled patch pairs with the same and different key points as 0 and 1, respectively. After image transformation, the patches differed significantly in gray distribution and morphology, thereby ensuring that image transformation significantly improves the generality of the training data.

E. Registration

After the training of the L2-Siamese model, we trained the DRE network, which was then used to extract the key points needed for multiresolution remote sensing image registration and the depth characteristics of its neighborhood. In the registration process, the key points of the two images are first determined through the SIFT algorithm. The patches centered on the key points of the two images are intercepted and inputted to the DRE to extract its deep features. The deep features output by the model and the shallow features of the standardized SIFT descriptor, which have 128 dimensions, are concatenated to obtain a 256-D feature vector.

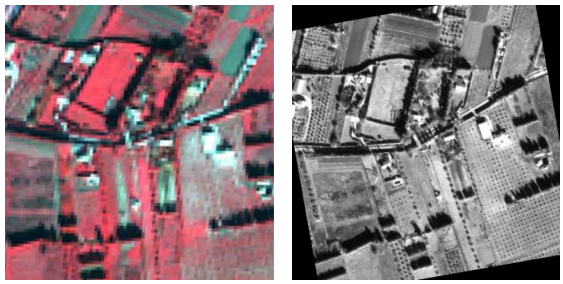
The final feature vector is used to determine the matching point through the distance-based brute force matching method. As the remote sensing image has a large number of targets in grayscale with similar textures, similar feature vectors will cause mismatched points. In this study, the random sample consensus algorithm (RANSAC) was employed to delete mismatched points globally.

F. Experimental Setting and Datasets

In the following subsections, we report the evaluation of the DRE proposed in this study, including the comparison with

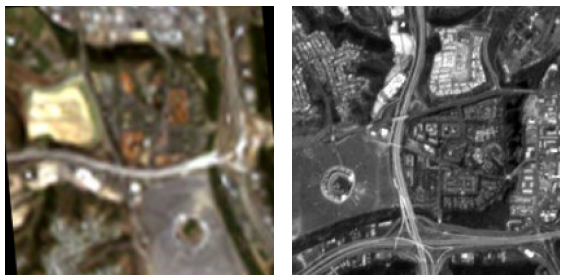
TABLE I
OVERVIEW OF THE EXPERIMENTAL DATA

Image	Size	Type	Sensor	Res
P-1	385×377	MS	QuickBird	2.44m
		PAN	QuickBird	0.61m
P-2	236×236	MS	ETM+	30m
		PAN	ETM+	15m
P-3	722×707	MS	MUX(ZY-3)	6m
		PAN	NAD(ZY-3)	2.1m
P-4	385×377	MS	PMS(GF-1)	8m
		PAN	PMS(GF-1)	2m
P-5	600×500	SAR	Radarsat-2	-
P-6	400×400	SAR	Radarsat-2	-



(a) (b)

Fig. 7. P-1 experimental image. (a) Low-resolution multispectral remote sensing image. (b) High-resolution panchromatic remote sensing image.

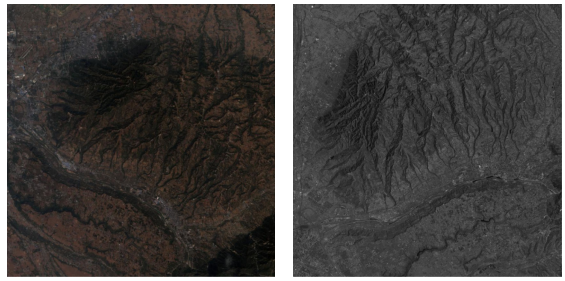


(a) (b)

Fig. 8. P-2 experimental image. (a) Low-resolution multispectral remote sensing image. (b) High-resolution panchromatic remote sensing image.

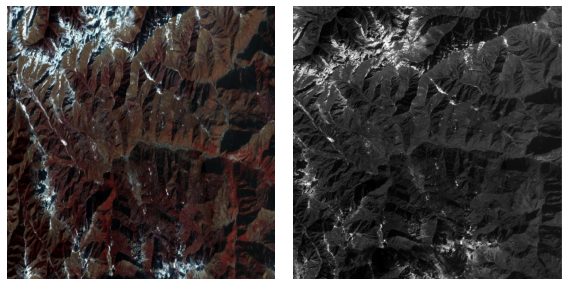
several state-of-the-art feature matching methods. The experiments were performed using different remote sensing images to evaluate the performance and robustness of the algorithm. Seven remote sensing registration algorithms were used as comparison groups: SIFT [14], SURF [16], FSC-SIFT [20], PSO-SIFT [17], SAR-SIFT [19], PSO-SIFT-CNN [23], and RF-Net [31].

Six experimental datasets were selected for the experiment: four sets of PAN and MS images, and two sets of SAR images. Among them, PAN and MS images prove the effect of the proposed multiresolution remote sensing image registration model, and SAR images serve as a supplementary experiment to prove the robustness of the proposed method. Table I lists the details of the experimental data.



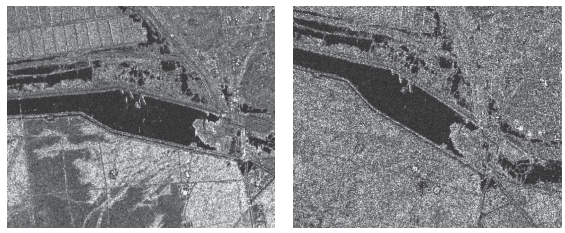
(a) (b)

Fig. 9. P-3 experimental image. (a) Low-resolution multispectral remote sensing image. (b) High-resolution panchromatic remote sensing image.



(a) (b)

Fig. 10. P-4 experimental image. (a) Low-resolution multispectral remote sensing image. (b) High-resolution panchromatic remote sensing image.



(a) (b)

Fig. 11. P-5 experimental image. (a) Low-resolution multispectral remote sensing image. (b) High-resolution panchromatic remote sensing image.

Figs. 7–12 show the experimental images. Owing to page layout limitations, the images shown in this article are scaled down. However, the image ratio is unchanged.

G. Evaluation Criteria

To experimentally compare the image registration performance of the proposed model, we employed the metrics proposed by Goncalves *et al.* [41] to evaluate the image registration results. The metrics are as follows.

1) N_{red}: number of control points.

2) RMS_{all}: root-mean-square error based on all control points (it should reach the sub-pixel level).

3) RMS_{Loo}: root-mean-square error computed by the control point residuals based on leave-one-out method. That is, calculate the RMS_{all} of the control points of $N_{\text{red}} - 1$ except for a certain pair of feature points, and then calculate the average value.

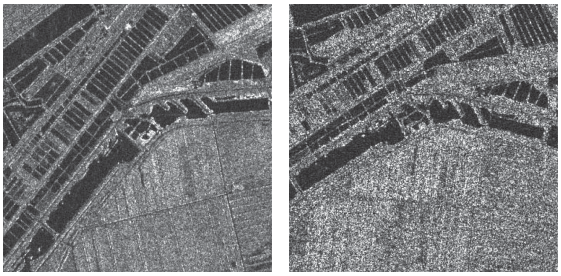


Fig. 12. P-6 experimental image. (a) Low-resolution multispectral remote sensing image. (b) High-resolution panchromatic remote sensing image.

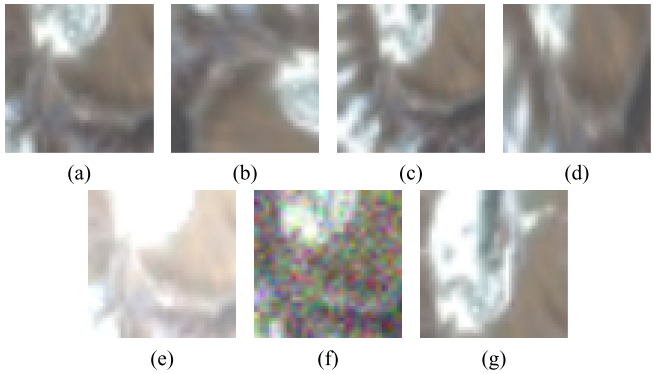


Fig. 13. Experimental images of robustness test for deep features. (a) Original. (b) Rotation. (c) Scaling. (d) Affine. (e) Brightness. (f) Noise. (g) Other images.

4) $P_{\text{quad}}(N_{\text{red}} > 20)$: statistical evaluation of residual distribution across the quadrant. The Chi-square distribution (degree of freedom: 1.0) is used to detect the distribution of feature points.

5) $BPP(1.0)$: bad points proportion with N_{red}. Points with a residual distance greater than 1.0 pixel are called bad points.

6) S_{kew}: statistical evaluation of preference axis on the residual scatterplot. When $N_{\text{red}} < 20$, use Spearman correlation coefficient; when $N_{\text{red}} \geq 20$, use Pearson correlation coefficient.

7) S_{cat}: statistical evaluation of the distribution of control points across the image.

8) $\emptyset$: weighted sum of the above seven measures. The weighted sum of the abovementioned seven measures $\emptyset$ is calculated as (9) and (10).

In the metrics present at the bottom of this page, except for N_{red}, the smaller the value of the other metrics, the better the performance. In addition, the qualitative evaluation was performed using a checkerboard image. The continuity of the image edges and overlapping region illustrate the registration performance.

H. Evaluation of the DRE Robustness

We investigated whether the deep features obtained by the DRE are more rotation-invariant and better suited for remote sensing image registration than those obtained with the autoencoder. Using the encoder with the same structure, the deep features obtained from the autoencoder were compared with the deep features obtained by the DRE. First, a key point was randomly selected in P-4. Then, the patch centered on it was intercepted, and various images of the patch were transformed through rotation, scaling, affinity and brightness change, and noise addition. Fig. 13 shows the transformation results.

The patches in 13 were inputted into the autoencoder and DRE to observe the distribution of their deep features. To distinguish the deep features of the patches of different key point neighborhoods, a patch from P-4 was randomly added as a reference in this experiment.

III. Experimental Results

Fig. 14 is a distribution map of deep features. It shows that the deep feature distribution of an image generated by the auto-encoder differs significantly after transformation, and its features are unsuitable for image registration. However, the distribution of deep features generated by the proposed DRE network is consistent, proving that the deep features obtained are more rotation-invariant and better suited for remote sensing image registration.

To verify the effect of patches of different sizes on the registration performance, we selected P-1 and P-6 as experimental data and conducted comparative experiments on patches of different sizes, including 20×20, 28×28, and 36×36. Subsequent to two down-samplings, the size of the convolution kernel of the final convolution layer was set to 5×5, 7×7, and 9×9. Finally, 128-D deep features were obtained. The experimental results are shown in Table II. The results indicate that the 28×28 patch size is the best; however, the difference effect produced by different patch sizes is minimal.

Some of compared algorithms are required to determine the ratio of the Euclidean distance between the nearest and second-nearest neighbors of the corresponding feature, denoted by $Dist_r$. In this experiment, $Dist_r$ from 0.7 to 0.95 was

$$N_{\text{red}} < 20:$$

$$\emptyset = \frac{2 \times (\frac{1}{N_{\text{red}}} + RMS_{\text{LOO}} + BPP(1.0) + S_{cat}) + RMS_{\text{all}} + 1.5 \times S_{\text{kew}}}{(2+2+2+2+1+1.5)} \tag{9}$$

$$N_{\text{red}} \geq 20:$$

$$\emptyset = \frac{2 \times (\frac{1}{N_{\text{red}}} + RMS_{\text{LOO}} + BPP(1.0) + S_{cat}) + RMS_{\text{all}} + 1.5 \times (P_{\text{quad}} + S_{\text{kew}})}{(2+2+2+2+1+1.5+1.5)}. \tag{10}$$

TABLE II
REGISTRATION PERFORMANCES OF P-1 AND P-2 IMAGES BY VARYING PATCH SIZE

Data	Size	$Dist_r$	N_{red}	RMS_{all}	RMS_{Loo}	P_{quad}	$BPP(1.0)$	S_{kew}	S_{cat}	$\emptyset$
P-1	20×20	0.90	101	0.551	0.551	0.432	0.070	0.070	0.999	0.380
	28×28	0.90	97	0.541	0.541	0.401	0.067	0.066	1.000	0.373
	36×36	0.90	88	0.545	0.545	0.421	0.056	0.065	1.000	0.375
P-2	20×20	0.85	140	0.537	0.538	0.135	0.025	0.162	0.999	0.343
	28×28	0.85	133	0.531	0.531	0.140	0.023	0.165	0.983	0.0.339
	36×36	0.85	120	0.535	0.535	0.141	0.029	0.165	0.989	0.345

TABLE III
QUANTITATIVE COMPARISON AMONG NINE METHODS ON P-1

Method	$Dist_r$	N_{red}	RMS_{all}	RMS_{Loo}	P_{quad}	$BPP(1.0)$	S_{kew}	S_{cat}	$\emptyset$
SIFT	0.75	112	0.993	0.993	0.852	0.277	0.047	0.999	0.575
SURF	0.80	108	1.435	1.435	0.054	0.463	0.237	0.576	0.570
SAR-SIFT	0.90	100	0.860	0.860	0.169	0.320	0.017	0549	0.383
PSO-SIFT-CNN	-	113	1.065	1.066	0.477	0.469	0.024	0.999	0.575
RF-Net	0.90	27	1.303	1.303	0.159	0.444	0.020	0.999	0.595
PSO-SIFT	0.90	114	0.585	0.585	0.113	0.035	0.072	1.000	0.343
FSC-SIFT	0.90	48	0.644	0.643	0.111	0.083	0.045	1.000	0.366
DRE*	0.90	99	0.594	0.595	0.489	0.088	0.078	1.000	0.402
DRE	0.90	97	0.541	0.541	0.401	0.067	0.066	1.000	0.373

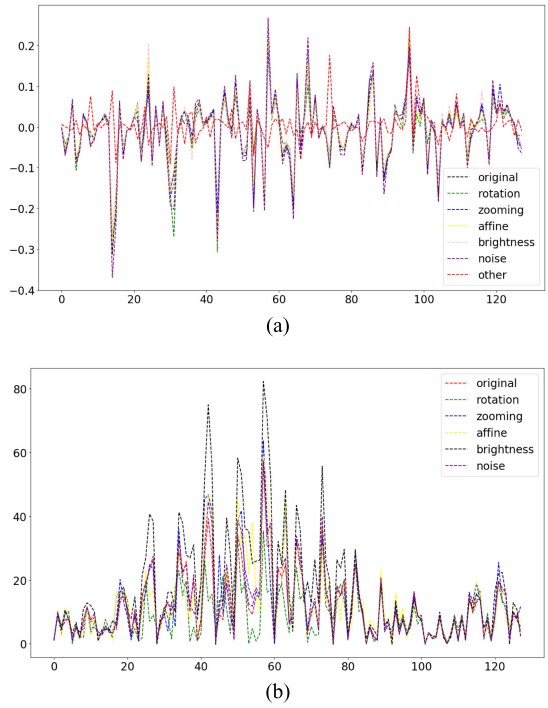


Fig. 14. Distribution map of deep features. (a) Deep features generated by the DRE. (b) Deep features generated by the autoencoder.

searched for the smallest RMS_{all} as its comparison parameter. We set d_r of the PSO-SIFT to 0.9. Because the features of the multiresolution remote sensing images vary significantly, there were numerous error points after registration. SIFT, SURF, and the proposed algorithm use RANSAC for error point screening.

Tables III–VIII show the experimental results of seven methods on P-1 to P-6, respectively. The symbol "–" in these tables indicates that the registration result does not meet the test conditions. DER* represents the DER model without mixed shallow features.

The proposed method achieved subpixel registration of the six pairs of experimental images, including multiresolution optical remote sensing images and SAR images. In addition, it achieved better performance on the six sets of experimental images.

In order to visually detect the performance of the image registration, the multispectral image P-2 and the SAR image P-6 with larger registration errors were selected as the data for the comparison experiment. Through the analysis of Table IV and Table VIII, PSO-SIFT and SAR-SIFT are the best models in the comparison method for P-2 and P-6 images, respectively. The comparison result is shown as a panel mosaic image, as shown in Fig. 15. It can be seen that the method proposed is continuous in edges and overlapping regions, while the frame area in the contrast experiment picture is not aligned. Since the registration and visualization results of other images have relatively small differences, this article does not compare and display other images. Fig. 16 directly shows mosaicked image results of other registered images.

IV. DISCUSSION

The DRE network was shown to extract deep features effectively. This is because the deep model can transform multiresolution data features to the same scale, significantly improving the robustness of the extracted features and achieving the registration of multiresolution remote sensing images. The results demonstrate that the proposed registration model can achieve subpixel registration. The relative registration accuracy improved by 1.6%–7.5%, whereas the overall performance improved by 4.5%–14.1%.

The shallow features used for the traditional manual design cannot effectively identify image data that are very similar but in a different form; this results in ineffective feature matching for

TABLE IV
QUANTITATIVE COMPARISON AMONG NINE METHODS ON P-2

Method	$Dist_r$	N_{red}	RMS_{all}	RMS_{Loo}	P_{quad}	$BPP(1.0)$	S_{kew}	S_{cat}	$\emptyset$
SIFT	0.75	35	1.286	1.286	0.401	0.486	0.078	0.999	0.634
SURF	0.90	47	1.433	1.433	0.369	0.511	0.140	0.999	0.677
SAR-SIFT	0.95	26	0.921	0.921	0.802	0.423	0.232	0.999	0.603
PSO-SIFT-CNN	-	29	1.381	1.380	0.621	0.517	0.014	0.999	0.682
RF-Net	-	-	-	-	-	-	-	-	-
PSO-SIFT	0.90	39	0.612	0.612	0.046	0.051	0.103	0.999	0.351
FSC-SIFT	0.80	34	0.701	0.701	0.374	0.088	1.172	1.000	0.430
DRE*	0.90	52	0.617	0.617	0.081	0.075	0.221	1.000	0.375
DRE	0.90	41	0.610	0.609	0.124	0.071	0.128	1.000	0.350

TABLE V
QUANTITATIVE COMPARISON AMONG NINE METHODS ON P-3

Method	$Dist_r$	N_{red}	RMS_{all}	RMS_{Loo}	P_{quad}	$BPP(1.0)$	S_{kew}	S_{cat}	$\emptyset$
SIFT	0.80	103	0.552	0.552	0.468	0.029	0.079	1.000	0.368
SURF	0.75	124	1.191	1.191	0.771	0.387	0.015	1.000	0.628
SAR-SIFT	-	-	-	-	-	-	-	-	-
PSO-SIFT-CNN	-	85	0.542	0.542	0.551	0.174	0.186	1.000	0.425
RF-Net	0.90	89	0.535	0.535	0.590	0.201	0.195	0.999	0.434
PSO-SIFT	0.90	25	0.593	0.593	0.239	0.004	0.318	0.999	0.392
FSC-SIFT	0.80	55	0.540	0.539	0.219	0.252	0.147	0.999	0.392
DRE*	0.85	101	0.560	0.561	0.210	0.051	0.192	0.995	0.366
DRE	0.90	133	0.531	0.531	0.140	0.023	0.161	0.983	0.339

TABLE VI
QUANTITATIVE COMPARISON AMONG NINE METHODS ON P-4

Method	$Dist_r$	N_{red}	RMS_{all}	RMS_{Loo}	P_{quad}	$BPP(1.0)$	S_{kew}	S_{cat}	$\emptyset$
SIFT	0.80	525	0.901	0.901	0.996	0.187	0.150	1.000	0.566
SURF	0.85	398	1.635	1.635	0.564	0.583	0.061	1.000	0.751
SAR-SIFT	0.95	289	0.896	0.896	0.049	0.343	0.078	1.000	0.464
PSO-SIFT-CNN	-	312	0.920	0.919	0.768	0.218	0.082	1.000	0.539
RF-Net	0.80	201	0.960	0.960	0.192	0.308	0.005	0.999	0.462
PSO-SIFT	0.95	15	0.519	0.518	-	0	0.768	0.999	0.461
FSC-SIFT	0.85	537	0.579	0.579	0.998	0.074	0.197	1.000	0.473
DRE*	0.90	540	0.520	0.521	0.440	0.201	0.102	1.000	0.397
DRE	0.90	576	0.507	0.507	0.435	0.173	0.003	1.000	0.377

TABLE VII
QUANTITATIVE COMPARISON AMONG NINE METHODS ON P-5

Method	$Dist_r$	N_{red}	RMS_{all}	RMS_{Loo}	P_{quad}	$BPP(1.0)$	S_{kew}	S_{cat}	$\emptyset$
SIFT	-	-	-	-	-	-	-	-	-
SURF	-	-	-	-	-	-	-	-	-
SAR-SIFT	0.90	52	0.836	0.836	0.291	0.269	0.214	1.000	0.487
PSO-SIFT-CNN	-	10	1.221	1.221	-	0.312	0.341	0.999	0.666
RF-Net	-	-	-	-	-	-	-	-	-
PSO-SIFT	0.90	22	0.541	0.541	0.576	0.091	0.015	0.999	0.398
FSC-SIFT	0.90	19	0.622	0.621	-	0.053	0.006	0.999	0.399
DRE*	0.90	42	0.401	0.401	0.236	0.043	0.210	0.999	0.333
DRE	0.90	45	0.392	0.392	0.230	0.071	0.263	0.999	0.341

TABLE VIII
QUANTITATIVE COMPARISON AMONG NINE METHODS ON P-6

Method	$Dist_r$	N_{red}	RMS_{all}	RMS_{Loo}	P_{quad}	$BPP(1.0)$	S_{kew}	S_{cat}	$\emptyset$
SIFT	0.85	16	1.498	1.497	-	0.812	0.277	0.999	0.824
SURF	-	-	-	-	-	-	-	-	-
SAR-SIFT	0.95	41	0.766	0.766	0.812	0.201	0.101	0.999	0.520
PSO-SIFT-CNN	-	18	1.212	1.212	-	0.412	0.312	0.999	0.670
RF-Net	0.90	21	1.016	1.016	0.752	0.324	0.225	0.999	0.600
PSO-SIFT	0.90	12	0.732	0.732	-	0	0.0.063	0.997	0.424
FSC-SIFT	-	-	-	-	-	-	-	-	-
DRE*	0.85	45	0.725	0.725	0.430	0.213	0.220	0.999	0.468
DRE	0.85	36	0.701	0.700	0.429	0.191	0.228	0.999	0.460

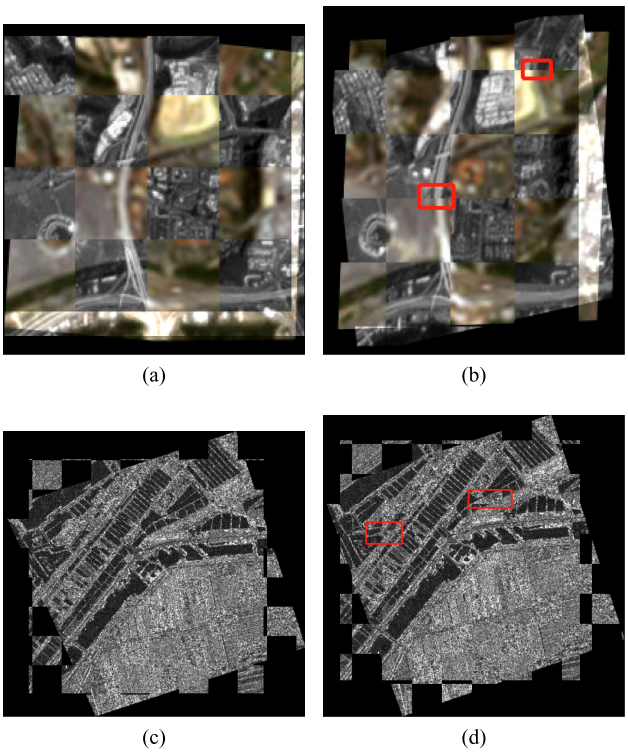


Fig. 15. (a) P-2 registration result based on DER model. (b) P-2 registration result based on PSO-SIFT model. (c) P-6 registration result based on DER model. (d) P-6 registration result based on SAR-SIFT.

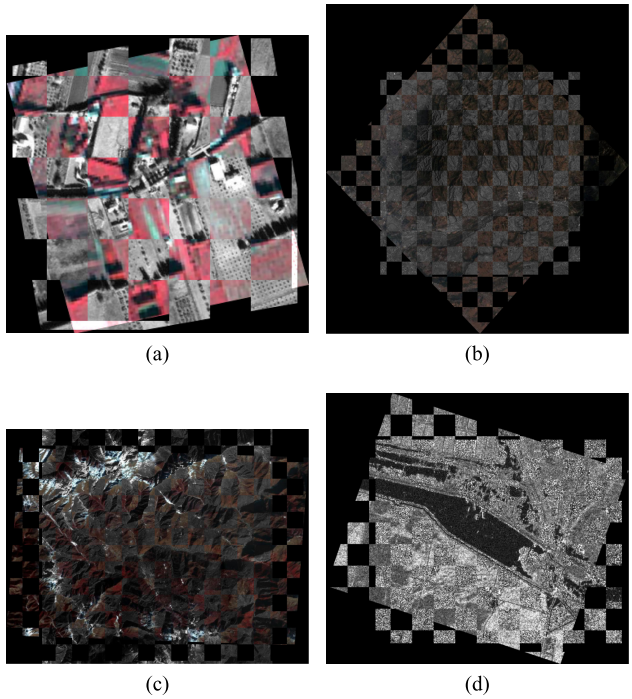


Fig. 16. The result of registration on the checkerboard mosaicked image: (a) Result of P-1. (b) Result of P-3. (c) Result of P-4. (d) The result of P-5.

two images with large differences. SIFT and SURF, which are the most commonly used methods, were shown to have relatively stable performance but poor matching accuracy. In most cases, the RMS_{all} value of the two methods is greater than 1, and subpixel registration is not achievable. PSO-SIFT performs well on optical remote sensing images, but its registration performance on SAR images lags behind the method proposed in this study. Conversely, SAR-SIFT performs poorly on optical images but performs better on SAR images. It can be seen that feature extraction algorithms based on manual design are significantly affected by the type of remote sensing image data.

The SIFT variants, such as PSO-SIFT, FSC-SIFT, and SAR-SIFT, have strong constraints on key point matching. These strong constraints can eliminate incorrect points better but might remove a large number of correct points on some images. PSO-SIFT performs better on P-1 and P-2. Nevertheless, in the experiments of P-3 and P-4, the value of N_{red} obtained by the algorithm is far lower than that obtained by the other algorithms. FSC-SIFT performs well in the optical image registration task, but, owing to its strong constraints, it has difficulty determining the matched point in SAR image registration. SAR-SIFT performs well on SAR images (P-5 and P-6) but poorly on multiresolution optical remote sensing images. It is worth noting that the N_{red} value of the SAR-SIFT algorithm on P-3 is insufficient to complete the registration task. The experimental results prove that it is difficult to determine the constraint strength and increase the robustness of the algorithm by simply improving the matching points screening capability of the algorithm.

During this research, we found that the design and training of deep feature extractors also directly affect the results of remote sensing image registration. The deep feature extractor used by PSO-SIFT-CNN is the VGG16 model for classification tasks. This kind of deep descriptor lacks a special design in the training process, so the deep features obtained by VGG16 have poor rotation invariance and are more sensitive to image deformation and rotation. Moreover, the model structure of VGG16 is larger, and the feature extraction requires considerable time. The experiments show that the performance of PSO-SIFT-CNN is poor.

RF-Net can not only extract the deep features of key points, but also automatically extract key points, which is the future development direction of image registration based on deep networks. Unfortunately, the performance of the RF-Net model is mediocre in the experimental data, especially when the image resolution is low and the feature difference is large.

This article presented a method to fuse deep and shallow features into multiresolution remote sensing image registration. To compare the registration performance of the hybrid feature and the single deep feature, a comparative experiment was conducted in the study, namely DER and DER*. The experimental results show that DER has better registration performance than DER*. These results prove that the hybrid feature is more significant and robust than the single deep feature.

V. Conclusion

We proposed an algorithm framework for generating deep features using a DRE combined with shallow features for multiresolution remote sensing image registration. The DRE was trained with the L2-Siamese model. Unlike traditional registration based on shallow features, the proposed deep features can

better utilize key point neighborhood information and possess strong robustness. The proposed algorithm achieved a relatively uniform feature expression for images of different resolutions in the same area.

The method based on the combination of deep features and shallow features for multiresolution remote sensing image registration performed better than state-of-the-art methods. The dataset required for the model training was designed to solve the problem of having only a few registered remote sensing images.

In the future, we plan to optimize the structure of deep feature extraction models and improve the robustness and saliency of deep features. In addition, we will attempt to establish the deep feature correspondence of multisource heterogeneous images and apply them to the registration of SAR and optical images.

REFERENCES

[1] B. Zitova and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, 2003.

[2] H. Chen, H. Zhang, J. Du, and B. Luo, "Unified framework for the joint super-resolution and registration of multiangle multi/hyperspectral remote sensing images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2369–2384, May. 2020.

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

[4] P. Zhang, M. Gong, L. Su, J. Liu, and Z. Li, "Change detection based on deep feature representation and mapping transformation for multi-spatial-resolution remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 24–41, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0924271616000563

[5] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosc. Remote Sens.*, vol. 48, no. 5, pp. 2403–2420, Mar. 2010.

[6] M. Chen and Z. Shao, "Robust affine-invariant line matching for high resolution remote sensing images," *Photogramm. Eng. Remote Sens.*, vol. 79, no. 8, pp. 753–760, Aug. 2013.

[7] T. Hu, H. Zhang, H. Shen, and L. Zhang, "Robust registration by rank minimization for multiangle hyper/multispectral remotely sensed imagery," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2443–2457, Jun. 2014.

[8] R. Feng, Q. Du, X. Li, and H. Shen, "Robust registration for remote sensing images by combining and localizing feature-and area-based methods," *ISPRS J. Photogramm. Remote Sens.*, vol. 151, pp. 15–26, Mar. 2019.

[9] J. Senthilnath, S. Omkar, V. Mani, and T. Karthikeyan, "Multiobjective discrete particle swarm optimization for multisensor image alignment," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1095–1099, Sep. 2013.

[10] H. Goncalves, L. Corte-Real, and J. A. Goncalves, "Automatic image registration through image segmentation and sift," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2589–2600, Jul. 2011.

[11] R. Feng, X. Li, and H. Shen, "Mountainous remote sensing images registration based on improved optical flow estimation," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 4, pp. 479–484, Jun. 2019.

[12] G. Brigot, E. Colin-Koeniguer, A. Plyer, and F. Janez, "Adaptation and evaluation of an optical flow method applied to coregistration of forest remote sensing images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 2923–2939, Jul. 2016.

[13] L. Zagorchev and A. Goshtasby, "A comparative study of transformation functions for nonrigid image registration," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 529–538, Mar. 2006.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[15] C. Wang, L. Wang, and L. Liu, "Progressive mode-seeking on graphs for sparse feature matching," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 788–802.

[16] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, Jun. 2008.

[17] W. Ma *et al.*, "Remote sensing image registration with modified sift and enhanced feature matching," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 3–7, Jan. 2017.

[18] S. Paul and U. C. Pati, "Remote sensing optical image registration using modified uniform robust sift," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1300–1304, Sep. 2016.

[19] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 453–466, Jan. 2015.

[20] Y. Wu, W. Ma, M. Gong, L. Su, and L. Jiao, "A novel point-matching algorithm based on fast sample consensus for image registration," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 43–47, Jan. 2015.

[21] N. Merkle, W. Luo, S. Auer, R. Müller, and R. Urtasun, "Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images," *Remote Sens.*, vol. 9, no. 6, p. 586, Jun. 2017.

[22] Z. Yang, T. Dan, and Y. Yang, "Multi-temporal remote sensing image registration using deep convolutional features," *IEEE Access*, vol. 6, pp. 38544–38555, Jul. 2018.

[23] F. Ye, Y. Su, H. Xiao, X. Zhao, and W. Min, "Remote sensing image registration using convolutional neural network features," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 232–236,Feb. 2018.

[24] H. Zhu, L. Jiao, W. Ma, F. Liu, and W. Zhao, "A novel neural network for remote sensing image matching," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2853–2865, Sep. 2019.

[25] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS J. Photogramm.*, vol. 145, pp. 148–164, Jan. 2018.

[26] X. Han, T. Leung, Y. Jia, R. Sukthankar, and A. C. Berg, "MatchNet: Unifying feature and metric learning for patch-based matching," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2015, pp. 3279–3286.

[27] H. Zhang *et al.*, "Registration of multimodal remote sensing image based on deep fully convolutional neural network," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3028–3042, Aug. 2019.

[28] W. Ma, J. Zhang, Y. Wu, L. Jiao, H. Zhu, and W. Zhao, "A novel two-step registration method for remote sensing images based on deep and local features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4834–4843, Jul. 2019.

[29] M. Dusmanu *et al.*, "D2-Net: A trainable CNN for joint detection and description of local features," 2019, in *Proc. IEEE Comput. Vision Pattern Recognit. (CVPR 2019)*, 2019.

[30] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2018, pp. 224–236.

[31] X. Shen *et al.*, "RF-Net: An end-to-end image matching network based on receptive field," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, 2019, pp. 8132–8140.

[32] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, 2005, pp. 539–546.

[33] H. He, M. Chen, T. Chen, and D. Li, "Matching of remote sensing images with complex background variations via siamese convolutional neural network," *Remote Sens.*, vol. 10, no. 2, p. 355, Feb. 2018.

[34] L. H. Hughes, M. Schmitt, L. Mou, Y. Wang, and X. X. Zhu, "Identifying corresponding patches in SAR and optical images with a pseudo-siamese CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 784–788, May 2018.

[35] E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, and F. Moreno-Noguer, "Discriminative learning of deep convolutional feature point descriptors," in *Proc. Int. Conf. Comput. Vis.*, Santiago, 2015, pp. 118–126.

[36] Y. Tian, B. Fan, and F. Wu, "L2-Net: Deep learning of discriminative patch descriptor in Euclidean space," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Honolulu, HI, 2017, pp. 661–669.

[37] A. Mishchuk, D. Mishkin, F. Radenovic, and J. Matas, "Working hard to know your neighbor's margins: Local descriptor learning loss," in *Proc. Adv. Neural Inform. Process. Syst.*, Dec. 2017, pp. 4826–4837.

[38] W. Xiong, B. Du, L. Zhang, L. Zhang, and D. Tao, "Denoising auto-encoders toward robust unsupervised feature representation," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Vancouver, BC, 2016, pp. 4721–4728.

[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 770–778.

[40] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state-of-the-art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Apr. 2017.

[41] H. Gonçalves, J. A. Gonçalves, and L. Corte-Real, "Measures for an objective evaluation of the geometric correction process quality," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 292–296, Apr. 2009.

Rongbo Fan received the bachelor's degree from the School of Automation, Harbin Engineering University, Harbin, China, in 2018. He is currently working toward the M.S. degree at Northwestern Polytechnical University, China.

His research interests include remote sensing image processing and deep learning.

Bochuan Hou received the master's degree in communication and information systems, from the Northwestern Polytechnical University, Xi'an, China, in 2018. He is currently working toward the Ph.D. degree at the Beijing University of Posts and Telecommunications, Beijing, China.

His research interests include Internet of vehicles security, mobile edge computing, blockchain, and communication security.

Jinbao Liu is currently working toward the Ph.D. degree at the Xi'an University of Technology, China.

He is an Engineer with the Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Company Ltd. His research interests include soil and vegetation spectrum analysis, image processing, machine learning, remote sensing, and signal processing.

Jianhua Yang received the B.S. degree from Xidian University, Xi'an, China, in 1989, the M.S. and Ph.D. degrees from the Northwestern Polytechnical University.

She is currently a Professer with Northwestern Polytechnical University, Xi'an, China, and also a Doctoral Supervisor. Her research interests include image processing, deep learning, bionic robot, sensor signal processing, detection, and control technology.

Zenglin Hong received the B.S. and the M.S. degrees from the Northwest Normal University, Lanzhou, China, in 1987 and 1999, respectively. He received the Ph.D. degree from the Northwestern Polytechnical University, Xi'an, China in 2007.

He is currently a Professor with the Northwestern Polytechnical University and Shaanxi Institute of Geological Survey, and also a Doctoral Supervisor. His research interests include resource management, regional economy, and systems engineering.