remote sensing

MDPI

Article

SD-CapsNet: A Siamese Dense Capsule Network for SAR Image Registration with Complex Scenes

Bangjie Li, Dongdong Guan *, Xiaolong Zheng, Zhengsheng Chen and Lefei Pan

High-Tech Institute of Xi'an, Xi'an 710025, China
* Correspondence: gdd@whu.edu.cn

Abstract: SAR image registration is the basis for applications such as change detection, image fusion, and three-dimensional reconstruction. Although CNN-based SAR image registration methods have achieved competitive results, they are insensitive to small displacement errors in matched point pairs and do not provide a comprehensive description of keypoint information in complex scenes. In addition, existing keypoint detectors are unable to obtain a uniform distribution of keypoints in SAR images with complex scenes. In this paper, we propose a texture constraint-based phase congruency (TCPC) keypoint detector that uses a rotation-invariant local binary pattern operator (RI-LBP) to remove keypoints that may be located at overlay or shadow locations. Then, we propose a Siamese dense capsule network (SD-CapsNet) to extract more accurate feature descriptors. Then, we define and verify that the feature descriptors in capsule form contain intensity, texture, orientation, and structure information that is useful for SAR image registration. In addition, we define a novel distance metric for the feature descriptors in capsule form and feed it into the Hard L2 loss function for model training. Experimental results for six pairs of SAR images demonstrate that, compared to other state-of-the-art methods, our proposed method achieves more robust results in complex scenes, with the number of correctly matched keypoint pairs (NCM) at least 2 to 3 times higher than the comparison methods, a root mean square error (RMSE) at most 0.27 lower than the compared methods.

Keywords: synthetic aperture radar (SAR); image registration; Siamese dense capsule network (SD-CapsNet)

check for updates

Citation: Li, B.; Guan, D.; Zheng, X.; Chen, Z.; Pan, L. SD-CapsNet: A Siamese Dense Capsule Network for SAR Image Registration with Complex Scenes. *Remote Sens.* **2023**, *15*, 1871. https://doi.org/10.3390/rs15071871

Academic Editor: Dusan Gleich

Received: 20 February 2023
Revised: 19 March 2023
Accepted: 20 March 2023
Published: 31 March 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

Synthetic aperture radar (SAR) is used in a variety of applications such as disaster monitoring [1], agricultural monitoring [2], and geological surveys [3] due to its all-weather, all-day Earth observation capability. These applications require a variety of technical tools to support them, such as change detection [4], image fusion [5] and 3D reconstruction [6]. Image registration plays a vital role in these technical tools. Image registration is the process of transforming multiple images acquired at different times or from different viewpoints into the same coordinate system. However, implementing robust and highly accurate geometric registration of SAR images is a challenging task because SAR sensors produce varying degrees of geometric distortion when imaging at different viewpoints. In addition, SAR images suffer from severe speckle noise.

Existing SAR image registration methods can be roughly divided into two categories, i.e., area-based methods and feature-based methods. The area-based methods, also known as intensity-based methods, realize registration of image pairs by a template-matching scheme. Classical template-matching schemes include mutual information (MI) [7] and normalized cross-correlation coefficient (NCC) [8]. However, these area-based methods generally require high computation cost. Although the computational cost can be reduced to a large extent by precoarse registration [9,10] or local search strategies [11,12], the area-based methods still perform poorly in scenes with large geometric distortions.

In contrast to the area-based methods, the feature-based methods achieve registration by extracting feature descriptors of keypoints, and their registration speed is usually better than that of the area-based methods. The feature-based methods can be further divided into handcrafted approaches and learning-based approaches. The traditional handcrafted approaches for SAR image registration include SAR-SIFT [13–15], speeded-up robust features (SURF) [16,17], KAZE-SAR [18], etc. Although these methods reduce the interference of speckle noise in SAR image registration, there is still a large number of incorrectly matched point pairs in SAR image registration with complex scenes. One reason is that due to the complexity of texture and structural information in SAR images with complex scenes, the accuracy and repeatability of the detected keypoints is not well supported. Eltanany et al. [19] used phase congruency (PC) and a Harris corner keypoint detector to detect meaningful keypoints in SAR images with complex intensity variations. Xiang et al. [20] proposed a novel keypoint detector based on feature interaction in which a local coefficient of variation keypoint detector was used to mask keypoints in mountainous and built-up areas. However, it is difficult to detect uniformly distributed keypoints in SAR images with complex scenes using these methods. Another reason is that the handcrafted feature descriptors are unable to extract significant features such as line or contour structures in complex scenes. With the development of deep learning, methods such as convolutional neural networks (CNNs) have achieved exciting success in the field of SAR image interpretation. Quan et al. [21] designed a deep neural network (DNN) for SAR image registration and confirmed that deep learning can extract more robust features and achieve more accurate matching. Zhang et al. [22] proposed a Siamese fully convolutional network to achieve SAR image registration and used the strategy of maximizing the feature distance between positive and hard negative samples to train the network model. Xiang et al. [20] proposed a Siamese cross-stage partial network (Sim-CSPNet) to extract feature descriptors containing both deep and shallow features, which increases the expressiveness of the feature descriptors.Fan et al. [23] proposed a transformer-based SAR image registration method, which exploits the powerful learning capability of the transformer to extract accurate feature descriptors of keypoints under weak texture conditions.

Although CNN-based methods extract feature descriptors that contain salient information useful for registration, such as structure, intensity, orientation, and texture information, CNNs improve feature robustness by virtue of "translation invariance" in convolution and pooling operations. "Translation invariance" means that the CNN is able to steadily extract salient features from an image patch even when the patch undergoes small slides (including small translations and rotations) [24]. However, this "translation invariance" is fatal in image registration, as it leads to small displacement errors in the matched point pairs, as shown in Figure 1. In addition, the feature descriptors obtained by the CNN-based methods are mostly described in a vector form. However, in traditional deep learning networks, the activation of one neuron can only represent one kind of information, and the singularity of its dimensionality dictates that the neuron itself cannot represent multiple kinds of information at the same time [25]. Therefore, the multiple pieces of information of keypoints and the relationship between the information hidden in a large number of network parameters pose many problems for SAR image registration. For instance, (1) the network requires a large number of training samples to violently decipher the relationships between multiple kinds of information. However, the scarcity of SAR image registration datasets often leads to incomplete training of models or causes data to fall into local optima. (2) The presence of complex non-linear relationships in feature descriptors detaches them from their actual physical interpretation, resulting in poor generalization performance. (3) The feature descriptors in vector form are not capable of representing fine information about keypoints in complex scenes. When two keypoints at different locations have similar structure and intensity information, their feature descriptors may be very similar in vector space, so that the registration algorithm is unable to recognize the differences between the feature descriptors, which may result in mismatched point pairs. As shown in Figure 2,

although the advanced registration algorithm Sim-CSPNet is used, keypoint *A* is still incorrectly matched to keypoint *C*.

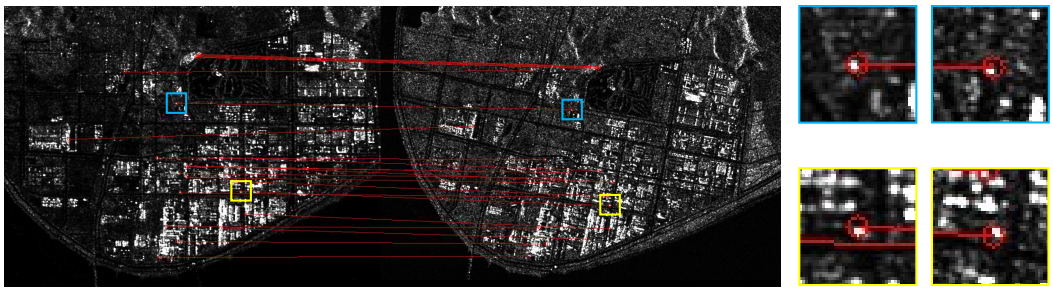
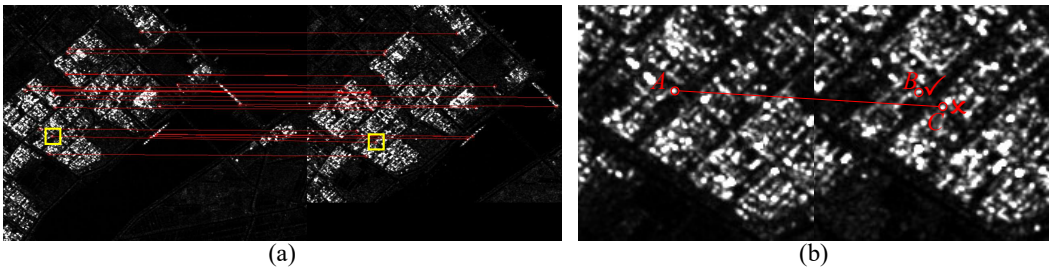


Figure 1. Illustration of small displacement errors in the matched point pairs obtained by Sim-CSPNet [20].



(a) (b)

Figure 2. Illustration of a mismatched point pair. (**a**) Registration results obtained by Sim-CSPNet [20]. (**b**) Enlarged regions of the mismatched point pair.

The advent of capsule network (CapsNet) [26] has broken the limitation that an active neuron in traditional neural networks can only represent certain kinds of information, achieving more competitive results than CNNs in various image application fields [25,27–29]. CapsNets encapsulate multiple neurons to form neuron capsules (NCs), enabling the evolution from "scalar neurons" to "vector neurons". In this case, a single NC can represent multiple messages at the same time, and the network does not need to additionally encode the complex relationships between multiple pieces of information. Therefore, CapsNets require far less training data than CNNs, whereas the effectiveness of the deep features is no less than that of CNNs. Despite the many advantages of CapsNets over CNNs, there are still some limitations to the application of CapsNets for image alignment tasks; therefore, no relevant research has been published.

One limitation is that the traditional CapsNet adopts a single convolutional layer to extract deep features, which is not sufficient to obtain significant features. The most direct solution is to add more convolutional layers to extract deep features [29,30]. However, the small amount of training data cannot support the training of a deep model, and the increased backpropagation distance leads to reduced convergence performance of the network. In addition, as mentioned earlier, most feature descriptors are in the form of vectors, whereas the feature descriptors extracted by CapsNets are in a special matrix form (the rows of the matrix represent the NCs, and the columns of the matrix represent a variety of useful information). Therefore, how to measure the distance between the feature descriptors is one of the key difficulties that constrain the application of CapsNet to image registration.

In this paper, we propose a novel registration method for SAR images with complex scenes based on a Siamese dense capsule network (SD-CapsNet), as shown in Figure 3. The proposed registration method consists of two main components, namely a texture constraint-based phase congruency (TCPC) keypoint detector and a Siamese dense capsule network-based (SD-CapsNet) feature descriptor extractor. In SAR data processing, phase information has been shown to be superior to amplitude information, reflecting the contour and structural information of the image [31,32]. However, the phase information contains

contour and structural information generated by geometric distortions such as overlays or shadows within complex scenes, which may affect the correct matching of keypoints. Therefore, we add a texture feature to the PC keypoint detector to constrain the keypoint positions. Specifically, when a keypoint has a large local texture feature, we consider that the keypoint may be located inside a complex scene, making it difficult to extract discriminative feature descriptors. Furthermore, considering the possible rotation between the reference and sensed images, the use of rotation-invariant texture features facilitates the acquisition of keypoints with high repeatability. Therefore, we adopt the rotation-invariant local binary pattern (RI-LBP) to describe the local texture features of the keypoints detected by the PC keypoint detector, and we discard a detected keypoint when its RI-LBP value is higher than the average global RI-LBP value. To break the limitation of traditional capsule networks in terms of deep feature extraction capability, we cascade three convolutional layers in front of the primary capsule layer and design a dense form of connection from the convolutional layer to the primary capsule layer, which makes the primary capsules contain both deep semantic information and shallow detail information. The dense connection form shortens the distance of loss backpropagation, which improves the convergence performance of the network and reduces the dependence of the model on training samples. In addition, in the parameter settings of most CapsNets, the dimension of the high-level capsules is greater than that of the primary capsules, allowing the high-level capsules to represent more complex entity information [29,30]. However, in our proposed method, high-level capsules are used as feature descriptors to describe important information of keypoints, such as structure information, intensity information, texture information, and orientation information; therefore, the feature descriptors extracted by SD-CapsNet require only four dimensions, greatly reducing the computational burden of the dynamic routing process. In addition, we define the L2-distance between capsules for feature descriptors in the capsule form. The effectiveness of the proposed method is verified using data for Jiangsu Province and Greater Bay Area in China obtained by Sentinel-1.

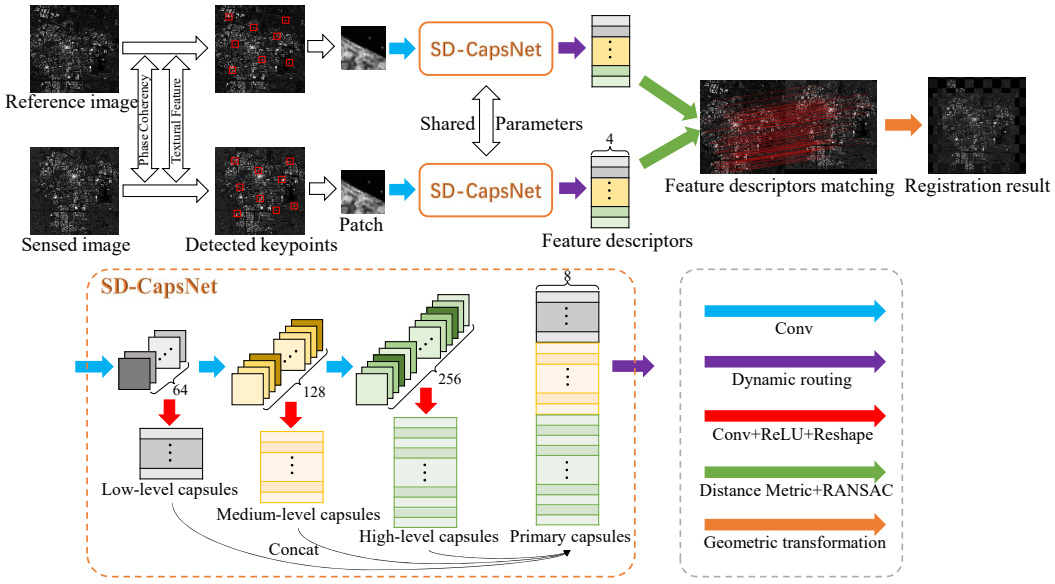


Figure 3. Schematic diagram of the proposed method.

The main contributions of this paper are briefly summarized as follows.

(1) We propose a texture constraint-based phase congruency (TCPC) keypoint detector that can detect uniformly distributed keypoints in SAR images with complex scenes and remove keypoints that may be located in overlay or shadow regions, thereby improving the high-repeatability of keypoints;

(2) We propose a Siamese dense capsule network (SD-CapsNet) to implement feature descriptor extraction and matching. SD-CapsNet designs a dense connection to construct

the primary capsules, which shortens the backpropagation distance and makes the primary capsules contain both deep semantic information and shallow detail information;

(3) We innovatively construct feature descriptors in the capsule form and verify that each dimension of the capsule corresponds to intensity, texture, orientation, and structure information. Furthermore, we define the L2 distance between capsules for feature descriptors in the capsule form and combine this distance with the hard L2 loss function to implement the training of SD-CapsNet.

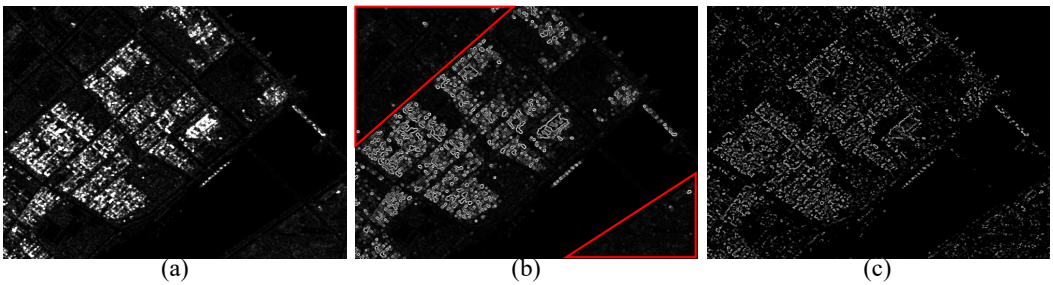
The rest of this paper is organized as follows. In Section 2, the proposed method is introduced. Section 3 presents the experimental results of our proposed method, as well as a comparisons with other state-of-the-art methods. Section 4 includes a discussion on the effectiveness of the keypoint detector and feature descriptor extractor. Finally, conclusions are presented in Section 5.

2. Methodology

Figure 3 shows the overall architecture of the proposed method. The proposed method consists of four steps: keypoint detection, feature descriptor extraction, feature descriptor matching, and geometric transformation. First, the PC keypoint detector is used to capture keypoints with significant structural features, followed by constrained keypoint detection results using the texture feature obtained by RI-LBP. Second, SD-CapsNet is proposed to extract feature descriptors of keypoints. Thirdly, a novel distance metric is defined to measure the distance between feature descriptors in the form of capsules. Finally, random sample consensus (RANSAC) is used to remove mismatched point pairs, and least squares is used to calculate the projection transformation matrix. Researchers generally agree that after removing mismatched point pairs by RANSAC or fast sample consensus (FSC), all that remains are correctly matched point pairs. In this section, we will introduce the details and contributions of the first three modules.

2.1. Texture Constraint-Based Phase Congruency Keypoint Detector

Many current keypoint detectors are based on gradient information, e.g., DoG [33], Sobel [34], Harris [35], etc. However, these operators are unable to extract accurate and well-distributed keypoints in weakly textured regions of SAR images, as shown in Figure 4b. As shown in Figure 4c, PC is a detector based on local phase information and is therefore able to extract structural information in weakly textured regions of SAR images.



(a) (b) (c)

Figure 4. Comparison of phase congruency and gradient. (**a**) SAR image. (**b**) Gradient feature. (**c**) Phase congruency feature.

PC can be calculated on multiple scales and in multiple directions using log Gabor wavelets according to the following formula [36].

$$\begin{cases} \mathrm{PC}(x) = \frac{\Sigma_n W(x) \lfloor A_n(x) \Delta\phi_n(x) - T \rfloor}{\Sigma_n A_n(x) + \varepsilon} \\ \Delta\phi_n(x) = \cos(\phi_n(x) - \bar{\phi}(x)) - |\sin(\phi_n(x) - \bar{\phi}(x))|, \end{cases} \tag{1}$$

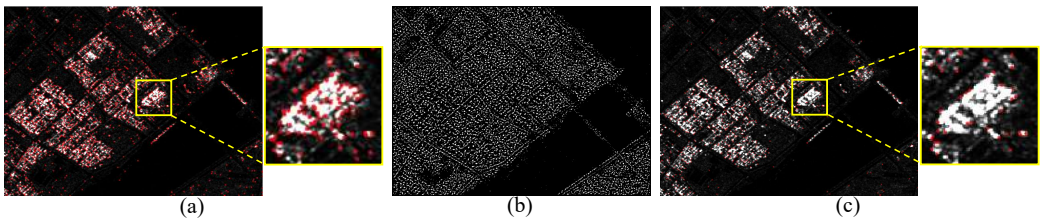
where A_n and $\phi_n(x)$ are the amplitude and phase angle of the n-th Fourier component at location x, T is the estimated noise threshold, $W(x)$ is the weighting factor for the frequency spread, ε is a small constant to avoid division by zero, the symbol $\lfloor\ \rfloor$ denotes that the

enclosed quantity is equal to itself if its value is positive and zero otherwise, and $\bar{\phi}(x)$ is the mean phase angle.

Following the classical moment analysis formulas, the maximum (M) and minimum (m) moment of PC are computed as

$$\begin{aligned} M &= \tfrac{1}{2}\left[a + c + \sqrt{b^2 + (a-c)^2}\right], \\ m &= \tfrac{1}{2}\left[a + c - \sqrt{b^2 + (a-c)^2}\right], \\ &\begin{cases} a = \sum [\mathrm{PC}(\theta)\cos(\theta)]^2 \\ b = 2\sum [\mathrm{PC}(\theta)\cos(\theta)] \cdot \sum [\mathrm{PC}(\theta)\sin(\theta)] \\ c = \sum [\mathrm{PC}(\theta)\sin(\theta)]^2 \end{cases} \end{aligned} \tag{2}$$

where $\mathrm{PC}(\theta)$ denotes the phase congruency value determined at orientation θ. The maximum and minimum moments of PC represent the edge and corner strength information, respectively. In this paper, we select the first 10,000 points of the sum of the PC maximum and minimum moment as the proposal keypoints, as shown in Figure 5a. Although the proposal keypoints are well distributed at the edges and corners, where the structural information is strong, many keypoints are also detected in the interior of complex scenes. However, complex scenes of SAR images may have varying degrees of geometric distortion (such as overlays or shadows) at different viewpoints, resulting in small possible displacement deviations between keypoint pairs. Therefore, we propose constraining the keypoint positions using texture information. Considering the possible rotation between the reference image and the sensed image, the RI-LBP is used to describe the local texture features of the keypoints.



(a) (b) (c)

Figure 5. Visualization of keypoint detection results. (**a**) Proposal keypoints obtained by the PC detector. (**b**) Texture features obtained by RI-LBP. (**c**) Keypoints obtained by the TCPC detector.

LBP is an operator used to describe the local texture features of images, which can be written as [37]

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3}$$

where g_c and g_p denote the gray values of the central pixel and its neighbors, respectively; R and P denote the radius of the circularly neighborand the number of neighbors, respectively; and p is the index of the neighbor.

Ojala et al. [38] proposed a rotation-invariant local binary pattern (RI-LBP) to extract a rotation-invariant texture feature.

$$LBP^{ri}_{P,R} = \min\{ROR(LBP_{P,R}, k) \mid k = 0, 1, \ldots, P-1\} \tag{4}$$

where $ROR(x|k)$ indicates that the binary number (x) is cyclically shifted to the right k times.

Finally, the proposal keypoints obtained by the PC detector are identified as the final keypoints by RI-LBP according to the following rules.

$$KP_{final} = KP_i \times \left\lfloor \overline{LBP^{ri}_{P,R}} - LBP^{ri}_{P,R}(KP_i) \right\rfloor, KP_i \in KP_{proposal} \quad (5)$$

where KP_{final} and $KP_{proposal}$ denote the final keypoints and proposal keypoints, respectively; $\overline{LBP^{ri}_{P,R}}$ denotes the global average value of the RI-LBP feature; and the symbol $\lfloor\ \rfloor$ denotes that the enclosed quantity is equal to itself if its value is positive and zero otherwise. Then, the keypoints inside complex scenes are removed, as shown in Figure 5c.

2.2. Siamese Dense Capsule Network-Based Feature Descriptor Extractor

In CNN-based feature descriptor extractors, various pieces of useful information are mixed together, resulting in feature descriptors losing their actual physical significance. CapsNet-based feature descriptor extractors encapsulate multiple neurons into neuron capsules that can represent multiple kinds of information at the same time. Figure 6 shows the difference between two types of feature descriptors. In this paper, we divide the information useful for SAR image registration into four categories, i.e., structure, intensity, texture, and orientation information. In the feature descriptors extracted by CapsNet-based methods, each row represents a capsule that contains various types of information, and each column represents a type of information that contains multiple features of this type of information.

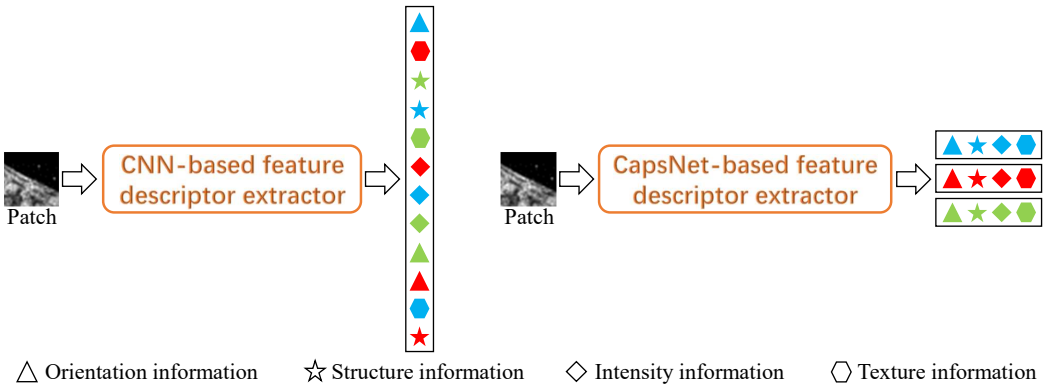


Figure 6. Illustration of the difference between two types feature descriptors. (The different shapes represent different types of information, and the different colors represent different features).

Inspired by a densely connected convolutional network (DenseNet) [39] and a cross-stage partial network (CSPNet) [40], we design a new way to connect capsule networks, whereby the convolutional layers are able to connect directly to the primary capsule layer. As shown in Figure 7a, in the traditional connection of CapsNets, losses need to be backpropagated layer by layer, and the backpropagation distance increases with the number of layers in the network. However, in our proposed densely connected CapsNet, the loss of the primary capsule layer can go straight to each convolutional layer, greatly reducing the backpropagation distance, and the backpropagation distance does not increase with the number of convolutional layers, as shown in Figure 7b. In addition, the primary capsules of SD-CapsNet can contain both detailed information from low-level convolutional layers and salient information from high-level convolutional layers.

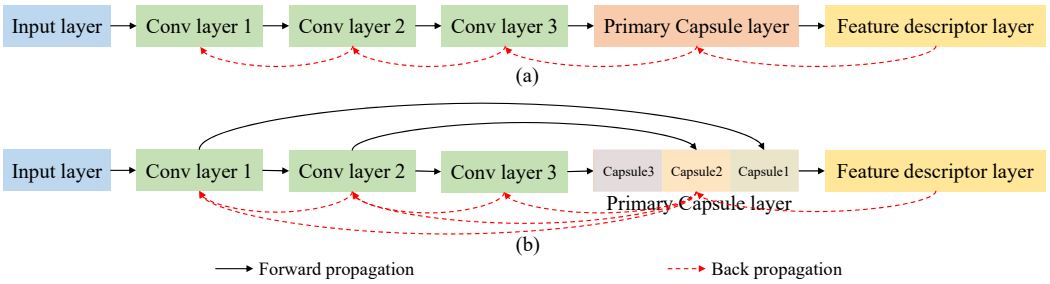


Figure 7. Illustration of the backpropagation pathway. (**a**) Traditional connected pathway. (**b**) The proposed densely connected pathway.

Next, we use a dynamic routing mechanism to deliver the output of the primary capsules to the feature descriptors. A diagram of the dynamic routing mechanism is shown in Figure 8. In the dynamic routing mechanism, the process of transferring the primary capsule (u_i) to the feature descriptor (f_j) is defined as

$$f_j = \sum_i c_{ij} \widehat{u}_{j|i} \tag{6}$$

where $\widehat{u}_{j|i} = W_{ij} u_i$. W_{ij} denotes the matrix of connection weights from the i-th capsule to the j-th capsule, c_{ij} represents coupling coefficient, which can be written as

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \tag{7}$$

where b_{ij} represents the prior probability of the i-th capsule relative to the j-th capsule.

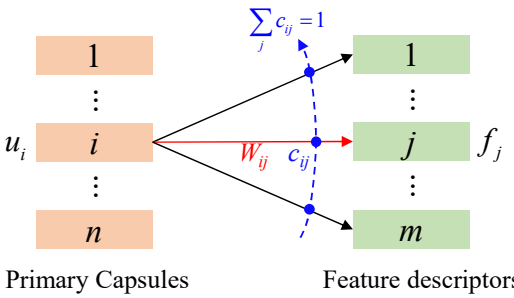


Figure 8. Illustration of the dynamic routing mechanism.

In most CapsNets, the dimensions of the high-level capsules are much larger than those of the primary capsules so that the high-level capsules represent more complex information. However, information with too many dimensions may be divorced from the actual physical interpretation and place an additional computational burden on the network. In SAR image registration tasks, the information that positively influences the registration results can be divided into four categories, i.e., intensity, orientation, texture, and structure information. Therefore, we set the information dimension of the feature descriptors to 4, and for consistency with the dimensions of other feature descriptors, the feature dimension is set to 32. Now, the size of the feature descriptors extracted by the proposed method is 32×4, as shown in Figure 9a. In other words, the common vector form (128×1) of the feature descriptors is replaced with a specific matrix form (32×4).

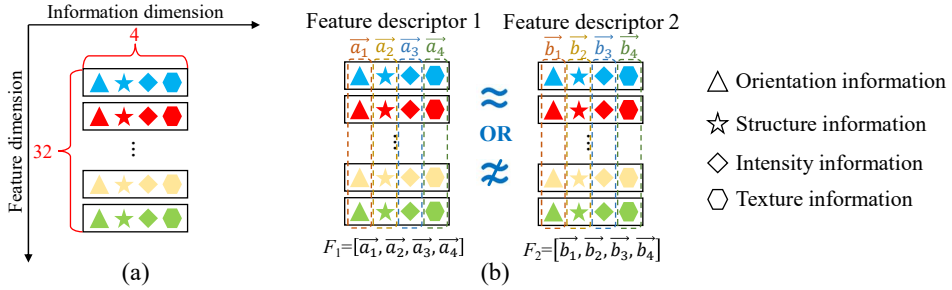


Figure 9. Illustration of the feature descriptor form. (**a**) The size of the feature descriptor. (**b**) The mathematical representation of feature descriptors.

The goal of training is to maximize the distance between the positive matches and hard negative matches. Before calculating the loss function, the distance metric between the feature descriptors in capsule form needs to be determined. Due to the special form of feature descriptors, the useful information of keypoints is confused after capsules are reshaped into one-dimensional vectors. Therefore, classical Euclidean distance, L2 distance, and cosine distance are not applicable in capsule networks. For the two given feature descriptors ($F_1 = \left[\vec{a_1}, \vec{a_2}, \vec{a_3}, \vec{a_4}\right]$ and $F_2 = \left[\vec{b_1}, \vec{b_2}, \vec{b_3}, \vec{b_4}\right]$), the distance between F_1 and F_2 can be defined as

$$\begin{aligned} d(F_1, F_2) &= \|F_1 - F_2\|_2 \\ &= \left\| \vec{a_1} - \vec{b_1}, \vec{a_2} - \vec{b_2}, \vec{a_3} - \vec{b_3}, \vec{a_4} - \vec{b_4} \right\|_2 \\ &= \left| \sqrt{\odot}\left(\sum_{i=1}^{4} \left(\vec{a_i} - \vec{b_i} \right)^2 \right) \right| \\ &= \left| \sqrt{\odot}\left(\sum_{i=1}^{4} \left(\left(\vec{a_i} - \vec{b_i} \right) \odot \left(\vec{a_i} - \vec{b_i} \right) \right) \right) \right| \end{aligned} \tag{8}$$

where $\odot$ denotes the Hadamard product, which can be expressed as $\vec{z} = \vec{x} \odot \vec{y} = [x_1 y_1, x_2 y_2, \ldots, x_n y_n]$. Similarly, $\sqrt{\odot}$ can be expressed as $\vec{p} = \sqrt{\odot}(\vec{z}) = [\sqrt{z_1}, \sqrt{z_2}, \ldots, \sqrt{z_n}]$. $|\cdot|$ denotes the module of the vector contained therein.

Then, we feed the distance function between feature descriptors into the hard L2 loss function [41]. For a given batch of SAR image pairs $((R_i, S_i)_{i=1\cdots n})$, feature descriptors $\left(\left(F_i^R, F_i^S\right)_{i=1\cdots n}\right)$ are extracted by SD-CapsNet. The non-matched feature descriptor $(F_{j_{\min}}^S)$ closest to F_i^R can be calculated as

$$j_{\min} = \arg\min_{j=1\cdots n, j\neq i} d\left(F_i^R, F_j^S\right) \tag{9}$$

The non-matched feature descriptor $(F_{k_{\min}}^R)$ closest to F_i^S can be calculated as

$$k_{\min} = \arg\min_{k=1\cdots n, k\neq i} d\left(F_k^R, F_i^S\right) \tag{10}$$

Finally, the triplet margin loss can be written as

$$L = \frac{1}{n}\sum_{i=1}^{n} \max\left\{0, 1 + d\left(F_i^R, F_i^S\right) - \min\left(d\left(F_i^R, F_{j_{\min}}^S\right), d\left(F_{k_{\min}}^R, F_i^S\right)\right)\right\} \tag{11}$$

As mentioned earlier, feature descriptors in capsule form contain intensity, texture, structure and orientation information. To verify this conclusion, we reconstruct the feature descriptors as images and apply separate perturbations to each type of information during the reconstruction process. The reconstruction process is implemented by a decoder, which

contains three fully connected layers, the structure and parameters of which are shown in Table 1.

Table 1. Model specifications and parameters of the decoder.

Block	Layer	Input Size	Output Size
Input layer	One column of the feature descriptor		32 × 1
Fully connected layer	Linear (32, 256); "RelU"	32 × 1	256 × 1
Fully connected layer	Linear (256, 512); "RelU"	256 × 1	512 × 1
Fully connected layer	Linear (512, 1024); "Sigmoid"	512 × 1	1024 × 1
Output layer	Reshape	1024 × 1	32 × 32

As shown in Figure 10, we add perturbations in the magnitude range of [−0.25, 0.25] to each column of the feature descriptor with a perturbation interval of 0.05. When perturbation is added to the column where the intensity information is located, the intensity information of the SAR image changes globally. When a perturbation is added to the column where the texture information is located, the texture information of the SAR image is changed locally. We add perturbations to the column where the orientation information is located, and as the perturbations are added, the SAR image orientation is gradually rotated. We add perturbations to the column where the structure information is located, and as the perturbations are added, the structure information of the SAR image changes to different degrees.

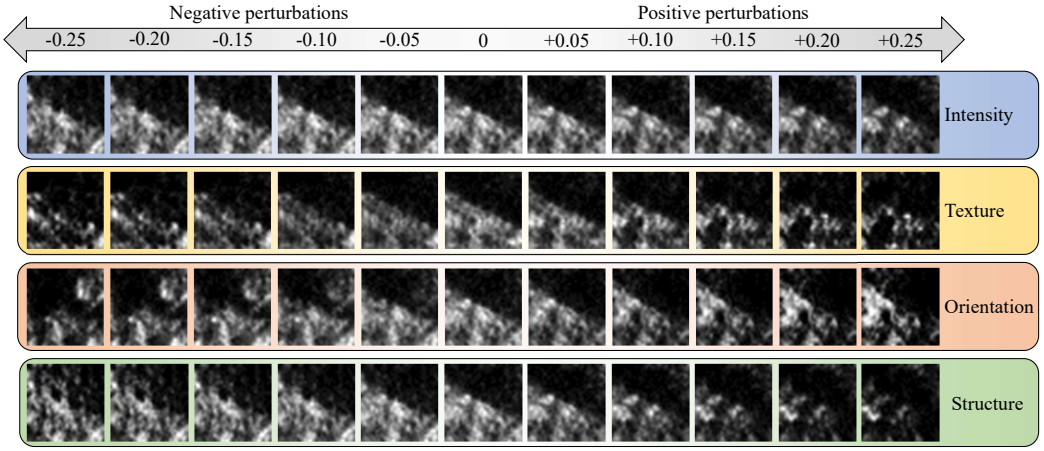


Figure 10. Visualization of reconstruction results for feature descriptors under different perturbations.

3. Experimental Results and Analysis

3.1. Data Description and Parameter Settings

In this paper, we collect large-scale SAR image pairs of four scenes for the production of a training dataset, as shown in Figure 11. The training dataset comprises C-band data from the Gaofen-3 system. To completely separate the test data from the training dataset, six SAR image pairs in the C-band from Sentinel-1 system are selected for testing. The incidence angle of the reference image is approximately 30.73° in the near range and 46.07° in the far range. The incidence angle of the sensed image is approximately 30.71° in the near range and 46.05° in the far range. Both the reference and sensed images have the same resolution of 10 m × 10 m. Detailed information about the six SAR image pairs is listed in Table 2.

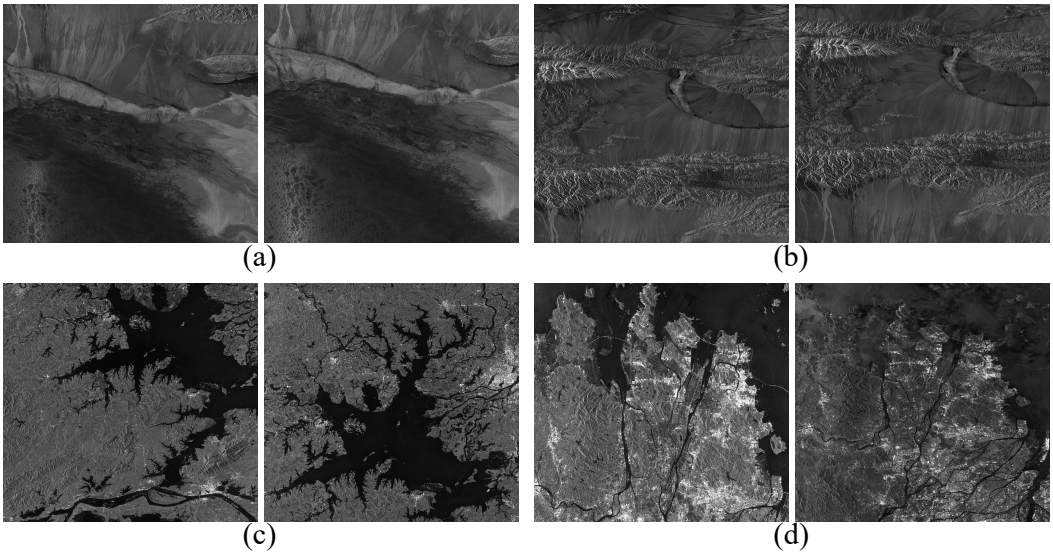


Figure 11. The training dataset of SAR images obtained by the Gaofen-3 system. (**a**) Desert. (**b**) Mountain. (**c**) Lake. (**d**) Urban.

Table 2. Detailed Information about SAR image pairs used for testing.

Site	Size	Resolution (Range × Azimuth)
Pair 1	836 × 583 663 × 488	10 m × 10 m
Pair 2	1027 × 752 1339 × 911	10 m × 10 m
Pair 3	1095 × 1201 1171 × 1143	10 m × 10 m
Pair 4	918 × 683 976 × 666	10 m × 10 m
Pair 5	1135 × 961 945 × 916	10 m × 10 m
Pair 6	500 × 499 600 × 600	11 m × 14 m

In our experiments, the PC is calculated at three scales and in six orientations, and the scaling factor between successive filters is set to 1.6. The RI-LBP is calculated in a circular neighborhood with a radius equal to three ($R = 3$), and the number of sampled neighbors is set to eight ($P = 8$). The threshold of RANSAC is set to 0.8. The number of kernels in the three convolutional layers is set to 64, 128, and 256, respectively. The dimension of the primary capsules is set to eight. "Adam" is employed as an optimizer, and the learning rate is set to 0.001. The detailed model specifications and parameters are listed in Table 3. The basic experimental environment settings are listed in Table 4. The comparison methods are introduced briefly as follows.

- SAR-SIFT [13] proposed SAR-Harris instead of DoG to detect keypoints and used the circular descriptor to describe spatial context information;
- SOSNet [42] used second-order similarity regularization for local descriptor learning, and proposed SOSLoss for registration model training with a small training sample;
- Sim-CSPNet [20] proposed a feature interaction-based keypoint detector and used a Siamese cross-stage partial network to generate feature descriptors.

Table 3. Model specifications and parameters of the SD-CapsNet.

Block	Layer	Input Size	Output Size (Width × Height × Channel)
Input layer			32 × 32 × 1
Conv layer 1	Conv (3 × 3), stride (1), ReLU	32 × 32 × 1	32 × 32 × 64
Low-level capsule Layer	Conv (9 × 9), stride (2), ReLU Reshape	32 × 32 × 64 12 × 12 × 64	12 × 12 × 64 1152 × 8
Conv layer 2	Conv (3 × 3), stride (1), ReLU	32 × 32 × 64	32 × 32 × 128
Medium-level capsule Layer	Conv (9 × 9), stride (2), ReLU Reshape	32 × 32 × 128 12 × 12 × 128	12 × 12 × 128 2304 × 8
Conv layer 3	Conv (3 × 3), stride (1), ReLU	32 × 32 × 128	32 × 32 × 256
High-level capsule layer	Conv (9 × 9), stride (2), ReLU Reshape	32 × 32 × 128 12 × 12 × 256	12 × 12 × 256 4608 × 8
Primary capsule layer	Concat	1152 × 8, 2304 × 8, 4608 × 8	8064 × 8
Feature descriptor layer	Dynamic routing	8064 × 8	32 × 4

Table 4. The basic experimental environment settings.

Platform	Linux
Torch	V 1.10.1
CPU	Intel(R) Xeon(R) Silver 4210R
Memory	64G
GPU	Nvidia GeForce RTX 3090
Video memory	24G

3.2. Registration Results

To evaluate the registration performance of the proposed method, we provide the keypoint matching results with different methods, as shown in Figures 12 and 13. Table 5 presents the quantitative comparison results. Note that for a fair comparison, SOSNet, Sim-CSPNet, and the proposed SD-CapsNet use the same training dataset.

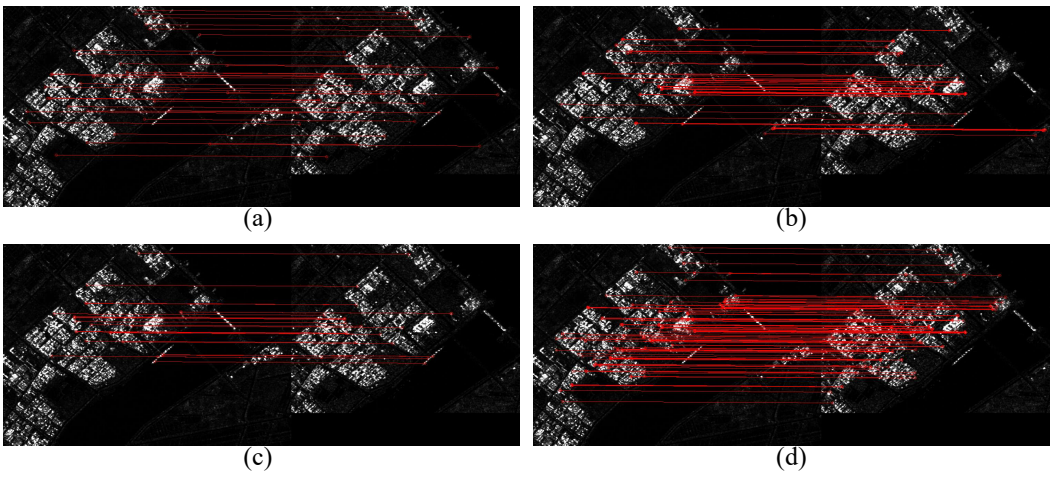


Figure 12. The feature descriptor matching results with different methods for pair 1. (**a**) SAR-SIFT. (**b**) SOSNet. (**c**) Sim-CSPNet. (**d**) Our proposed SD-CapsNet.

As shown in Figure 12, the distribution of matched keypoint pairs is non-uniform in the results of SOSNet and Sim-CSPNet, and the number of correctly matched keypoint pairs (NCM) in SAR-SIFT results is low. However, our proposed method obtains the most uniform and the largest number of correctly matched keypoint pairs. Figure 13 shows that

the NCM of the registration results obtained by SAR-SIFT, SOSNet, and Sim-CSPNet is much lower than that of the registration results obtained by our proposed method because the comparison methods are not adjusted for complex scenes and therefore cannot be adapted to the registration task of SAR images with complex scenes. From Figures 14–17, we can obtain the same conclusion that the proposed method is able to obtain the most uniform and the largest number of correctly matched keypoint pairs.

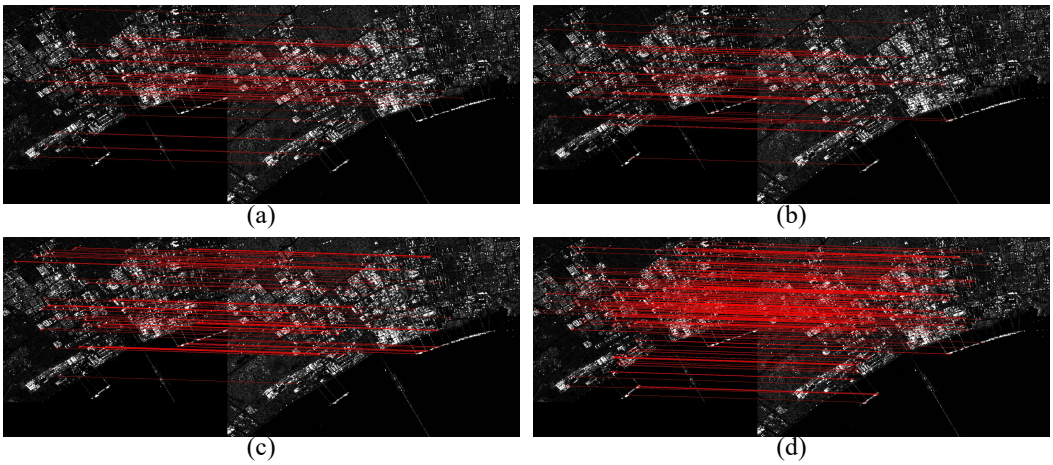


Figure 13. The feature descriptor matching results with different methods for pair 2. (**a**) SAR-SIFT. (**b**) SOSNet. (**c**) Sim-CSPNet. (**d**) Our proposed SD-CapsNet.

Table 5. Comparison of the NCM, RMSE, and Time (s) of Six SAR Image Pairs with Different Methods.

Method	Pair 1			Pair 2			Pair 3			Pair 4			Pair 5			Pair 6		
	NCM	RMSE	Time	NCM	RMSE	Time	NCM	RMSE	Time	NCM	RMSE	Time	NCM	RMSE	Time	NCM	RMSE	Time
SAR-SIFT	33	0.77	718.51	70	0.68	2496.3	46	0.77	2807.37	23	0.89	975.43	40	0.76	1625.38	14	0.87	369.9
SOSNet	94	0.77	28.78	60	0.72	40.42	61	0.73	39.89	84	0.79	18.06	118	0.84	19.19	17	0.82	15.39
Sim-CSPNet	24	0.61	3.15	121	0.63	5.63	105	0.65	7.01	48	0.76	5.08	43	0.74	5.68	51	0.71	3.60
SD-CapsNet	**156**	**0.59**	**2.05**	**391**	**0.45**	**5.43**	**369**	**0.46**	**6.75**	**310**	**0.74**	**3.60**	**325**	**0.73**	**5.39**	**93**	**0.66**	**3.39**

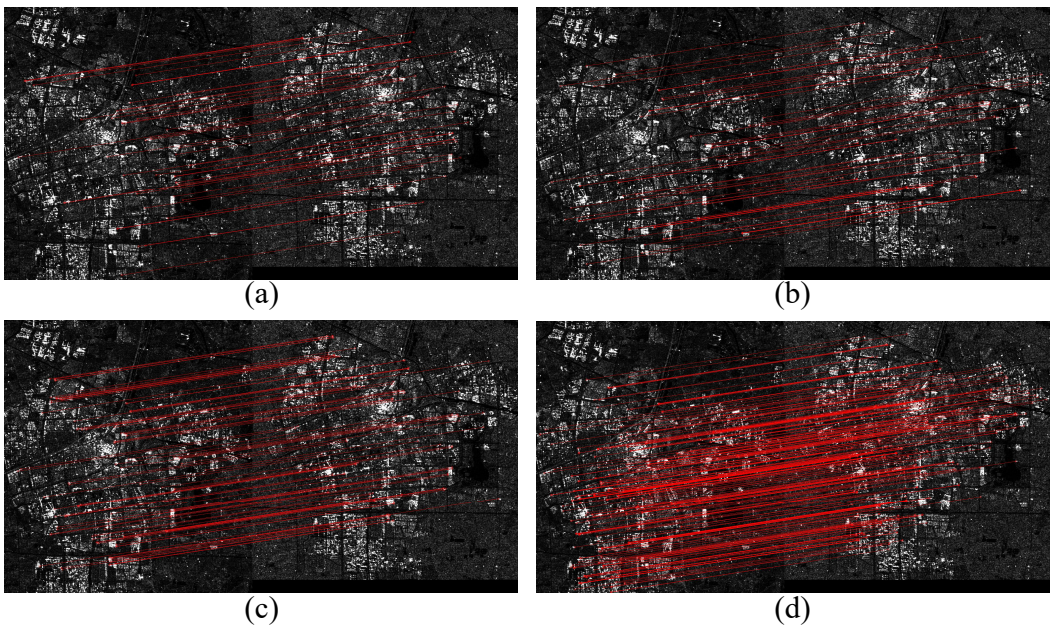


Figure 14. The feature descriptor matching results with different methods for pair 3. (**a**) SAR-SIFT. (**b**) SOSNet. (**c**) Sim-CSPNet. (**d**) Our proposed SD-CapsNet.

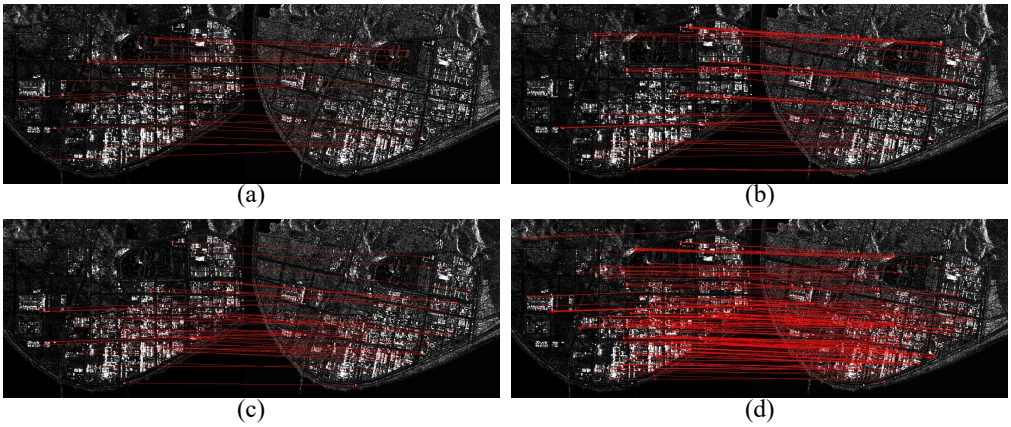


Figure 15. The feature descriptor matching results with different methods for pair 4. (**a**) SAR-SIFT. (**b**) SOSNet. (**c**) Sim-CSPNet. (**d**) Our proposed SD-CapsNet.

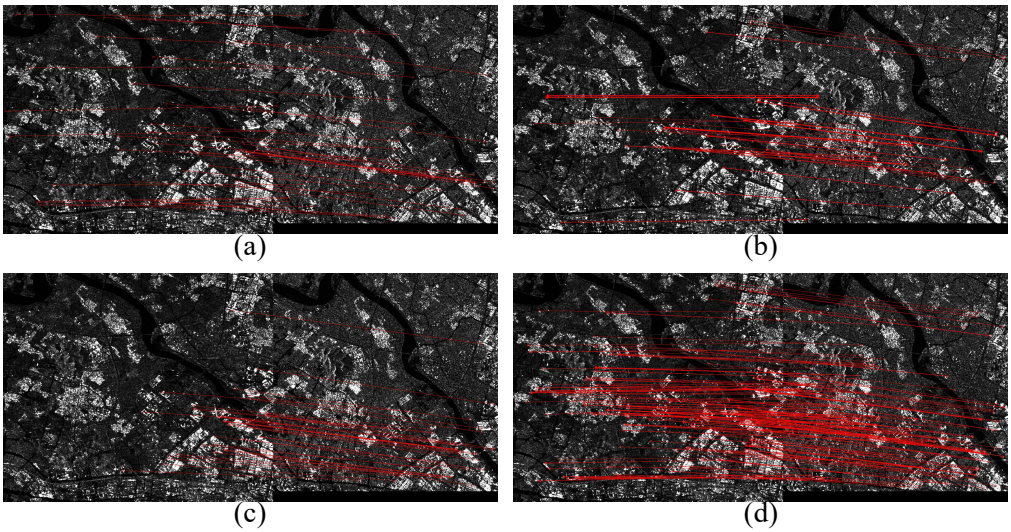


Figure 16. The feature descriptor matching results with different methods for pair 5. (**a**) SAR-SIFT. (**b**) SOSNet. (**c**) Sim-CSPNet. (**d**) Our proposed SD-CapsNet.

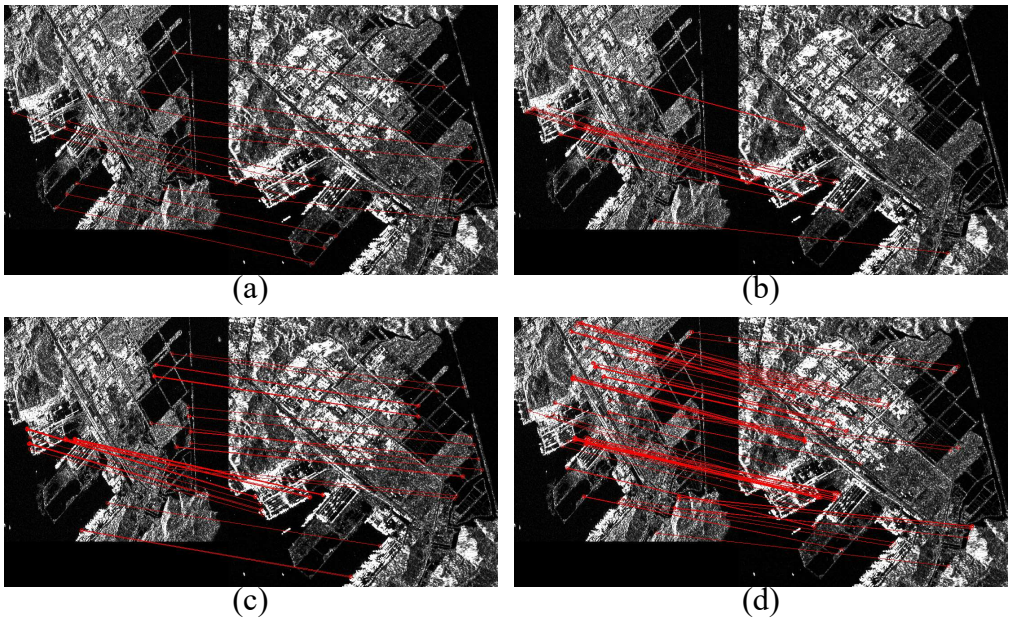


Figure 17. The feature descriptor matching results with different methods for pair 6. (**a**) SAR-SIFT. (**b**) SOSNet. (**c**) Sim-CSPNet. (**d**) Our proposed SD-CapsNet.

The quantitative comparison results of different methods are listed in Table 5. The NCM of SAR-SIFT is lower than that of the other comparison methods, whereas the running time is much longer than that of the other comparison methods. Although SOSNet is able to achieve larger NCMs, its RMSE is higher than that of the other comparison methods, and its running time is much longer than that of Sim-CSPNet and SD-CapsNet. Since Sim-CapsNet is designed with a keypoint detector based on feature intersection, it is able to detect keypoints with high repeatability and therefore has a lower RMSE than SAR-SIFT and SOSNet. However, Sim-CSPNet is unable to detect uniformly distributed matched keypoint pairs in SAR images with complex scenes. Our proposed SD-CapsNet achieves the highest NCM and lowest RMSE and uses the least amount of runtime compared to other comparison methods. Experimental results show that our proposed method is able to detect uniformly distributed keypoints in complex scenes and to extract accurate feature descriptors for keypoints.

To show the registration results more intuitively, we present final registration results in terms of checkboard overlay, as shown in Figure 18. In the registration results of SAR image pairs 1–3, SAR-SIFT and SOSNet produce incoherent regions, as shown in the red box regions in Figure 18. In the registration results for SAR image pairs 4 and 5, SAR-SIFT and Sim-CapsNet are unable to estimate accurate transformation models because they have fewer NCMs, resulting in incoherent regions, as shown in Figure 18(a4,c4,a5,c5). As shown in Figure 18(b5), although SOSNet obtains more NCMs, its RMSE is higher than that of other comparison methods; therefore, its results show incoherent regions. In the registration results of SAR image pair 6, the NCMs of SAR-SIFT and SOSNet are too small to accurately estimate the transformation model, resulting in significant incoherent regions, as shown in Figure 18(a6,b6). Our method outperforms other comparison methods in terms of NCM and RSME, so the checkerboard results of the proposed method have no obvious errors, as shown in Figure 18(d1–d6).

Finally, we show a enlarged map of a selected region of the registration results for pair 5 to explain the link between the keypoint matching result and the registration result, as shown in Figure 19. In the keypoint matching results of SOSnet and Sim-CSPNet, no correctly matched keypoint pairs appear in the selected region; therefore, incoherent roads are produced in the region of the registration results. In the keypoint matching results of SAR-SIFT, only a few correctly matched keypoint pairs appear in the selected region, which still produce incoherent roads, despite improved registration results compared to SOSnet and Sim-CSPNet. However, our proposed method achieves more correctly matched keypoint pairs in selected regions, so the transformation model is calculated more accurately, and no incoherent roads emerge.

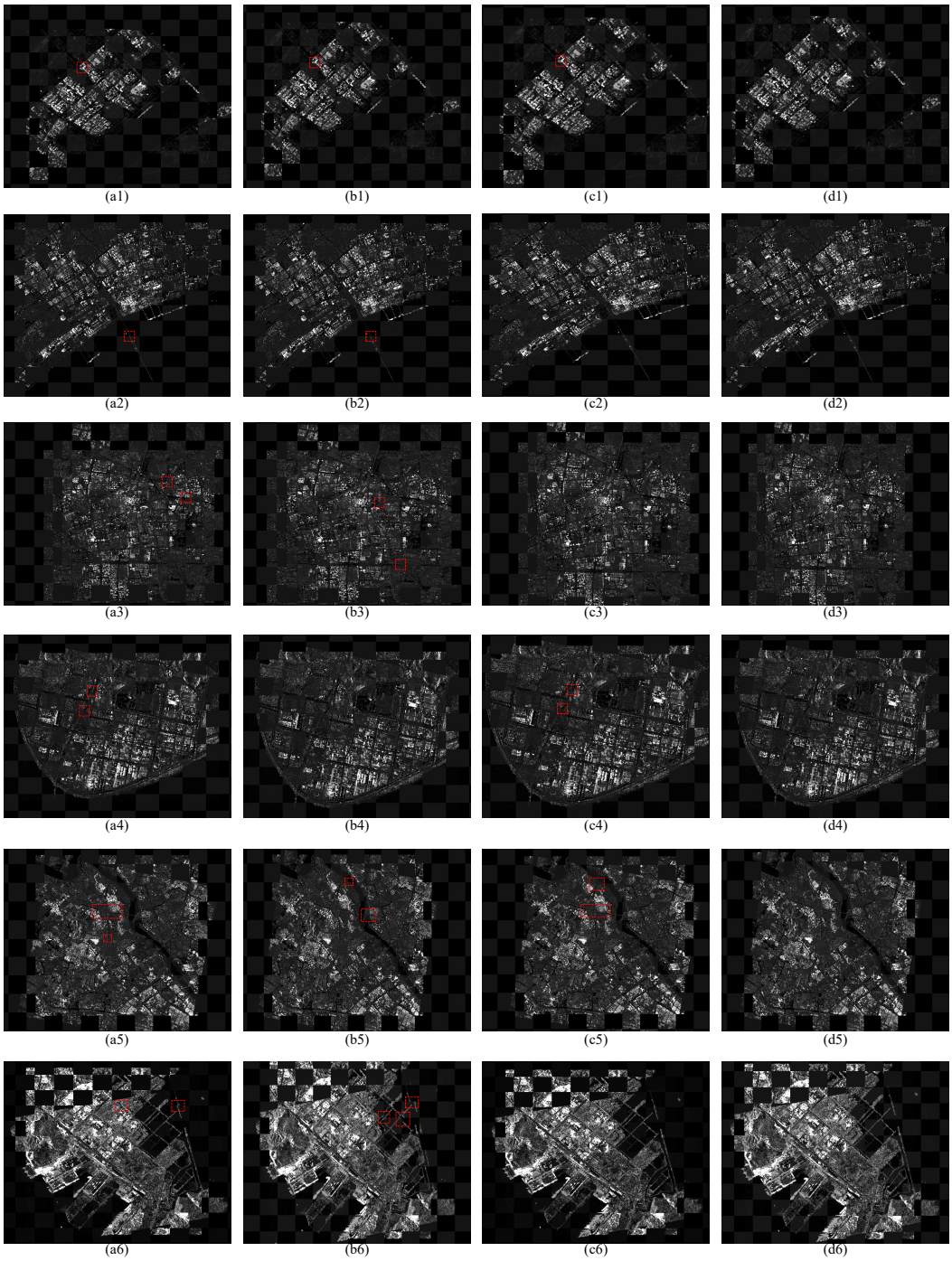


Figure 18. Checkboard overlay maps of the different methods. (**a1**–**a6**) Registration results of SAR-SIFT. (**b1**–**b6**) Registration results of SOSNet. (**c1**–**c6**) Registration results of Sim-CSPNet. (**d1**–**d6**) Registration results of SD-CapsNet.

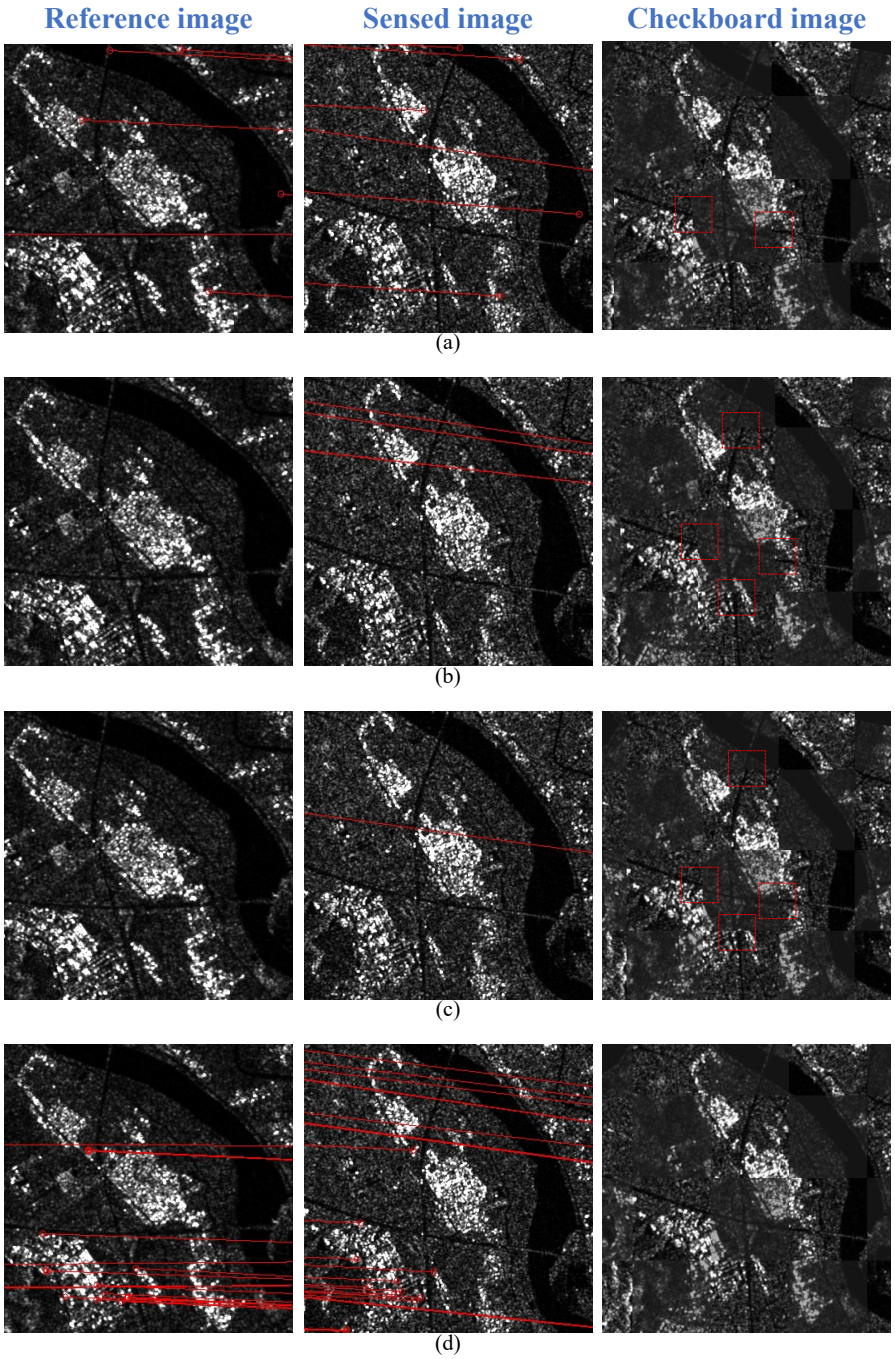


Figure 19. Enlarged maps of the registration results for pair 5 with different methods. (**a**) SAR-SIFT. (**b**) SOSNet. (**c**) Sim-CSPNet. (**d**) SD-CapsNet.

4. Discussion

4.1. Effectiveness of the TCPC Keypoint Detector

To verify the effectiveness of the proposed TCPC keypoint detector, ablation experiments are conducted on SAR image pair 6. Specifically, five different keypoint detectors are selected, and the same SD-CapsNet is used to implement feature descriptor extraction and matching. The feature descriptor matching results are shown in Figure 20. We can observe that the matched keypoint pairs obtained by the comparison method are not uniformly distributed in the SAR image. The NCMs of the comparison method are also much smaller than those of the proposed method according to visual inspection. Our proposed TCPC keypoint detector obtains more uniform and a larger number of correctly matched keypoint pairs.

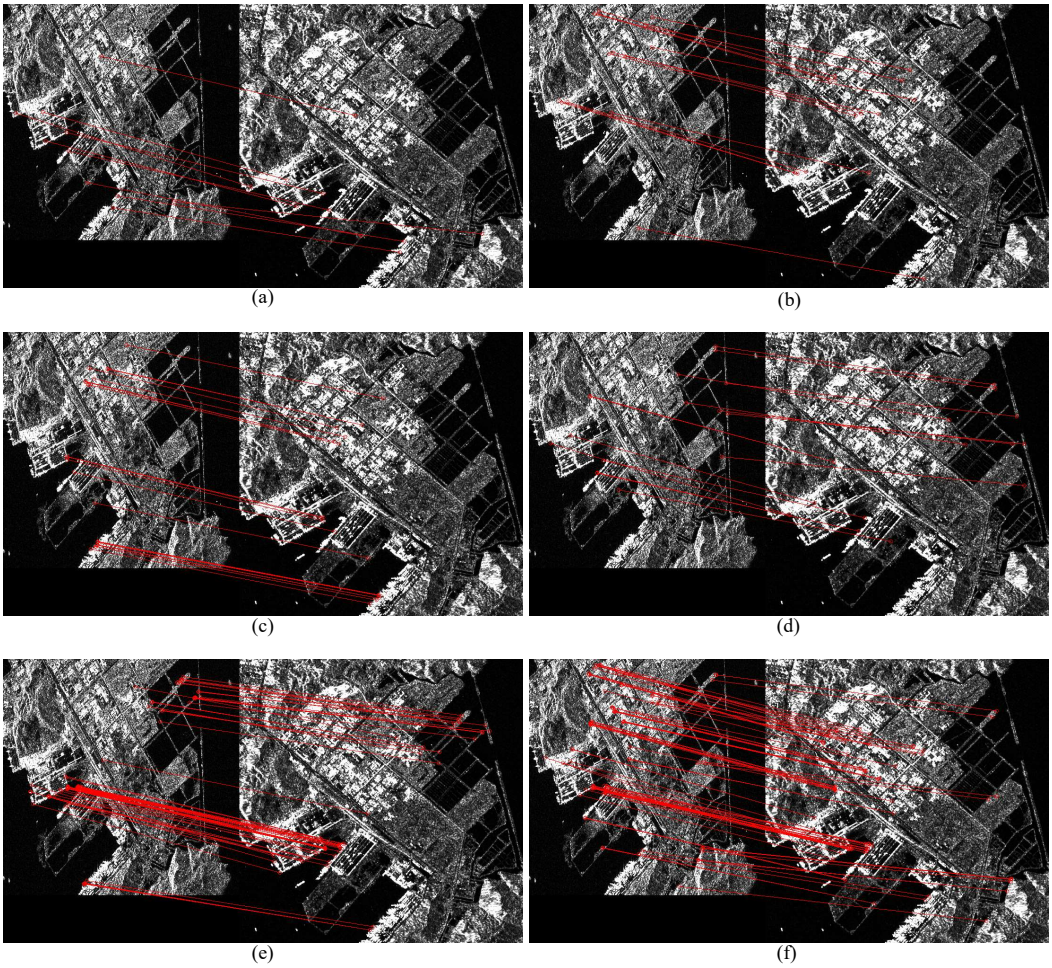


Figure 20. The feature descriptor matching results with different keypoint detectors. (**a**) DoG. (**b**) FAST. (**c**) Harris. (**d**) SAR-Harris. (**e**) Feature interaction. (**f**) Our proposed TCPC.

We present quantitative comparison results of different keypoint detectors in Table 6. The network takes more registration time when using the DoG, FAST, Harris, and SAR-Harris detectors, indicating that they can detect more keypoints than feature interaction (FI) and TCPC keypoint detectors. However their NCM is much smaller than that of FI and TCPC keypoint detectors, indicating that the repeatability of keypoints is not high. The FI detector detects a smaller number of keypoints and therefore yields a smaller NCM than that of TCPC. Our proposed TCPC detector detects keypoints with the highest repeatability in an acceptable detection time, thereby obtaining the largest NCM.

Table 6. Quantitative comparison of feature descriptor matching with different keypoint detectors.

Method	DoG [33]	FAST [43]	Harris [35]	SAR-Harris [13]	FI [20]	Proposed TCPC
Detection time	0.10	0.01	0.57	44.27	2.40	1.17
Matching time	2.65	3.28	2.44	2.81	1.14	2.22
Total time	2.75	3.29	3.01	47.08	3.54	3.39
NCM	14	14	22	33	74	93
RMSE	0.78	0.74	0.75	0.69	0.66	0.66

4.2. Effectiveness of the SD-CapsNet Feature Descriptor Extractor

To verify the effectiveness of the proposed SD-CapsNet, ablation experiments are conducted on SAR image pair 6. Specifically, three different registration networks are selected, and the same TCPC is used to implement keypoint detection. In particular, CapsNet represents a capsule network without dense connections, with the same number

of network layers and parameter settings as SD-CapsNet. The feature descriptor matching results with different feature descriptor extractors are shown in Figure 21. CapsNet and SOSNet are unable to obtain a uniformly distributed and sufficient number of correctly matched keypoint pairs.

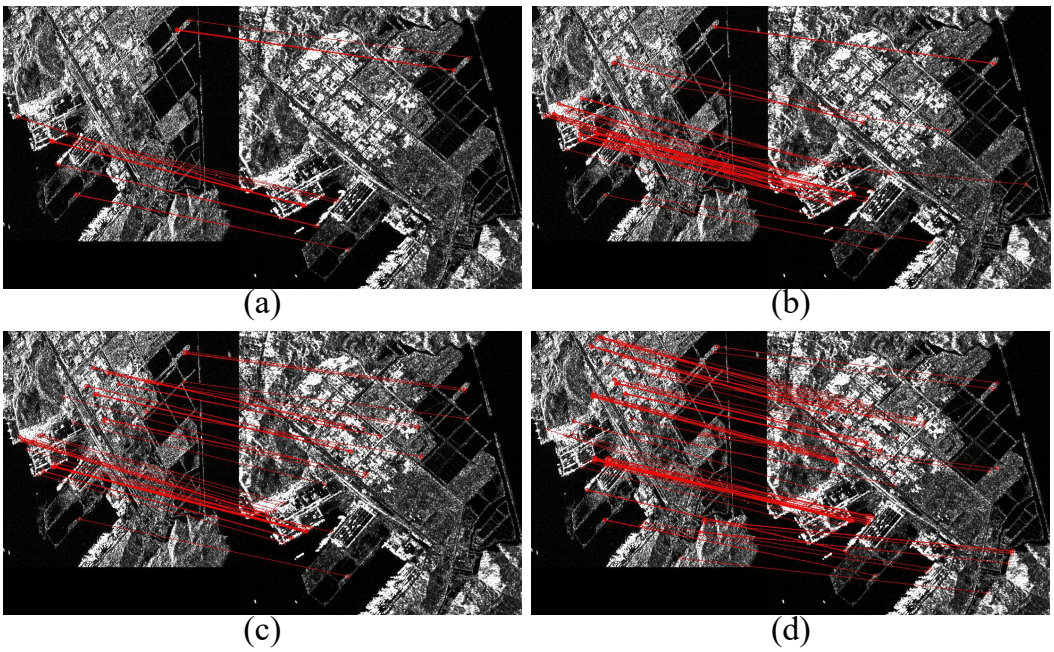


Figure 21. The feature descriptor matching results with different feature descriptor extractors. (**a**) TCPC+SOSNet. (**b**) TCPC+Sim-CSPNet. (**c**) TCPC+CapsNet. (**d**) TCPC+SD-CapsNet.

We present quantitative comparison results of different feature descriptor extractors in Table 7. The traditional capsule network has a much smaller NCM than that of SD-CapsNet due to the outdated connection method. Experimental results show that the proposed method achieves the largest NCM and the lowest RMSE with the shortest running time compared to the other comparison methods.

Table 7. Quantitative comparison of feature descriptor matching with different feature descriptor extractors.

Method	TCPC+SOSNet	TCPC+Sim-CSPNet	TCPC+CapsNet	SD-CapsNet
RMSE	0.74	0.71	0.71	0.66
NCM	22	55	57	96
Time(s)	8.85	3.47	3.23	3.39

5. Conclusions

In this paper, we propose a TCPC keypoint detector and a SD-CapsNet feature descriptor extractor to address the problems of inaccurate keypoint detection and inadequate feature descriptor extraction in SAR image registration of complex scenes. First, we extract the SAR image texture features using RI-LBP to remove the keypoints detected by the PC detector in the strong texture regions. Secondly, we propose SD-CapsNet to extract feature descriptors of keypoints, which can shorten the backpropagation distance and make the primary capsules contain both deep semantic information and shallow detail information. In addition, we construct feature descriptors in the capsule form and verify that each dimension of the capsule corresponds to intensity, texture, orientation, and structure information. Finally, we define the L2-distance between capsules for the feature descriptors and adopt a hard L2 loss function to train the model. The extensive experimental results demonstrate that the proposed method can maintain stable registration performance on

SAR images with complex scenes, and its NCM, RMSE, and running time are better than those of other state-of-the-art methods. However, because the proposed method masks keypoints in strongly textured regions, the distribution of keypoints may not be uniform in SAR images with large mountainous regions. In future work, we hope to use digital elevation models to improve registration accuracy in areas with large geometric distortions.

Author Contributions: Conceptualization, B.L. and D.G.; Methodology, B.L. and D.G.; Software, D.G., X.Z. and Z.C.; Resources, X.Z. and L.P.; Writing, B.L. and D.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Natural Science Basic Research Plan of Shaanxi Province, China, under grant 2022JQ-694.

Acknowledgments: The authors would like to thank the Deliang Xiang for providing the SAR testing images and sharing the codes for comparison.

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

References

1. Rambour, C.; Audebert, N.; Koeniguer, E.; Le Saux, B.; Crucianu, M.; Datcu, M. Flood detection in time series of optical and sar images. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *43*, 1343–1346. [CrossRef]
2. Cheng, J.; Zhang, F.; Xiang, D.; Yin, Q.; Zhou, Y. PolSAR image classification with multiscale superpixel-based graph convolutional network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5209314. [CrossRef]
3. Zhu, Y.; Yao, X.; Yao, L.; Yao, C. Detection and characterization of active landslides with multisource SAR data and remote sensing in western Guizhou, China. *Nat. Hazards* **2022**, *111*, 973–994. [CrossRef]
4. Sun, Y.; Lei, L.; Li, X.; Tan, X.; Kuang, G. Structure consistency-based graph for unsupervised change detection with homogeneous and heterogeneous remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4700221. [CrossRef]
5. Gong, M.; Zhou, Z.; Ma, J. Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering. *IEEE Trans. Image Process.* **2011**, *21*, 2141–2151. [CrossRef]
6. Fornaro, G.; Pauciullo, A.; Reale, D.; Verde, S. Multilook SAR tomography for 3-D reconstruction and monitoring of single structures applied to COSMO-SKYMED data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2776–2785. [CrossRef]
7. Xie, H.; Pierce, L.E.; Ulaby, F.T. Mutual information based registration of SAR images. In Proceedings of the IGARSS 2003—2003 IEEE International Geoscience and Remote Sensing Symposium, Proceedings (IEEE Cat. No. 03CH37477), Toulouse, France, 21–25 July 2003; Volume 6, pp. 4028–4031.
8. Wang, Y.; Yu, Q.; Yu, W. An improved Normalized Cross Correlation algorithm for SAR image registration. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, 22–27 July 2012; pp. 2086–2089.
9. Ye, Y.; Shen, L. Hopc: A novel similarity metric based on geometric structural properties for multi-modal remote sensing image matching. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *3*, 9. [CrossRef]
10. Xiang, Y.; Wang, F.; You, H. An automatic and novel SAR image registration algorithm: A case study of the Chinese GF-3 satellite. *Sensors* **2018**, *18*, 672. [CrossRef]
11. Cordón, O.; Damas, S. Image registration with iterated local search. *J. Heuristics* **2006**, *12*, 73–94. [CrossRef]
12. Wu, Y.; Ma, W.; Miao, Q.; Wang, S. Multimodal continuous ant colony optimization for multisensor remote sensing image registration with local search. *Swarm Evol. Comput.* **2019**, *47*, 89–95. [CrossRef]
13. Dellinger, F.; Delon, J.; Gousseau, Y.; Michel, J.; Tupin, F. SAR-SIFT: A SIFT-like algorithm for SAR images. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 453–466. [CrossRef]
14. Paul, S.; Pati, U.C. SAR image registration using an improved SAR-SIFT algorithm and Delaunay-triangulation-based local matching. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 2958–2966. [CrossRef]
15. Wang, M.; Zhang, J.; Deng, K.; Hua, F. Combining optimized SAR-SIFT features and RD model for multisource SAR image registration. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5206916. [CrossRef]
16. Durgam, U.K.; Paul, S.; Pati, U.C. SURF based matching for SAR image registration. In Proceedings of the 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 5–6 March 2016; pp. 1–5.
17. Liu, R.; Wang, Y. SAR image matching based on speeded up robust feature. In Proceedings of the 2009 WRI Global Congress on Intelligent Systems, Xiamen, China, 19–21 May 2009; Volume 4, pp. 518–522.
18. Pourfard, M.; Hosseinian, T.; Saeidi, R.; Motamedi, S.A.; Abdollahifard, M.J.; Mansoori, R.; Safabakhsh, R. KAZE-SAR: SAR image registration using KAZE detector and modified SURF descriptor for tackling speckle noise. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5207612. [CrossRef]
19. Eltanany, A.S.; Amein, A.S.; Elwan, M.S. A modified corner detector for SAR images registration. *Int. J. Eng. Res. Afr.* **2021**, *53*, 123–156. [CrossRef]
20. Xiang, D.; Xu, Y.; Cheng, J.; Hu, C.; Sun, X. An Algorithm Based on a Feature Interaction-based Keypoint Detector and Sim-CSPNet for SAR Image Registration. *J. Radars* **2022**, *11*, 1081–1097.

21. Quan, D.; Wang, S.; Ning, M.; Xiong, T.; Jiao, L. Using deep neural networks for synthetic aperture radar image registration. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 2799–2802.
22. Zhang, H.; Ni, W.; Yan, W.; Xiang, D.; Wu, J.; Yang, X.; Bian, H. Registration of multimodal remote sensing image based on deep fully convolutional neural network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3028–3042. [CrossRef]
23. Fan, Y.; Wang, F.; Wang, H. A transformer-based coarse-to-fine wide-swath SAR image registration method under weak texture conditions. *Remote Sens.* **2022**, *14*, 1175. [CrossRef]
24. Kayhan, O.S.; Gemert, J.C.v. On translation invariance in cnns: Convolutional layers can exploit absolute spatial location. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 14274–14285.
25. Cheng, J.; Zhang, F.; Xiang, D.; Yin, Q.; Zhou, Y.; Wang, W. PolSAR image land cover classification based on hierarchical capsule network. *Remote Sens.* **2021**, *13*, 3132. [CrossRef]
26. Sabour, S.; Frosst, N.; Hinton, G.E. Dynamic routing between capsules. In Proceedings of the NIPS 2017, Advances in Neural Information Processing Systems 30, Long Beach, CA, USA, 4–9 December 2017.
27. Deng, F.; Pu, S.; Chen, X.; Shi, Y.; Yuan, T.; Pu, S. Hyperspectral image classification with capsule network using limited training samples. *Sensors* **2018**, *18*, 3153. [CrossRef]
28. Xiang, C.; Zhang, L.; Tang, Y.; Zou, W.; Xu, C. MS-CapsNet: A novel multi-scale capsule network. *IEEE Signal Process. Lett.* **2018**, *25*, 1850–1854. [CrossRef]
29. Yang, S.; Lee, F.; Miao, R.; Cai, J.; Chen, L.; Yao, W.; Kotani, K.; Chen, Q. RS-CapsNet: An advanced capsule network. *IEEE Access* **2020**, *8*, 85007–85018. [CrossRef]
30. Phaye, S.S.R.; Sikka, A.; Dhall, A.; Bathula, D. Dense and diverse capsule networks: Making the capsules learn better. *arXiv* **2018**, arXiv:1805.04001 .
31. Fan, J.; Wu, Y.; Wang, F.; Zhang, Q.; Liao, G.; Li, M. SAR image registration using phase congruency and nonlinear diffusion-based SIFT. *IEEE Geosci. Remote Sens. Lett.* **2014**, *12*, 562–566.
32. Wang, L.; Sun, M.; Liu, J.; Cao, L.; Ma, G. A robust algorithm based on phase congruency for optical and SAR image registration in suburban areas. *Remote Sens.* **2020**, *12*, 3339. [CrossRef]
33. Goncalves, H.; Corte-Real, L.; Goncalves, J.A. Automatic image registration through image segmentation and SIFT. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 2589–2600. [CrossRef]
34. Li, Y.; Liu, L.; Wang, L.; Li, D.; Zhang, M. Fast SIFT algorithm based on Sobel edge detector. In Proceedings of the 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang, China, 21–23 April 2012; pp. 1820–1823.
35. Ye, Y.; Wang, M.; Hao, S.; Zhu, Q. A novel keypoint detector combining corners and blobs for remote sensing image registration. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 451–455. [CrossRef]
36. Kovesi, P. Phase congruency detects corners and edges. In Proceedings of the Australian Pattern Recognition Society Conference: DICTA 2003, Sydney, Australia, 10–12 December 2003.
37. Guo, Z.; Zhang, L.; Zhang, D. A completed modeling of local binary pattern operator for texture classification. *IEEE Trans. Image Process.* **2010**, *19*, 1657–1663.
38. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [CrossRef]
39. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
40. Wang, C.Y.; Liao, H.Y.M.; Wu, Y.H.; Chen, P.Y.; Hsieh, J.W.; Yeh, I.H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 390–391.
41. Wang, Z.; Liu, T. Two-stage method based on triplet margin loss for pig face recognition. *Comput. Electron. Agric.* **2022**, *194*, 106737. [CrossRef]
42. Tian, Y.; Yu, X.; Fan, B.; Wu, F.; Heijnen, H.; Balntas, V. Sosnet: Second order similarity regularization for local descriptor learning. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11016–11025.
43. Rosten, E.; Drummond, T. Machine learning for high-speed corner detection. In Proceedings of the Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; Part I 9, pp. 430–443.

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.