

# Cross-Rejective Open-Set SAR Image Registration

Shasha Mao<sup>1</sup>, Shiming Lu<sup>1</sup>, Zhaolong Du<sup>1</sup>, Licheng Jiao<sup>1\*</sup>, Shuiping Gou<sup>1</sup>, Luntian Mou<sup>2</sup>,  
Xuequan Lu<sup>3</sup>, Lin Xiong<sup>4</sup>, Yimeng Zhang<sup>1\*</sup>

<sup>1</sup>Xidian University, <sup>2</sup>Beijing University of Technology, <sup>3</sup>University of Western Australia, <sup>4</sup>ShensiliCon

## Abstract

Synthetic Aperture Radar (SAR) image registration is an essential upstream task in geoscience applications, in which pre-detected keypoints from two images are employed as observed objects to seek matched-point pairs. In general, the registration is regarded as a typical closed-set classification, which forces each keypoint to be classified into the given classes, but ignoring an essential issue that numerous redundant keypoints are beyond the given classes, which unavoidably results in capturing incorrect matched-point pairs. Based on this, we propose a Cross-Rejective Open-set SAR Image Registration (CroR-OSIR) method. In this work, these redundant keypoints are regarded as out-of-distribution (OOD) samples, and we formulate the registration as a special open-set task with two modules: supervised contrastive feature-tuning and cross-rejective open-set recognition (CroR-OSR). Unlike traditional open-set recognition, all samples, including OOD samples, are available in the CroR-OSR module. CroR-OSR conducts the closed-set classifications in individual open-set domains from two images, meanwhile employing the cross-domain rejection during training, to exclude these OOD samples based on confidence and consistency. Moreover, a new supervised contrastive tuning strategy is incorporated for feature-tuning. Especially, the cross-domain estimation labels obtained by CroR-OSR are fed back to the feature-tuning module for feature-tuning, to enhance feature discriminability. The experimental results illustrate that the proposed method achieves more precise registration than the state-of-the-art methods. The code is released at <https://github.com/XDYaoshi/CroR-OSIR-main>.

## 1. Introduction

Synthetic Aperture Radar (SAR) [20] has rich imaging characteristics with all-weather and high-resolution, and it has been widely applied in military and civilian fields. Nevertheless, SAR images are generally captured in different

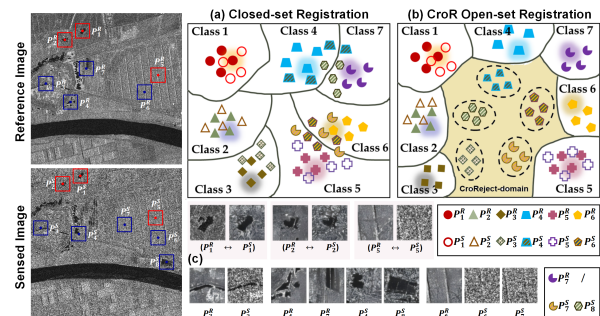


Figure 1. An illustration of Closed-Set Registration and CroR Open-Set Registration. Here, there are eight and seven pre-detected keypoints respectively marked on two SAR images, including three matched-point pairs labeled in red boxes and nine redundant keypoints labeled in blue boxes. It is seen that these redundant keypoints are forcefully classified to the seven classes given in closed-set registration, but they would be excluded into the reject-domain in CroR Open-Set Registration.

views or phases, resulting in inevitable diversities among them even in the same region, which increases the difficulty of SAR image processing [9]. Therefore, SAR image registration [31] has become an indispensable task in SAR image processing, especially when different images are analyzed simultaneously, and the registration precision of SAR images is essential for other downstream tasks, such as change detection [23], target tracking [8], image fusion [28], etc.

In SAR image registration, a common workflow involves first detecting keypoints on two SAR images and then employing them as observed objects to seek out the matched-point pairs between two images for registration. The matched-point pairs are more accurate, the registration is more precise. At present, the deep learning-based (DL-based) registration has become a research hotspot of SAR image registration. Earlier, in most DL-based registration methods [17, 26, 39, 43, 45], SAR image registration is treated as a classical binary classification problem that predicts whether a pair of keypoints are matched or not, and the predicted matched-point pairs are used to estimate the registration matrix. Recently, researchers have considered each keypoint as an independent category [6, 18] and then

\*Corresponding Authors.

E-mail: lchjiao@mail.xidian.edu.cn; zhangyimeng@xidian.edu.cn

directly constructed a multiple classification model to seek the matched-point pairs.

In most existing methods, the classification model constructed for image registration is generally specified in a classical closed-set space, here shortened as the closed-set registration. An illustration of the closed-set registration is given in Figure 1. Suppose  $m$  and  $n$  keypoints respectively pre-detected from two SAR images (the reference and sensed images), as shown in Figure 1. Based on these points, a closed-set classification model with  $m$  classes is constructed to classify  $n$  keypoints into the given classes and then obtain matched-point pairs, as shown in Figure 1(a). It is seen that three matched points can be classified into their matched classes, specifically  $p_1^S \rightarrow \text{Class-1 } (p_1^R)$ ,  $p_2^S \rightarrow \text{Class-2 } (p_2^R)$ , and  $p_5^S \rightarrow \text{Class-5 } (p_5^R)$ .

However, these redundant points are also classified into the given classes at the same time, such as  $p_3^S \rightarrow \text{Class-3 } (p_3^R)$ ,  $p_4^S$  and  $p_8^S \rightarrow \text{Class-4 } (p_4^R)$ ,  $p_6^S$  and  $p_7^S \rightarrow \text{Class-5 } (p_5^R)$ , etc, but these keypoints are essentially unmatched-point pairs. In practice, we normally obtain numerous redundant keypoints that do not have an inherent matching relationship to others by existing pre-detection algorithms [16], and their quantity may exceed 90% of all keypoints. Regrettably, *in closed-set registration, the model forces all keypoints to be classified into the given  $m$  categories, which causes these redundant points to be inevitably incorrectly matched.* Moreover, the regions around some unmatched points inherently share high similarity in SAR images, such as  $p_5^R, p_6^R, p_6^S$  and  $p_7^S$ , which demands higher discriminability of the features to distinguish them, whereas *binding all of them in a close-set space only with partial classes increases the challenge of extracting discriminative features.*

Noticeably, the essence of image registration is to seek out  $k$  matched-point pairs from  $m$  and  $n$  keypoints, meanwhile filtering out  $r$  redundant points,  $r = m + n - 2k$ . It means that there should be only  $c$  independent categories in all keypoints,  $c = k + r$  and  $c < m + n$ , rather than  $m$  or  $n$  categories. But, it is miserable that  $k$  and  $r$  keypoints are unknown, which also brings the inextricable problem for the close-set registration. Interestingly, if we refer to these redundant points ( $r$  keypoints) as Out-Of-Distribution (OOD) samples, the image registration can be regarded as an Open-Set Recognition (OSR) [33] problem. In the open-set scenario, the existing model normally identifies unseen OOD samples with unknown classes during the testing phase. Nevertheless, different from existing OSR, all samples (including OOD samples) are available in registration under the open-set scenario. And we need to catch these matched-point pairs and exclude these redundant points (OOD samples) during the training phase.

Based on this, we propose a Cross-Rejective Open-set SAR Image Registration method (*CroR-OSIR*) that aims at excluding OOD samples (unmatched points) meanwhile

seeking out accurately matched points during the training process. In the proposed method, a special supervised contrastive module *SupCon* is deployed to extract the features of keypoints. In the registration module, we further construct two independent classification domains  $\mathcal{R}$  and  $\mathcal{S}$  for two images, respectively corresponding to  $m$ -class and  $n$ -class classification problems. By uniting  $\mathcal{R}$  and  $\mathcal{S}$ , a Cross-Rejective OSR module (*CroR-OSR*) is constructed, which performs the rejection process within the classification domains  $\mathcal{R}$  and  $\mathcal{S}$ , meanwhile considering the processing of the combined domain  $\mathcal{D}$  with  $c$  classes. Specifically, the *CroR-OSR* module conducts the closed-set classification training within each domain and provides cross-domain estimated labels for each keypoint. Based on these cross-domain estimated labels, we further exclude unreliable matched-point pairs, ultimately obtaining  $k$  matched-point pairs with high confidence. The cross-domain rejection processes form the foundation of *CroR-OSR*. The cross-domain estimations are further fed back into *SupCon* for additional feature-tuning. As shown in Figure 1(b), the Cross-Rejective OSR algorithm figures out and rejects these redundant keypoints during training in the cross-rejection domain, with additional efforts in tuning the features.

Compared with most existing methods, the main contributions of the proposed method are shown as follows:

- We introduce a cross-rejective open-set recognition concept into SAR image registration for the first time, which provides a novel viewpoint on the registration problem from an open-set perspective, namely excluding the OOD samples from all known/unknown classes during training to promote seeking more accurate matched-points.
- This work proposes a cross-rejective OSR module with two cross-domain rejection strategies, which can filter out numerous redundant points during training and meanwhile obtain matched-point pairs with higher consistency, without the requirement of additional post-processing.
- This work presents a closed-set *SupCon* feature-tuning module with the guidance of the cross-domain estimation iteratively, to enhance the distinguishability of features among redundant points and matched points.

## 2. Related Works

### 2.1. SAR Image Registration

SAR image registration represents the foundation for information fusion of SAR images captured at different times, from different angles, or even in different modalities. The objective is to find the accurate pairs of matched points between two images (the reference image and the sensed image), and many methods have been proposed.

Initially, pixel intensity-based methods, such as NCC (Normalized Cross-Correlation), are implemented for template matching scenarios [32]. However, NCC-based meth-

ods are sensitive to rotation and affine transformations between two SAR images. The Sum of Squared Differences (SSD) [10] is another fast but unstable metric that directly compares image intensity values between images, but it is susceptible to intensity variations such as speckle noise. The Scale-Invariant Feature Transform (SIFT) [16] is one of the most influential methods in handcrafted feature-based registration. Consequently, numerous extensions of SIFT have been proposed, including Speeded-Up Robust Features (SURF) [3], PCA-SIFT (Principal Component Analysis SIFT) [12], Affine-SIFT (ASIFT) [21], Oriented FAST and Rotated BRIEF (ORB) [30], and so on.

At present, the DL-based registration has been a research hotspot. Zhang et al. [44] employed Convolutional Neural Networks (CNNs) to extract deep features from adaptively downsampled reference and sensed images. Huang et al. [11] proposed a two-stage registration method with two cascaded CNNs, which facilitated the coarse-to-fine registration of discontinuous frames in video SAR images. Given the indistinct textures and severe speckle noise in SAR images, some researchers have attempted to leverage attention mechanisms to extract richer features from SAR image patches. Deng et al. [6] proposed to treat each keypoint as an independent category and constructed a dual-branch classification network, where only the overlap of classes was considered as matched points. In practice, we observe that most of the pre-detected keypoints (potentially more than 90%) are unmatched points, due to the limitation of the pre-detection algorithm. These unmatched points heavily limit the existing classification deep models, with most of which take additional post-processing algorithms to filter the result to guarantee performance [14, 27, 43].

## 2.2. Open Set Recognition

Open Set Recognition (OSR) was introduced by Scheirer et al. [33, 34], aimed at minimizing open space risk while ensuring the classification accuracy of known classes, where unknown-class samples appeared only during the testing phase. In OSR, based on their probability distribution, samples are classified either as one of the known classes or as a rejection domain, becoming an unknown class.

Traditional OSR methods are typically implemented using some classic learning models, such as SVMs, nearest neighbours, sparse representation, etc. Popular classification methods include One-Class SVM [35] and Support Vector Data Description (SVDD) [36], where One-Class SVM separates the training samples from the origin feature space with the maximum margin and SVDD encloses the training data within a hypersphere of minimal volume. Scheirer et al. [34] incorporated a non-linear kernel into the solution, combined with statistical extreme value theory and proposed a Weibull-calibrated SVM model.

In recent years, there have been many advancements in

deep models for OSR in many fields [41][19]. Among the most recent research, OpenMax [4] is one of the pioneering works in considering the use of DNNs to estimate the probability that an input sample belongs to an unknown class. OpenMax seeks a meta-recognition model for each category. It uses the *libMR* library to perform a Weibull fit on the maximum distances between all correctly classified positive training instances and the corresponding centroid. ConOSR [42] proposes the Mixup strategy, which linearly combines two augmented samples of an image to construct negative pairs, and it utilizes contrastive learning to train a neural network for determining the rejection threshold. SLAN [40] is proposed based the formulation of multi-label open set recognition (MLOS) problem to recognize unknown labels by using sub-labeling information.

## 3. The Proposed Method

### 3.1. Overview

To eliminate the impact of numerous redundant points in the registration problem, we propose the Cross-Rejective Open-set SAR Image Registration method, shortened as *CroR-OSIR*, and the framework is shown in Figure 2, mainly including two modules: closed-set supervised contrastive feature-tuning and cross-rejective open-set recognition.

In *CroR-OSR*, we first detect  $m$  and  $n$  keypoints respectively from two images by using a traditional manner. Each keypoint is regarded as an independent category to construct the given dataset  $\{\mathcal{X}_R, \mathcal{X}_S\}$ . Then, a supervised contrastive module (*SupCon*) is utilized to tune the feature representation of all keypoints in the embedding space. In particular, the *SupCon* module is supervised by the dynamic cross-domain estimation inferred by the *CroR-OSR* module, which is beneficial to enhance the feature discriminability. In the *CroR-OSR* module, two distinct classification domains are deployed to learn the classification hyperplanes for all keypoints ( $m+n$ ). To reject out-of-distribution points in the whole space, the two domains are individually trained to classify and estimate all keypoints in their unified domain simultaneously. During the estimation process, the rejection rule is updated iteratively based on statistical information, to exclude more redundant keypoints. A mutual prediction process is constructed between two domains, to generate the mapping of corresponding categories ( $m-n$ ) and the cross-domain estimation for all keypoints. The cross-domain estimation is further filtered by a domain-interactive checking process and then fed back into *SupCon*.

In short, the *SupCon* and *CroR-OSR* modules are executed in a sequential and iterative manner, and the model eventually converges when the classification mapping relationship reaches a high level of confidence. The stabilized mapping relationship (or cross-domain estimation) presents the final registering result.

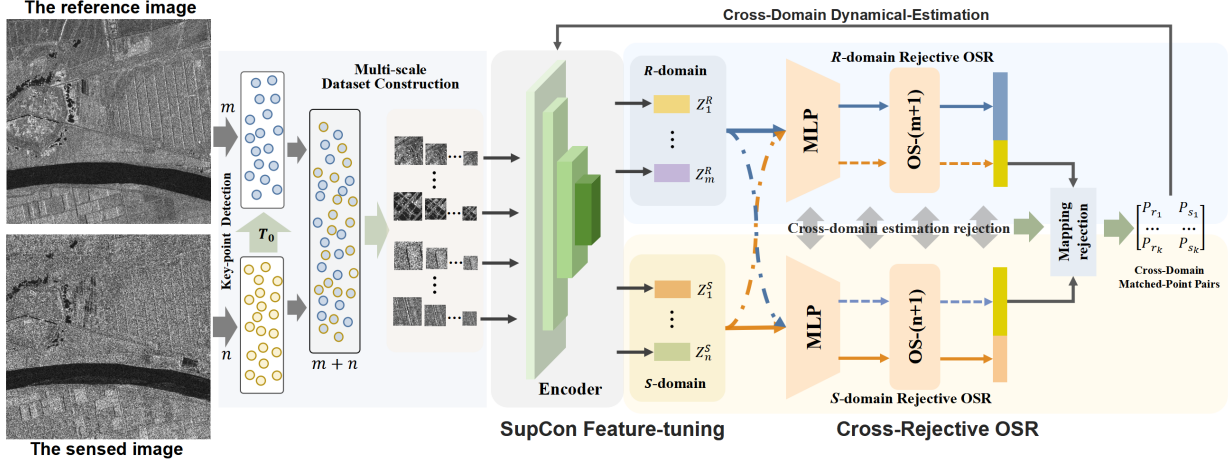


Figure 2. The framework of the proposed method (CroR-OSIR).

### 3.2. Formulation on Open-Set Image Registration

In the traditional OSR, given an entire data space denoted as  $\mathcal{D}$ , it is composed of a closed space ( $\mathcal{C}$ ) and an open space ( $\mathcal{O}$ ), where only the categories in  $\mathcal{C}$  are given, but  $\mathcal{O}$  is unknown and needs to be perceived in the learning process,  $\mathcal{D} = \mathcal{C} + \mathcal{O}$ . Suppose a closed space  $\mathcal{C}$  contain  $N$  categories, the sample set in  $\mathcal{C}$  is denoted as  $\mathcal{X}_{\mathcal{C}}$ . Meanwhile, suppose that the open space  $\mathcal{O}$  contains  $H$  categories, its sample set is denoted as  $\mathcal{X}_{\mathcal{O}}$ , and the entire space  $\mathcal{D}$  is of one sample set  $\mathcal{X}_{\mathcal{D}}$  containing  $N + H$  categories. Note that the number  $H$  of categories is unknown.

Different from the traditional closed-set task, the open-set task minimizes not only the open space risk but also the empirical classification risk, formulated as

$$\min_f \mathbb{E}_{x \in \mathcal{X}_{\mathcal{D}}} [\alpha L_{\mathcal{C}}(f(x), y) + \beta L_{\mathcal{O}}(f(x), y)], \quad (1)$$

where  $y$  expresses the label of the sample  $x$ ,  $f(x)$  represents the model prediction for  $x$ ,  $L_{\mathcal{C}}$  denotes the classification error risk,  $L_{\mathcal{O}}$  denotes the open set risk,  $\alpha$  and  $\beta$  are balancing weights between two risks, and  $\mathbb{E}_{x \in \mathcal{X}_{\mathcal{D}}} [\cdot]$  denotes the expectation over the sample space  $\mathcal{X}_{\mathcal{D}}$ .

**Formulation.** During the training phase, we treat the  $m$  keypoints from the reference image as an individual classification domain  $\mathcal{R}$  with  $m$  classes, and the  $n$  keypoints from the sensed image as another domain  $\mathcal{S}$  with  $n$  classes. Note that during testing, the sensed image (to be registered) will be fed only into the  $\mathcal{R}$  domain, where the classification results within the  $\mathcal{R}$  domain will be used to obtain the final matching relationships. Due to the sparse nature of the matching points, among the  $m + n$  points from all the images, typically we expect only  $k$  points to have a matching relationship, where  $0 \leq k \ll \min(m, n)$ . For the  $\mathcal{R}$  and  $\mathcal{S}$  domains, we denote the data as follows, respectively:

$$\mathcal{R} = \mathcal{C}_{\mathcal{R}} + \mathcal{O}_{\mathcal{R}}: \mathcal{C}_{\mathcal{R}} = \{\mathcal{X}_{\mathcal{C}}|m\} \text{ and } \mathcal{O}_{\mathcal{R}} = \{\mathcal{X}_{\mathcal{O}}|n-k\} \quad (2)$$

$$\mathcal{S} = \mathcal{C}_{\mathcal{S}} + \mathcal{O}_{\mathcal{S}}: \mathcal{C}_{\mathcal{S}} = \{\mathcal{X}_{\mathcal{C}}|n\} \text{ and } \mathcal{O}_{\mathcal{S}} = \{\mathcal{X}_{\mathcal{O}}|m-k\}. \quad (3)$$

When we merge the two independent domains to construct a closed-set case, we should have

$$\mathcal{D} = \{\mathcal{X}_{\mathcal{D}}|c\}, \text{ where } c = m + n - k, \quad (4)$$

but  $k$  is unknown and needs to be solved. As shown in Eqs.2 and 3, we treat the  $(m - k)$  keypoints and  $(n - k)$  keypoints as out-of-distribution (OOD) samples for corresponding individual domain, respectively. In the traditional OSR problem, samples with unknown categories are not available to the model during training, and the purpose is generally to recognize those unknown categories. Differently, the feature distribution of all  $m + n$  keypoints is provided in image registration. Interestingly, the challenge of image registration turns to be identifying the known categories (matched  $k$  keypoints) among all keypoints, rather than the unknown categories. Therefore, we transform the problem of searching out-of-distribution keypoints into finding overlapping ones in both domains, as shown in Eq.4, and we can further model it as follows

$$\min_f \mathbb{E}_{x \in \mathcal{X}_{\mathcal{D}}} (L_{\mathcal{R}} + L_{\mathcal{S}} + \gamma L_{crd}), \quad (5)$$

where  $L_{\mathcal{R}}$  and  $L_{\mathcal{S}}$  represent the traditional closed-set loss for two classification domains  $\mathcal{R}$  and  $\mathcal{S}$ , respectively. And the  $L_{crd}$  term represents the open space risk in the special combined domain,  $\mathcal{D}$ , and  $\gamma$  is the hyperparameter. Specifically,  $L_{crd}$  contains open-set rejection and cross-domain estimation process in our implementation, which constrains the keypoints from Image- $\mathcal{R}$  and Image- $\mathcal{S}$  to present a robust matching relationship. In experiments,  $\gamma$  is set as 1.

### 3.3. Supervised Contrastive Feature-tuning

First, we pre-detect keypoints respectively from the reference image (Image- $\mathcal{R}$ ) and the sensed image (Image- $\mathcal{S}$ ) us-



ing SIFT [16], and the sets of keypoints are denoted as

$$\rho_{\mathcal{R}} = \{p_1^r, p_2^r, \dots, p_m^r\}, \quad (6)$$

$$\rho_{\mathcal{S}} = \{p_1^s, p_2^s, \dots, p_n^s\}, \quad (7)$$

where  $m$  and  $n$  express the numbers of keypoints detected from the reference and sensed images, respectively. Moreover, we apply augments and the multi-scale cropping strategy to expand the samples, and more details can be found in the Supplementary Material.

Recently, supervised contrastive learning has been proven to be effective in providing discriminative features in classification tasks, including in the open-set recognition [13] [38, 42]. Thus, we adopt a contrastive module to tune the features of keypoints here as follows:

**Procedure 1.** Here we first input all the data from both classification domains  $\mathcal{R}$  and  $\mathcal{S}$  into a feature encoder, where Swin Transformer [15] is employed, which is effective at extracting global features [1]. Assuming a batch size of  $b$ , we denote a batch of data as  $\rho_B$ . First, we adopt a single fixed-scale cropping augmentation, where we resize image patches of different scales to a unified size of  $66 \times 66$ , and then randomly crop them to a size of  $64 \times 64$ . After the augmentation phase, each image patch corresponding to a keypoint in  $\rho_B$  generates two different augmented versions, denoted as  $\tilde{\rho}_B$ , which is now composed of  $2b$  image patches. With a projection head  $proj$ , we process  $\tilde{\rho}_B$  as follows:

$$\tilde{Z}_B = proj(f(\tilde{\rho}_B)) = \{z_1, z_2, \dots, z_{2b}\}. \quad (8)$$

Here the projection head transforms the dimension of encoded keypoints to 128. After projection, a traditional supervised contrastive objective in a batch is used:

$$L_{\text{sup}} = \frac{-1}{|I| \cdot |P(i)|} \sum_{i \in I} \sum_{p \in P(i)} \log \frac{\exp(z_i^T z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i^T z_a / \tau)}, \quad (9)$$

where  $I$  is the keypoints in the batch,  $|I|$  is the number of keypoints in the batch,  $|I| = 2b$ ,  $P(i)$  denotes the set of keypoints belonging to the same category as keypoint  $i$ , and  $A(i)$  represents the set of all keypoints different from point  $i$ .  $\tau$  is the temperature hyperparameter used to scale the similarity of the projected embedding vectors.

**Procedure 2.** After training the module as shown above for several epochs, we can obtain features with certain discriminative information. Differently from the traditional supervised contrastive manner, such features will be fed to the upcoming classification module in both domains to provide the category of each keypoint in the counterpart classification domain. For example, a keypoint  $i$  from the reference image is supposed to be classified in the domain  $\mathcal{R}$ , while we take the prediction result in classification domain  $\mathcal{S}$  here to generate a cross-domain estimation label for the point. After taking counterpart domain predictions for all samples

in the batch, we obtained the cross-domain labels for all samples and the category mapping relationship between the two classification domains,  $\mathcal{R}$  and  $\mathcal{S}$ . The total loss  $L_{\text{fea}}$  is given by

$$L_{\text{fea}} = L_{\text{sup}} + \mu L_{\text{crd}}^*, \quad (10)$$

$$L_{\text{crd}}^* = \frac{-1}{|I| \cdot |\hat{P}(i)|} \sum_{i \in I} \sum_{p \in \hat{P}(i)} \log \frac{\exp(z_i^T z_p / \tau)}{\sum_{a \in A(i)} \exp(z_i^T z_a / \tau)}, \quad (11)$$

where  $\hat{P}(i)$  denotes the set of positive keypoints belonging to the same cross-domain category as point  $i$ , and  $\mu$  is the hyperparameter. By employing additional cross-domain guided loss (with the cross-domain estimation setting to empty for the initial few batches in **Procedure 1** as warming up), we can increase the similarity between image patches that are similar but originate from different keypoints in the embedding space.

### 3.4. Cross-Rejective Open-Set Registration

After embedding all the keypoints, here we construct a Cross-Rejective OSR module with two classification heads  $h_{\mathcal{R}}(\cdot)$  and  $h_{\mathcal{S}}(\cdot)$ , corresponding to the two different classification domains  $\mathcal{R}$  and  $\mathcal{S}$ .

**Procedure 3.** After inputting the respective points to the corresponding classification domain, we have

$$p^{\mathcal{R}} = h_{\mathcal{R}}(f(\rho_{\mathcal{R}})) \text{ and } p^{\mathcal{S}} = h_{\mathcal{S}}(f(\rho_{\mathcal{S}})). \quad (12)$$

Here  $p^{\mathcal{R}}$  and  $p^{\mathcal{S}}$  represents the corresponding probability predictions in Image- $\mathcal{R}$  and Image- $\mathcal{S}$  respectively. The classification losses are computed as follows:

$$L_{\mathcal{R}} = CE(p^{\mathcal{R}}, y^{\mathcal{R}}) \text{ and } L_{\mathcal{S}} = CE(p^{\mathcal{S}}, y^{\mathcal{S}}), \quad (13)$$

where  $y^{\mathcal{R}}$  and  $y^{\mathcal{S}}$  represent the one-hot ground-truth labels in the two images respectively,  $y^{\mathcal{R}} \in \mathbb{R}^m$ ,  $y^{\mathcal{S}} \in \mathbb{R}^n$ .  $CE$  represents a conventional Cross-Entropy loss function.

**Procedure 4.** After we train  $h_{\mathcal{R}}(\cdot)$  and  $h_{\mathcal{S}}(\cdot)$  as shown above, a domain-interactive open-set rejection process is performed. First, we input all the points to their counterpart domain to get a cross-domain estimation probability result for each point. The estimated cross-domain result essentially contains information of the matching confidence. We apply a common-used statistical open-set rejection rule here [37][42] to filter the possible erroneous point pairs, considering the huge amount of out-of-distribution points in registration. Specifically, for category  $k$ ,  $\eta_k$  is the rejection threshold for open set recognition, which is determined by the  $\lambda$ -th percentile of correctly classified samples. We denote the cross-domain logits output for any keypoint as  $z$  here, and  $\hat{y}$  as a cross-domain ground-truth label:

$$\hat{y} = \begin{cases} \text{reject} & \text{if } \max(z) \leq \eta_k \\ k & \text{if } \max(z) > \eta_k \end{cases}, \quad (14)$$

where  $k = \text{argmax}(z)$ . By the cross-domain estimation rejection process shown in Eq.14, some keypoints with high matching confidence are left over. These matching pairs imply the cross-domain category mapping meanwhile, and we can further reject the unreliable mapping here. For example, a keypoint from Image- $\mathcal{R}$  with ground-truth label  $i$  is given a cross-domain ground-truth label  $j$  by Eq.14. When the matched keypoint from Image- $\mathcal{S}$  with ground-truth label  $j$  is given a cross-domain ground-truth label  $i$ , the  $i$ - $j$  mapping would be thought reliable, or it will be rejected.

Given the reliable mapping or cross-domain ground truth, we could finally present an overall loss with an additional cross-domain term  $L_{\text{crd}}$ . The total loss is given by:

$$L = L_{\mathcal{R}} + L_{\mathcal{S}} + \gamma L_{\text{crd}}, \quad (15)$$

$$L_{\text{crd}} = CE(\hat{y}_{\mathcal{R}}, h_{\mathcal{R}}(f(\rho_{\mathcal{S}}))) + CE(\hat{y}_{\mathcal{S}}, h_{\mathcal{S}}(f(\rho_{\mathcal{R}}))) \quad (16)$$

In Eq.16,  $\hat{y}_{\mathcal{R}}$  and  $\hat{y}_{\mathcal{S}}$  represent the cross-domain ground-truth label given in  $\mathcal{R}$  and  $\mathcal{S}$  respectively, and the keypoints mentioned in  $\rho_{\mathcal{R}}$  and  $\rho_{\mathcal{S}}$  only contain the ones remaining after open-set rejection and filter process here. Moreover, note that these points with cross-domain ground-truth labels will be fed back into **Procedure 2** to further tune the features. Generally, **Procedure 2** and **Procedure 4** form a sequential and iterative manner, and the whole algorithm will eventually converge until the mapping stabilizes. The stabilized matching relationship would be presented as our final registration result.

## 4. Experiments and Analyses

### 4.1. Datasets and Experiment Settings

In experiments, we use four datasets to validate the performance of the proposed method, including Wuhan, YellowR1/R2, and Yama datasets, where each dataset includes two SAR images. For multi-scale data construction, we use SIFT [16] to pre-detect keypoints on two images. Then, we employ five scales [17] to capture image patches of each keypoint, meanwhile, each point's eight nearest neighbors are also used to capture the image patches. Finally, 45 image patches are obtained for each keypoint. More details are given in our Supplementary Material.

During the network training stage, we initiate a warm-up training phase of 50~500 epochs for Encoder, and then additional 50 epochs of training are conducted with updates to the cross-domain estimation occurring every  $t$  epochs (such as 5, 10). In experiments, all input image patches are resized to  $64 \times 64$  and fed into the encoder. The batch size is set as 1024, the rejection threshold  $\lambda$  is set as 0.95~0.99, and the learning rate is 0.0001. Mixed-precision training is applied along with the Adam optimizer, and a cosine annealing schedule is used to adjust the learning rate. The coefficient  $\tau$  is set to 0.5. Furthermore, three crucial metrics

are applied to evaluate the registration performance, including  $N_{\text{red}}$ ,  $RMS_{\text{all}}$  and  $RMS_{\text{LOO}}$ , among these,  $RMS_{\text{all}}$  is the primary metric, representing the root mean square error (RMSE) of the registration, and a lower value indicates a more accurate registration.  $RMS_{\text{LOO}}$  is the  $RMS_{\text{all}}$  metric calculated using the leave-one-out method.  $N_{\text{red}}$  represents the number of matched-point pairs sought by one method.

### 4.2. Registration Performance of SAR Images

In this part, we compare the proposed method to 11 existing registration methods, including SIFT [16], SAR-SIFT [5], DNN [39], SuperPoint [7], Sparse-NCNet [29], MSDF-Net [17], STDT-Net [6], AdaSSIR [18], DALF [24], XFeat [25], and DBMDF [22], where the first two are handcrafted feature-based methods and the last nine are DL-based methods. Note that all compared methods use a post-processing strategy to obtain more accurate matched-point pairs. The registration results are shown in Table 1.

From Table 1, it is seen that the proposed method (*CroR-OSIR*) obtains higher registration performance than 11 compared methods on four datasets, meanwhile seeking out fewer matched-point pairs ( $N_{\text{red}}$ ). It indicates that the matched-point pairs obtained by *CroR-OSIR* are more accurate than others. Specifically, for YellowR1 datasets, it is observed that the number of matched-point pairs obtained by *CroR-OSIR* is comparable to SIFT and MSDF-Net, with 11 matched-point pairs, whereas the registration of our method is more accurate than them, with the improvements of 0.64 and 0.69 on  $RMS_{\text{all}}$  and  $RMS_{\text{LOO}}$ . It illustrates that the matching relationships established by our method have a higher confidence level. For Yama dataset, *CroR-OSIR* seeks out 25 matched-points pairs more than DNN (8), Sparse-NCNet (17), SuperPoint(19), XFeat(19), DALF(22), and MSDF-Net (12), but our result is better than theirs, which also illustrates that our sought matched-points are more accurate. Furthermore, it is noticed that these matched point pairs are reprocessed by the post-processing strategy in the compared methods, but our results are directly obtained without any additional strategies.

In addition, we also present the changes of the obtained matched-point pairs during the iterative process. Figure 3 shows the obtained matched-point pairs in the 5th, 10th, 15th, 25th, 30th and 50th epochs and their numbers ( $N_{\text{red}}$ ) for YellowR1 dataset, where all obtained matched points are labelled in images and alphabetized, and ( $\alpha$ ,  $\beta$ ) expresses a pair of two matched points. From Figure 3, it is first seen that the number of the obtained matched-point pairs exhibits a downward trend from the beginning to convergence. According to the coordinates of matched points, it is observed that some imprecise matched-point pairs (labelled by blue) are rejected in iterations, such as ( $B_R$ ,  $B_S$ ), ( $E_R$ ,  $E_S$ ), ( $M_R$ ,  $M_S$ ), ( $N_R$ ,  $N_S$ ), ( $G_R$ ,  $G_S$ ). Specifically for two difficult keypoints  $B_R$  and  $B_S$ , although the hor-

Table 1. Performance comparisons of SAR images registration on four datasets

Datasets	YellowR1			YellowR2			Yama			Wuhan		
Methods	$RMS_{all}$	$RMS_{LOO}$	$N_{red}$	$RMS_{all}$	$RMS_{LOO}$	$N_{red}$	$RMS_{all}$	$RMS_{LOO}$	$N_{red}$	$RMS_{all}$	$RMS_{LOO}$	$N_{red}$
SIFT [16]	0.9015	0.9436	11	1.1696	1.1711	88	1.1768	1.1806	69	1.2076	1.2139	17
SAR-SIFT [5]	1.0998	1.1424	31	1.1903	1.1973	301	1.2487	1.2948	151	1.2455	1.2491	66
DNN [39]	0.8024	0.8518	10	0.5784	0.5906	10	0.7293	0.7582	8	0.6471	0.6766	8
SuperPoint [7]	1.0801	1.1350	6	0.8950	0.8965	26	0.7947	0.8364	19	0.9199	0.9905	13
Sparse-NCNet [29]	0.6043	0.6126	17	0.6468	0.6595	67	0.6484	0.6591	17	0.6565	0.6777	44
MSDF-Net [17]	0.5923	0.6114	11	0.5051	0.5220	52	0.4645	0.4835	12	0.4345	0.4893	39
STDT-Net [6]	0.5487	0.5531	24	<u>0.4808</u>	<u>0.4954</u>	79	0.4604	0.4732	115	0.4490	0.4520	78
AdaSSIR [18]	0.5534	0.5720	20	0.5051	0.5220	52	0.4637	0.4707	71	<u>0.4217</u>	<u>0.4459</u>	47
DALF [24]	<u>0.3779</u>	<u>0.4026</u>	14	0.7385	0.8048	11	0.4264	0.4451	22	0.4588	0.4707	38
XFeat [25]	0.4516	0.4786	16	0.5819	0.6291	12	<u>0.3295</u>	<u>0.3453</u>	19	0.5039	0.5145	47
DBMDF [22]	0.8841	0.8966	17	0.9144	0.9261	32	0.4428	0.4634	59	0.8711	0.8745	71
CroR-OSIR	<b>0.2533</b>	<b>0.2565</b>	11	<b>0.0952</b>	<b>0.1386</b>	6	<b>0.3014</b>	<b>0.3022</b>	25	<b>0.1339</b>	<b>0.1402</b>	16

Table 2. Restroration results on optical images

Methods	DALF [24]	XFeat [25]	DBMDF [22]	CroR-OSIR
$RMS_{all}$	0.7043	0.7763	2.7918	<b>0.6600</b>
$RMS_{LOO}$	0.7095	0.7802	2.7946	<b>0.6605</b>
$N_{red}$	677	670	79	213

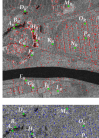



Reference Image	Epochs					
	5	10	15	25	30	50
	$N_{red} = 15$	$N_{red} = 12$	$N_{red} = 11$	$N_{red} = 10$	$N_{red} = 11$	$N_{red} = 11$
	(A <sub>R</sub> , A <sub>S</sub> ) (J <sub>R</sub> , J <sub>S</sub> )	(A <sub>R</sub> , A <sub>S</sub> ) (J <sub>R</sub> , J <sub>S</sub> )	(A <sub>R</sub> , A <sub>S</sub> ) (J <sub>R</sub> , J <sub>S</sub> )	(C <sub>R</sub> , C <sub>S</sub> ) (J <sub>R</sub> , J <sub>S</sub> )	(A <sub>R</sub> , A <sub>S</sub> ) (J <sub>R</sub> , J <sub>S</sub> )	(A <sub>R</sub> , A <sub>S</sub> ) (J <sub>R</sub> , J <sub>S</sub> )
	(B <sub>R</sub> , B <sub>S</sub> ) (K <sub>R</sub> , K <sub>S</sub> )	(C <sub>R</sub> , C <sub>S</sub> ) (K <sub>R</sub> , K <sub>S</sub> )	(C <sub>R</sub> , C <sub>S</sub> ) (K <sub>R</sub> , K <sub>S</sub> )	(D <sub>R</sub> , D <sub>S</sub> ) (K <sub>R</sub> , K <sub>S</sub> )	(C <sub>R</sub> , C <sub>S</sub> ) (K <sub>R</sub> , K <sub>S</sub> )	(C <sub>R</sub> , C <sub>S</sub> ) (K <sub>R</sub> , K <sub>S</sub> )
	(C <sub>R</sub> , C <sub>S</sub> ) (L <sub>R</sub> , L <sub>S</sub> )	(D <sub>R</sub> , D <sub>S</sub> ) (L <sub>R</sub> , L <sub>S</sub> )	(D <sub>R</sub> , D <sub>S</sub> ) (L <sub>R</sub> , L <sub>S</sub> )	(F <sub>R</sub> , F <sub>S</sub> ) (L <sub>R</sub> , L <sub>S</sub> )	(D <sub>R</sub> , D <sub>S</sub> ) (L <sub>R</sub> , L <sub>S</sub> )	(D <sub>R</sub> , D <sub>S</sub> ) (L <sub>R</sub> , L <sub>S</sub> )
	(D <sub>R</sub> , D <sub>S</sub> ) (M <sub>R</sub> , M <sub>S</sub> )	(F <sub>R</sub> , F <sub>S</sub> ) (O <sub>R</sub> , O <sub>S</sub> )	(F <sub>R</sub> , F <sub>S</sub> ) (O <sub>R</sub> , O <sub>S</sub> )	(H <sub>R</sub> , H <sub>S</sub> ) (O <sub>R</sub> , O <sub>S</sub> )	(F <sub>R</sub> , F <sub>S</sub> ) (O <sub>R</sub> , O <sub>S</sub> )	(F <sub>R</sub> , F <sub>S</sub> ) (O <sub>R</sub> , O <sub>S</sub> )
	(E <sub>R</sub> , E <sub>S</sub> ) (N <sub>R</sub> , N <sub>S</sub> )	(G <sub>R</sub> , G <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )	(H <sub>R</sub> , H <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )	(I <sub>R</sub> , I <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )	(H <sub>R</sub> , H <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )	(H <sub>R</sub> , H <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )
	(F <sub>R</sub> , F <sub>S</sub> ) (O <sub>R</sub> , O <sub>S</sub> )	(H <sub>R</sub> , H <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )	(I <sub>R</sub> , I <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )	(I <sub>R</sub> , I <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )	(I <sub>R</sub> , I <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )	(I <sub>R</sub> , I <sub>S</sub> ) (P <sub>R</sub> , P <sub>S</sub> )
	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )
	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )	(J <sub>R</sub> , J <sub>S</sub> )

Figure 3. The changes of numbers of matched-point pairs obtained by CroR-OSIR in iterations on YellowR1 dataset.

horizontal axis is consistent between them, there is some deviation (with 50 pixels) in the vertical axis, since there are matched deviations (8/9 pixels, 49/50 pixels) among final matched points. Moreover, although an imprecise matched-point pair ( $G_R$  and  $G_S$ ) is added in the 10th epoch, it is still rejected in subsequent iterations. Meanwhile, the pair ( $A_R$  and  $A_S$ ) is further sought out in later iterations.

Actually, although CroR-OSIR is proposed for SAR image registration to address the problem of redundant key-points, we find that it may be regarded as a new image registration framework based cross-rejective open-set recognition, which is adaptive for other data. Based on this, we make an experiment on the registration of optical images from HPatches-sequences-release [2]. In this experiment, we still used SIFT to detect keypoints from two optical images (obtaining 1451 and 1489 keypoints), the rejection threshold  $\lambda$  is set to 0.99, the learning rate is 0.00002, and the other parameters are the same as before. From the results shown in Table 2, it is seen that our CroR-OSIR ob-

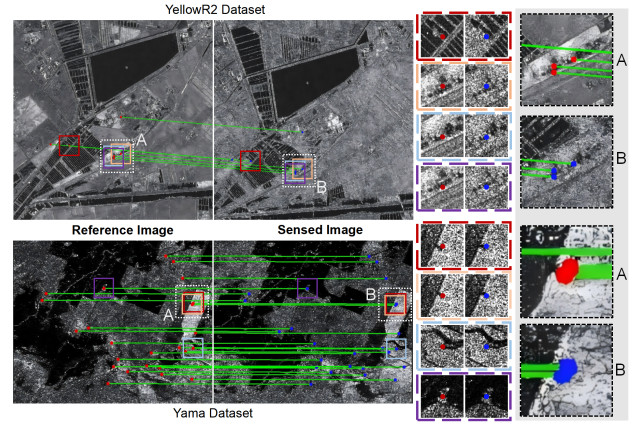


Figure 4. Visualization of the registration result given based on the matched point pairs for YellowR2 and Yama datasets.

tains better performance than the compared methods, with less matched-point pairs (213), which indicates our method is also effective for optical image registration.

### 4.3. Visualization of Registration Results

To better visually exhibit the registration results, we draw the connecting lines between the obtained matched-point pairs in this part. The visualization results on the YellowR2 and Yama datasets are illustrated in Figure 4, respectively, where the green line represents the connecting line between two matched points sought by the proposed method.

From Figure 4, it is obvious that the matching lines on both pairs of SAR images are parallel, demonstrating the accuracy of the proposed method. Meanwhile, it is seen that image patches of matched points are very similar, which illustrates the registration accuracy of our method. Moreover, we also find that there are high similarities among image patches of some adjacent points (densely distributed matching keypoints, as shown in A and B), such as those in the orange, blue and purple boxes from YellowR2, in the red and

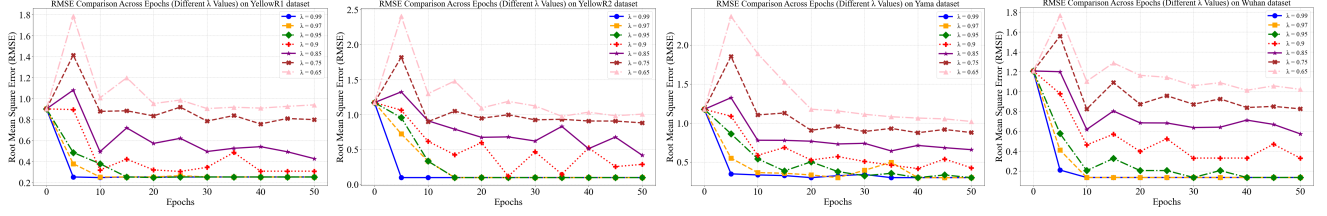


Figure 5. The registration performances ( $RM S_{all}$ ) under different  $\lambda$  (0.65, 0.75, 0.85, 0.90, 0.95, 0.97, 0.99) for four datasets.

orange boxes from Yama. High similarity increases the incorrectly matched risk, but our method can still achieve stable and precise matches in regions of high similarity, which validates the effectiveness of the cross-domain estimation labels in fine-tuning vectors within the feature space.

#### 4.4. Ablation Studies

In this part, we provide an analysis on the effect of different rejection thresholds ( $\lambda$ ) for registration. Here, we set the values of  $\lambda$  as 0.65, 0.75, 0.85, 0.90, 0.95, 0.97 and 0.99. Figure 5 shows the registration result ( $RM S_{all}$ ) obtained based on different values for four datasets in iterations.

From Figure 5, it is observed that the registration results become more accurate with the increase of the rejection threshold ( $\lambda$ ). For  $\lambda = 0.95, 0.97$  and  $0.99$ , their final registration performance is comparable, which indicates that a higher threshold is more effective for seeking out more accurate matched points. Moreover, it is seen that the model can converge faster when a higher threshold is set. For example, the model can converge fast when  $\lambda = 0.99$ , whereas the convergence may be reached at the 45th epoch for  $\lambda = 0.65$ . Empirically, a small  $\lambda$  leads to redundant matched point pairs (e.g., one-to-many or incorrect matches), resulting in performance degradation, while an excessively high  $\lambda$  prevents acquisition of matched-point pairs.  $\lambda$  is essentially related to matching confidence, and a higher value corresponds to a more correct matching when the encoder is better trained. Noticeably, although a higher threshold seems more effective, we found that it might bring higher risks in abortively seeking matched-point pairs, meanwhile a preminent encoder is also crucial.

Furthermore, to demonstrate the effectiveness of our proposed feature-tuning strategy for the *SupCon* module, we conduct an ablation experiment with a rejection threshold set to 0.95 and 0.90. In this experiment, we retain the original settings but remove the cross-domain estimation tuning term ( $L_{crd}^*$  in Section 3.3) for supervised contrastive feature-tuning, and  $\mu = 1$ . The experimental results are presented in Tabel.3. As shown in Table 3, it is observed that removal of the cross-domain estimation tuning term poses a significantly negative impact, especially when the rejection threshold  $\lambda = 0.90$ . The algorithm fails to exclude imprecise keypoint pairs. It implies that the feature

Table 3. Ablation studies on feature-tuning strategy in *SupCon*

Datasets		YellowR1			Yama		
$\lambda$	fine-tune	$RM S_{all}$	$RM S_{LOO}$	$N_{red}$	$RM S_{all}$	$RM S_{LOO}$	$N_{red}$
0.90	×	1.3697	1.3717	30	8.0312	8.0404	94
	✓	<b>0.3085</b>	<b>0.3120</b>	12	<b>0.4274</b>	<b>0.4442</b>	31
0.95	×	0.9104	0.9121	27	0.9462	0.9466	63
	✓	<b>0.2533</b>	<b>0.2565</b>	11	<b>0.3014</b>	<b>0.3022</b>	25

distribution of difficult keypoints in the combined classification domain remains mixed, and relying solely on training classifiers for the two individual domains is insufficient for effective feature-tuning.

## 5. Conclusion

To deal with the numerous redundant keypoints in registration, we propose to solve the SAR image registration problem from an open-set perspective. Different from traditional OSR, we propose a cross-rejective open-set registration method that identifies and excludes the redundant keypoints during the training process. Through cross-domain open-set rejection, the proposed method successfully filters out most redundant points during training. Additionally, the resulting cross-domain estimates are further applied to a supervised contrastive feature-tuning strategy, improving feature discriminability and thus ensuring the model performance on difficult keypoints with higher similarities. Experimental results on four datasets demonstrate that our method can seek out more accurate matched-point pairs and then obtain more precise registration than the state-of-the-art methods, without using any post-processing. The visualization results and ablation studies also validate that our method enhances the feature discriminability of difficult keypoints with high similarities by feature fine-tuning with the guidance of cross-domain estimation labels provided via the cross-rejection open set recognition module.

## 6. Acknowledgement

This work was supported in part by the State Key Program of National Natural Science of China under Grant 62234010.



## References

- [1] Abdulaziz Amer Aleissae, Amandeep Kumar, Rao Muhammad Anwer, Salman Khan, Hisham Cholakkal, Gui-Song Xia, and Fahad Shahbaz Khan. Transformers in remote sensing: A survey. *Remote Sensing*, 15(7):1860, 2023. 5
- [2] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5173–5182, 2017. 7
- [3] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008. 3
- [4] Abhijit Bendale and Terrance E Boulton. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1563–1572, 2016. 3
- [5] Flora Dellinger, Julie Delon, Yann Gousseau, Julien Michel, and Florence Tupin. Sar-sift: a sift-like algorithm for sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1):453–466, 2014. 6, 7
- [6] Xiaozheng Deng, Shasha Mao, Jinyuan Yang, Shiming Lu, Shuiping Gou, Youming Zhou, and Licheng Jiao. Multi-class double-transformation network for sar image registration. *Remote Sensing*, 15(11):2927, 2023. 1, 3, 6, 7
- [7] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018. 6, 7
- [8] Hui Fang, Guisheng Liao, Yongjun Liu, and Cao Zeng. Shadow-assisted moving target tracking based on multidiscriminant correlation filters network in video sar. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. 1
- [9] Giorgio Franceschetti and Riccardo Lanari. *Synthetic aperture radar processing*. CRC press, 2018. 1
- [10] MB Hisham, Shahruil Nizam Yaakob, RAA Raof, AB A Nazren, and NM Wafi. Template matching using sum of squared difference and normalized cross correlation. In *2015 IEEE student conference on research and development (SCoReD)*, pages 100–104. IEEE, 2015. 3
- [11] Xuejun Huang, Jinshan Ding, and Qinghua Guo. Unsupervised image registration for video sar. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:1075–1083, 2020. 3
- [12] Yan Ke and Rahul Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, pages II–II. IEEE, 2004. 3
- [13] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020. 5
- [14] Bangjie Li, Dongdong Guan, Xiaolong Zheng, Zhengsheng Chen, and Lefei Pan. Sd-capsnet: A siamese dense capsule network for sar image registration with complex scenes. *Remote Sensing*, 15(7):1871, 2023. 3
- [15] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 5
- [16] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004. 2, 3, 5, 6, 7
- [17] Shasha Mao, Jinyuan Yang, Shuiping Gou, Licheng Jiao, Tao Xiong, and Lin Xiong. Multi-scale fused sar image registration based on deep forest. *Remote Sensing*, 13(11):2227, 2021. 1, 6, 7
- [18] Shasha Mao, Jinyuan Yang, Shuiping Gou, Kai Lu, Licheng Jiao, Tao Xiong, and Lin Xiong. Adaptive self-supervised sar image registration with modifications of alignment transformation. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023. 1, 6, 7
- [19] Kirill Mazur, Edgar Sucar, and Andrew J Davison. Feature-realistic neural fusion for real-time, open set scene understanding. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8201–8207. IEEE, 2023. 3
- [20] Alberto Moreira, Pau Prats-Iraola, Marwan Younis, Gerhard Krieger, Irena Hajnsek, and Konstantinos P Papathanassiou. A tutorial on synthetic aperture radar. *IEEE Geoscience and remote sensing magazine*, 1(1):6–43, 2013. 1
- [21] Jean-Michel Morel and Guoshen Yu. Asift: A new framework for fully affine invariant image comparison. *SIAM journal on imaging sciences*, 2(2):438–469, 2009. 3
- [22] Javid Norouzi, Mohammad Sadegh Helfroush, Alireza Liaghat, and Habibollah Danyali. A deep-based approach for multi-descriptor feature extraction: Applications on sar image registration. *Expert Systems with Applications*, 254: 124291, 2024. 6, 7
- [23] Eleonora Jonasova Parelius. A review of deep-learning methods for change detection in multispectral remote sensing images. *Remote Sensing*, 15(8):2092, 2023. 1
- [24] Guilherme Potje, Felipe Cadar, André Araujo, Renato Martins, and Erickson R Nascimento. Enhancing deformable local features by jointly learning to detect and describe keypoints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1306–1315, 2023. 6, 7
- [25] Guilherme Potje, Felipe Cadar, André Araujo, Renato Martins, and Erickson R Nascimento. Xfeat: Accelerated features for lightweight image matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2682–2691, 2024. 6, 7
- [26] Dou Quan, Shuang Wang, Mengdan Ning, Tao Xiong, and Licheng Jiao. Using deep neural networks for synthetic aperture radar image registration. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2799–2802. IEEE, 2016. 1
- [27] Dou Quan, Huiyuan Wei, Shuang Wang, Ruiqi Lei, Baorui Duan, Yi Li, Biao Hou, and Licheng Jiao. Self-distillation

- feature learning network for optical and sar image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18, 2022. 3
- [28] Yinghui Quan, Yingping Tong, Wei Feng, Gabriel Dauphin, Wenjiang Huang, and Mengdao Xing. A novel image fusion method of multi-spectral and sar images for land cover classification. *Remote Sensing*, 12(22):3801, 2020. 1
- [29] Ignacio Rocco, Relja Arandjelović, and Josef Sivic. Efficient neighbourhood consensus networks via submanifold sparse convolutions. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 605–621. Springer, 2020. 6, 7
- [30] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. Ieee, 2011. 3
- [31] Eugenio Sansosti, Paolo Berardino, Michele Manunta, Francesco Serafino, and Gianfranco Fornaro. Geometrical sar image registration. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10):2861–2870, 2006. 1
- [32] Jignesh N Sarvaiya, Suprava Patnaik, and Salman Bombaywala. Image registration by template matching using normalized cross-correlation. In *2009 international conference on advances in computing, control, and telecommunication technologies*, pages 819–822. IEEE, 2009. 2
- [33] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boulton. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35(7):1757–1772, 2012. 2, 3
- [34] Walter J Scheirer, Lalit P Jain, and Terrance E Boulton. Probability models for open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 36(11):2317–2324, 2014. 3
- [35] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001. 3
- [36] David MJ Tax and Robert PW Duin. Support vector data description. *Machine learning*, 54:45–66, 2004. 3
- [37] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need? 2021. 5
- [38] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7492–7501, 2022. 5
- [39] Shuang Wang, Dou Quan, Xuefeng Liang, Mengdan Ning, Yanhe Guo, and Licheng Jiao. A deep learning framework for remote sensing image registration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:148–164, 2018. 1, 6, 7
- [40] Yibo Wang, Jun-Yi Hang, and Min-Ling Zhang. Multi-label open set recognition. *Advances in Neural Information Processing Systems*, 37:5739–5756, 2024. 3
- [41] Zhenyu Wang, Yali Li, Xi Chen, Ser-Nam Lim, Antonio Torralba, Hengshuang Zhao, and Shengjin Wang. Detecting everything in the open world: Towards universal object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11433–11443, 2023. 3
- [42] Baile Xu, Furao Shen, and Jian Zhao. Contrastive open set recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10546–10556, 2023. 3, 5
- [43] Famao Ye, Yanfei Su, Hui Xiao, Xuqing Zhao, and Weidong Min. Remote sensing image registration using convolutional neural network features. *IEEE Geoscience and Remote Sensing Letters*, 15(2):232–236, 2018. 1, 3
- [44] Han Zhang, Weiping Ni, Weidong Yan, Deliang Xiang, Junzheng Wu, Xiaoliang Yang, and Hui Bian. Registration of multimodal remote sensing image based on deep fully convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8):3028–3042, 2019. 3
- [45] Bin Zou, Haolin Li, and Lamei Zhang. Self-supervised sar image registration with sar-superpoint and transformation aggregation. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2022. 1