# Transfer Learning for Optical and SAR Data Correspondence Identification With Limited Training Labels

Mengmeng Zhang, *Member, IEEE*, Wei Li, *Senior Member, IEEE*, Ran Tao, *Senior Member, IEEE*, and Song Wang, *Senior Member, IEEE*

***Abstract*—Recent advancements in sensor technology have reflected promise in collaborative utilization; specifically, multisource remote sensing data correspondence identification attracts increasing attention. In this article, a domain-transfer learning based generative correspondence analysis (DT-GCA) scheme is proposed, which enables identifying corresponding data in optical and synthetic aperture radar (SAR) images with small-sized reference data. In the proposed architecture, an adversarial domain-translator is investigated as general-purpose domain transference solution to learn cross domain features. The optical-aided implicit representation, which is regarded as the clone of SAR, is adopted to estimate the correlation with SAR images. Particularly, the designed GCA integrates optical-generated features with SAR tightly instead of treating them separately and eliminates the discrepancy influence of different sensors. Experiments on cross-domain remote sensing data are validated, and extensive results demonstrate that the proposed DT-GCA yields substantial improvements over some state-of-the-art techniques when only limited training samples are available.**

***Index Terms*—Multisource correspondence identification, pattern recognition, remote sensing, transfer learning.**

## I. Introduction

With the rapid development of observation technologies, collaborative utilization of information from varied sensors has attracted considerable attention recently [1], [2]. Many tasks benefit from multisource information collaboration, where collaborative observation and monitoring applications with operational requirements have been extensively studied [3]–[5]. However, multisource images reflect great diversity in data volume, scale, and measurement system; thus, any subsequent fusion requests may fail without alignment and proper correlation of measurements for the same scene. Therefore, correspondence analysis serves a reasonably wide demand base in many areas.

In remote sensing area, the acquisition time for multisource data varies enormously according to corresponding collection purpose. Heterogeneous sensors represent potentially varied temporal, radiometric, geometric, and other properties. Therefore, heterogeneous image matching and correspondence identification are a common requirement for remote sensing applications, and the correspondence analysis has played significant role in specific applications, such as multisource coregistration, heterogeneous sensor data retrieval, and stereogrammetry [6]–[8]. There have already been some successful correspondence estimation investigations in computer vision setting. Metric learning, as a straightforward correlation analysis method, achieves improved results on simple similarity functions, e.g., Euclidean distance [9]. Nevertheless, with totally different imaging mechanism of remote sources and the scarcity of labeled patches, similarity function gets unwieldy to learn on multisource data [10]; the effectiveness of parameters optimization involved in similarity estimation would be uncontrollable, due to insufficient analysis of the data volume diversity and nonlinear radiometric differences under limited training data. For multisource remote sensing data correlation estimation, conventional hand-crafted approaches focus on matching human-engineering features. Gradient methods generally use both magnitudes and orientations for catching image properties. For example, researchers took advantage of spatial constrains to refine scale-invariant feature transform (SIFT) features [11], and the modified SIFT features were utilized for correlation analysis. However, gradient magnitudes may not be correlated due to different contrasts between images [12]. Ye *et al.* [13] further proposed a structural properties based feature descriptor for multimodal remote sensing image similarity metric. Aforementioned approaches involve massive hand-crafted features, which depend heavily on experts' experience and parameter setting. Nevertheless, it is difficult to find appropriate parameters to generate features for multisensor data similarity estimation.

Recently, deep-learning-based methods have broadly replaced hand-engineered approaches in image matching and correlation estimation [14], [15]. In [16], the in-depth analysis showed that the convolutional descriptors outperform SIFT in most cases except for the blur cases. Zbontar *et al.* [17] further

Manuscript received August 4, 2020; revised September 21, 2020 and November 15, 2020; accepted December 7, 2020. Date of publication December 14, 2020; date of current version January 8, 2021. This work was supported by the National Natural Science Foundation of China under Grant 62001023 and Grant U1833203 and in part by the China Postdoctoral Science Foundation under Grant 2020M670163 and Grant BX20200058. *(Corresponding author: Wei Li.)*

Mengmeng Zhang, Wei Li, and Ran Tao are with the College of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: 7520200002@bit.edu.cn; liwei089@ieee.org; rantao@bit.edu.cn).

Song Wang is with the Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208 USA (e-mail: songwang@cec.sc.edu).

Digital Object Identifier 10.1109/JSTARS.2020.3044643

demonstrated that convolutional neural network (CNN) was a good fit for computing the stereo matching cost. Inspired by the successful use of Siamese networks for image matching, Merkle *et al.* [18] adopted the Siamese network to learn the spatial shift between optical and synthetic aperture radar (SAR) image patches. Nevertheless, Zagoruyko *et al.* [19] pointed out that when compared to restricted Siamese networks, pseudo-Siamese networks are more flexible and adaptable. Thus in [20], a pseudo-Siamese architecture was deployed to learn the identification of corresponding patches in SAR and optical images. However, compared with Siamese CNN, the number of parameters involved in pseudo-Siamese network expands accordingly, and it is difficult to avoid the overfitting issue due to the low available quantity of annotated samples in a deep CNN. Moreover, in [21], generative adversarial networks (GANs) were utilized for controlling radiometry difference between optical and SAR images. Multisource analysis and correspondence estimation can be achieved using different supervisory information, alleviating the incomplete analysis issue caused by limited training patches. But this method is limited to deal with images containing objects, which exhibit the same geometric appearance in optical and corresponding SAR image. Hughes *et al.* [22] then proposed a false positive rate (FPR) suppression scheme for SAR-optical matching by augmenting hard negative data. However, hard negative mining requires that the original annotated dataset is large enough such that sufficient negative samples can be explored and then generated for matching task; also, the FPR suppression scheme lacks a comprehensive statistical measure of cross-domain heterogeneity on comparable image content.

When measuring the degree of correspondence between multisources, the heterogeneous information is not equally represented or measured in general matching algorithms under the situation of annotated data scarcity, which heavily restricts final correlation identification effect. To achieve better correspondence identification for optical and SAR data, a domain-transfer learning based generalized correlation analysis (DT-GCA) is proposed. Both data-driven supervisory information and annotated supervisory information are used for the correspondence analysis task, alleviating the inherent limitation of training data scarcity. Specifically, the proposed framework is formulated by optimizing both SAR-adapted domain translator-based CNN (DT-CNN) and generative correspondence analysis (GCA) module. First, inspired from [23], we remove the noise term in conditional adversarial network and reduce the number of transfer layers to construct an optical-SAR transfer network, promoting the control of heterogeneous information following data-constrained supervisory information. Accordingly, the information from optical domain is transferred to SAR domain by DT-CNN, and the transformed optical data are then fed into GCA module for further correspondence identification. It is noteworthy to mention that joint representation has shown to be effective in the fusion of multimedia information [24], and GCA module is designed for multisource data correspondence identification using the joint representation pattern for the first time.

The main contributions can be summarized as follows.

1) Considering both geometric and radiometric differences among optical and SAR data, the proposed DT-GCA integrates the heterogeneous information control, knowledge transfer, and correspondence identification in a unified framework. Besides, an efficient alternative optimization strategy is exploited to train the entire transfer learning model.
2) DT-CNN can align transformed optical data with SAR data, minimizing the discrepancy, and the correspondence of optical and SAR data can be comprehensively measured in terms of any arbitrary probability distribution in the mapped feature space. Therefore, DT-CNN is a successful and adaptive application of adversarial learning in domain discrepancy control.
3) Furthermore, GCA module is specifically designed to select informative features behaving similarly in two sources via multilevel integration, where the transformed optical data are tightly aligned with SAR, further eliminating the diversity influence of different sensors.

The rest of this article is organized as follows. Some related works are briefly reviewed in Section II. The methodology is described in Section III. The experimental results are reported in Section IV. Finally, Section V concludes this article.

## II. Related Method

With the increasing use of correspondence identification for applications such as object recognition [25], wide-baseline stereo [26], multisensor image registration [12], and multiview reconstruction [27], many research works have been conducted to improve some components in the detector-descriptor-similarity pipeline for precise correspondence estimation. In this section, we specifically review related works on correspondence analysis.

Based on the number of data source, existing correspondence identification methods could be roughly divided into two categories: single-source correspondence analysis (SS-CA) and multisource correspondence analysis (MS-CA). SS-CA focuses on finding accurate correspondence from the same source [28], [29], and MS-CA aims at identifying corresponding images from different data sources. Most existing studies are focused on SS-CA.

*SS-CA*: Specifically, SS-CA uses embedding learning methods to learn nonlinear mappings for generating intermediate representations. For example, feature learning methods [30]–[32] investigate descriptors to achieve satisfying correspondence estimation with appropriate distance metric. In [30] and [32], multiple parameterized blocks were integrated for discriminative feature descriptor learning, and different learning algorithms were proposed to find optimal parameters. Rather than simply computing Euclidean distance between descriptor vectors [30], [32], boosting strategy was applied to learn Mahalanobis distance metrics in [31]. In general, conventional methods for correspondence identification involve descriptors and similarity metrics, where most feature descriptors are human-engineering outcomes, such as SIFT [33] or DAISY [34]. While deep learning-based methods construct nonlinear

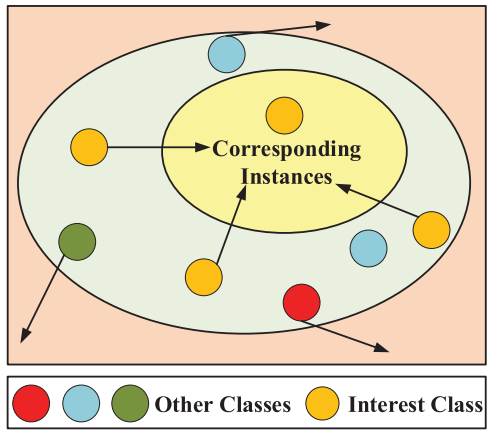
mappings to generate low dimensional representations, whose similarity is computed in some easy-to-compute distance metric [9], [28], [29], realizing accurate correspondence estimation. Moreover, novel elements were developed to compute cross-input differences, which captured local relationships between two input images based on mid-level features from each input image [35].

*MS-CA*: Despite that excellent identification performance has been achieved, aforementioned methods focused on inputs produced by the same sensor, hence their effects may not be satisfactory on multisources correspondence identification with extremely small size of labeled dataset. Generative networks [1], [22], which work as data augmentation processors, have been designed for overcoming data scarcity of MS-CA task. However, these methods focused on data augmentation, and only realized conclusive results as to whether generated data improved classification performance based on existing matching algorithm; hence, these generative networks cannot be regarded as integrated solutions. In [36], this work constructed an integrated framework of heterogeneous data analysis and classification, while it has less effect on binary classification problem, as in the correspondence identification task. In [20], a multisensor correspondence predictor is learned for SAR and optical images. It intended to address the heterogeneous nature of multisensor data, and a specific pseudo-Siamese network rather than a weight-shared Siamese network was designed. However, the pseudo-Siamese network treats multisource data correspondence identification as a simple parameterized block based binary classification problem; therefore, the robustness and accuracy of this straightforward correspondence identification strategy are still behind the expectation for real multisource correspondence estimation applications, especially in the situation of limited training dataset.

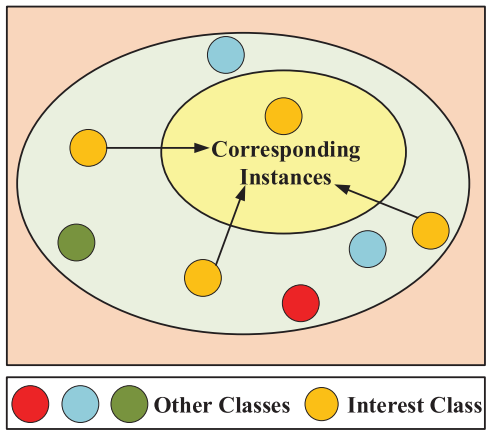
Apparently, finding appropriate embedding feature descriptors, as in SS-CA, contributes to more effective correspondence estimation. However, latent embedding space exploration for MS-CA can be difficult due to strikingly different local changes and global distribution of each source data. In this article, we use transfer learning, which is an improvement of learning in a new task through the transfer of knowledge from a related task that has already been learned. Furthermore, different from existing MS-CA, which only pays attention to either straightforward similarity estimation or simple parameterized block based classification, we mainly aim at domain-transfer learning to integrate two-domain translation (optical $\rightarrow$ SAR) before performing multisource image correspondence identification, which enables the minimization of marginal and conditional distribution discrepancy of multiple sources, ultimately resulting in successful correspondence identification.

## III. Proposed DT-GCA Framework

As illustrated in Fig. 2, the proposed DT-GCA framework includes a domain-transfer module (Part I), called DT-CNN, and a GCA module (Part II), called GCA. In the following, a robust correspondence predictor for optical and SAR images is described in details, and network training is elaborated.




(a)




(b)

Fig. 1. Constraints for adjusting interclass and intraclass distances: (a) feature-matching techniques, and (b) DT-CNN.

### A. Domain Translator-Based CNN (DT-CNN)

Correspondence identification comprises two core challenges in integrative processing of multisource images with limited training samples. First, various intrinsic and extrinsic sensing conditions may lead to image nonhomogeneity for each data source. Second, multisource images reflect great diversity in data volume, scale, and measurement system; the same content of varied sources exhibits different intensity values. Therefore, it is usually difficult to measure similarity based on their intensity values, and the disparity between intensity values of multisource images may lead to coincidental local intensity matching between noncorresponding content [37].

Current focuses of correspondence identification include contextual invariant representations, common structure extraction from varying conditions, and robust similarity measures. Some feature-matching techniques use deep CNN to explore the contextual invariant representation across a shared space (i.e., embedding space), which considers that source instances may not be relevant to target instances. Therefore, it minimizes the distribution differences by reweighting both source and target samples and then applying matching operation based on reweighted features. In spite of the fact that such feature-matching based approaches are intuitively designed [38] and effective for some

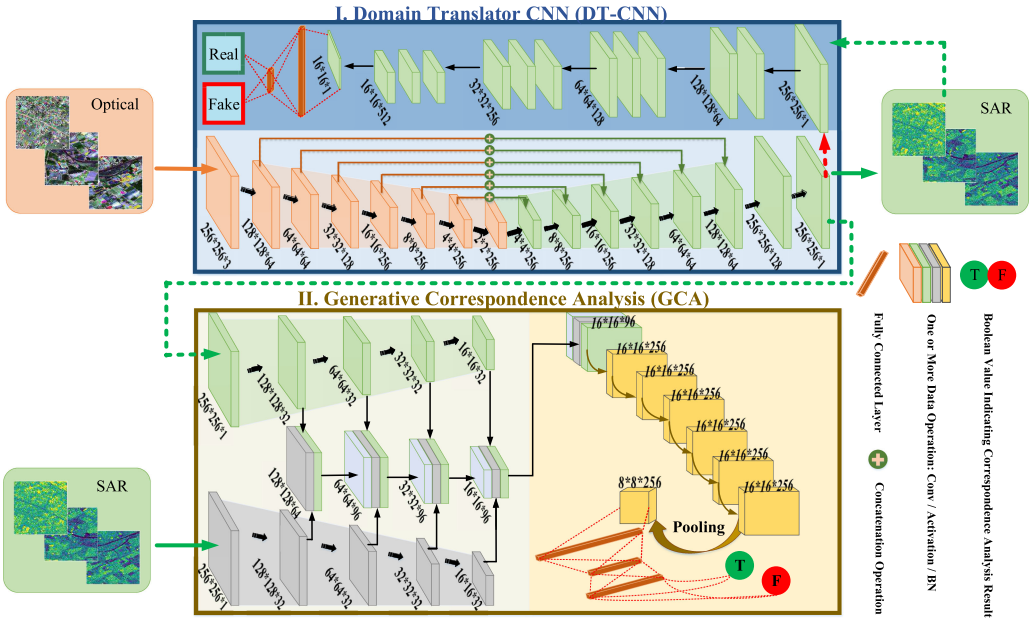



Fig. 2. Overall flowchart of the proposed DT-GCA for multisource data correspondence identification operation process.

correspondence identification problems, they may not be effective for large domain shift problems, where the distribution difference cannot be appropriately reduced in the shared feature space [39]. Hence, feature-matching techniques contain great limitations to assess the correlation across multisource data with strong appearance differences and limited training samples.

The major causes for the feature-matching techniques to have little success on multisource data correspondence analysis are two folded. 1) For feature-matching techniques, our first goal is to identify an ideal embedding space, and then different sources are projected to the unified embedding space, where features of the same scene are expected to have small distance while features of distinct scenes are expected to have large distance value. As shown in Fig. 1(a), to search for the embedding space that satisfies the distance criterion, both interclass and intraclass distance are simultaneously adjusted in the optimization process of feature-matching methods. Multiple constraints can increase the uncertainty in stochastic optimization, leading to weak network generalization ability. 2) In feature-matching methods, it is hard to identify an ideal embedding space guided only by squared $L_2$ distances or specific boolean label (+1/-1, indicating correlation) within limited input instances, which may result in incomplete semantic and contextual characteristic utilization, causing information loss, and nonaccurate identification.

Based on the above analysis, we attempt to transfer optical data to SAR data first, and each learning phase is under the guidance of a complete SAR image as shown in Part I of Fig. 2. By doing so, we eliminate the defect that comes from merely manipulating boolean label to explore appropriate subspace for quite different image pairs, and heterogeneous information between domains could be fully measured and integrated. Besides, since multisource correspondence identification is a typical binary classification issue, the distance of dissimilar patches increases as the distance of similar patches decreases

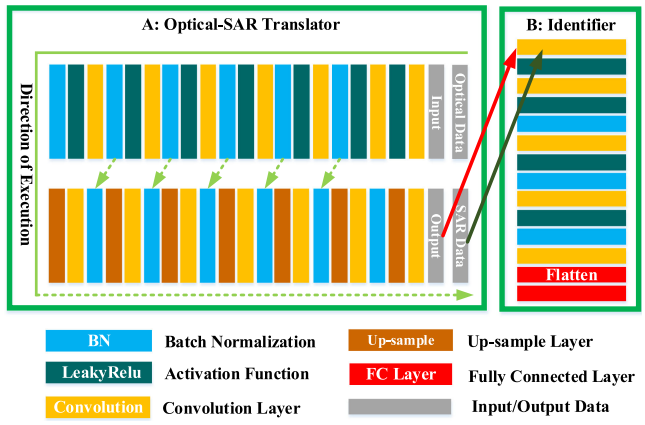



Fig. 3. Overall parameter configuration of the proposed DT-CNN.

undoubtedly, as shown in Fig. 1(b). This way, single distance constraint is adequate to meet the requirement of interclass and intraclass distances adjustment; hence, the optical-SAR domain translator focuses solely on reducing the distance between multisource images containing identical scene. DT-CNN concentrates on enhancing the similarity of correspondence instances across multisource images with one certain sensor-orientation constraint, which promotes more stable parameters optimization and robust feature representations.

Details of DT-CNN (Part I of Fig. 2) is illustrated in Fig. 3. The common CNN model always involves in multiple functional groups, each of which attains specific function. While in the proposed DT-CNN, all the convolution operations are executed with zero padding as in [23], and the convolution stride is set as 2. DT-CNN is a significant component of whole transfer learning framework, which is different from the widely used ordinary GAN, and can be viewed as a successful GAN variant with both domain-translator and identifier. GAN learns a mapping

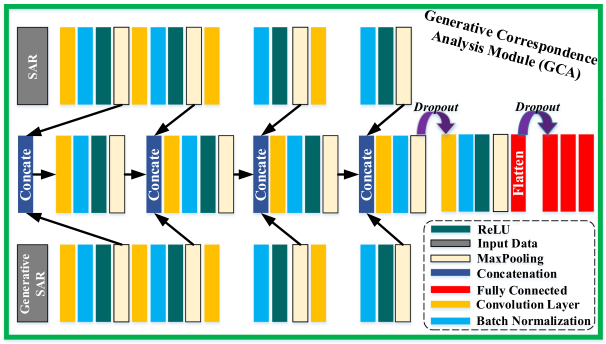



Fig. 4. Overall parameter configuration of the designed GCA.

from random noise vector $\mathbf{z}$ to output image $\mathbf{y}$, $\mathbf{G}: \mathbf{z} \rightarrow \mathbf{y}$ [40]. While conditional GAN learns a mapping from observed image $\mathbf{x}$ to output $\mathbf{y}$, and the random noise vector $\mathbf{z}$ is indispensable, $\mathbf{G}: \{\mathbf{x}, \mathbf{z}\} \rightarrow \mathbf{y}$ [23]. However, the optical-SAR translator module in Fig. 3 concentrates on mapping from observed optical image to corresponding deterministic SAR image, and *DT-CNN does not introduce any noise term so as to obtain a stable optical-SAR translator*, playing an important role in optical knowledge-transfer. Considering about the quantity of training set, the number of hidden layers is reduced referring to [23]. Also, the identifier of DT-CNN is optimized following the guidance of cross-entropy loss instead of mean absolute error loss used in [23]. Besides, the translator of DT-CNN is optimized following the guidance of mean squared loss. The training procedure of DT-CNN is shown in Fig. 6.

### B. GCA Module

Since SAR and optical images lie on different manifolds, it is not advisable to compare them directly by using existing descriptors, especially for a few labeled samples. Using the abovementioned well-trained DT-CNN, transfer learning can be effectively implemented for optical-SAR correspondence identification. Here, transfer learning is an optimization that allows improved performance when modeling the second task (i.e., correspondence identification) based on the first task (i.e., domain-transfer).

As shown in Fig. 4, we use both original SAR and optical-generated SAR (shorted as generative SAR) for correspondence analysis in a joint representation pattern [24]. All the convolution operations are executed with zero padding, and the kernel size is set as 3. Besides, GCA module consists of three network channels, i.e., SAR channel, generative SAR channel, and concatenation channel. Representative characteristics of SAR and generative SAR are excavated through SAR channel and generative SAR channel, respectively. Additionally, concatenation channel is designed for texture comparison of SAR and generative SAR. It is worth mentioning that concatenation channel does not perform the texture comparison on constant observation scale, but on transforming scale from fine to coarse. Moreover, the forward operation of the network links the comparisons of varied scales together, and all the comparisons are adjusted in the process of backpropagation following the guidance of cross-entropy loss. GCA keeps the correlation of different spatial scales, which are expected to be helpful for robust multisource image correspondence identification.

### C. Training Details of Proposed DT-GCA

The training process of the proposed method can be separated into two stages. Specifically, the training strategy of DT-GCA is summarized in Algorithm 1. First, for DT-CNN, only positive image pairs are involved for domain translation. Positive image pairs are optical and SAR images collected over the same area, which ensures the tight correlation between optical and SAR data for robust cross-domain transfer. Particularly, a simple but effective data augmentation method is utilized, which produces additional data without introducing extra labeling costs. In this method, additional training samples are generated by rotating 90°, flipping, and adding noise (Gaussian noise of 0 mean and 0.01–0.03 standard deviation) in the training phase. In doing so, the number of training samples can be increased by a factor of four, ensuring the setting of well-tuned parameters. The imperfect, generated samples act as a form of regularization, which leads to a more discriminative model and reduces the risk of over-fitting [22], [41]. For example, there is no satellites traveling in an orbit, which follows the geometry of a 90° rotation; hence, the parameters optimization procedure needs to cope with the tradeoff between diverse training samples.

As shown in Fig. 6, DT-CNN training consists of two phases, i.e., translator activated phase and identifier activated phase. These two phases are iterated in turns. The training phase I is trained on every data batch, while the training phase II is trained at eight-batch intervals. Accordingly, for training phase I, the number of iterations in each epoch equals to $N/B$ ($N$ denotes the number of training samples, $B$ denote the batch size). For training phase II, the number of iterations in each epoch equals to $N/(B*8)$. Apparently, the number of training iterations for DT-CNN is set based on the training set capacity, rather than using a fixed presetting number. The convergence performance with different training set size is illustrated in Fig. 5. From the convergence presentation, it's clear that the oscillating descent process is stable. Besides, for small-sized training set, it takes much less iterations to reach the convergence state; therefore, the dynamic setting of iterations for training DT-CNN is necessary. Besides, based on the comparison of the convergence situation between Fig. 5(b) and (d), we can see that the convergence fluctuation range of optimization process with regard to the augmented training set can be enlarged, while the oscillating descent and the convergence process are still stable. Also, the generative loss of the network trained on the augmented training set (i.e., 0.2019) is comparable with (even smaller than) that on the nonaugmented set (i.e., 0.2253). Further validation of the data augmentation operation will be provided in Section IV-B.

The weights and bias involved in DT-CNN are initialized with glorot normalization [42], and both translator and identifier are trained using Adam optimization algorithm [43] as it is computationally efficient. The optimization hyperparameters for the translator are fixed to $\beta_1 = 0.9$, $\beta_2 = 0.999$ with a learning rate

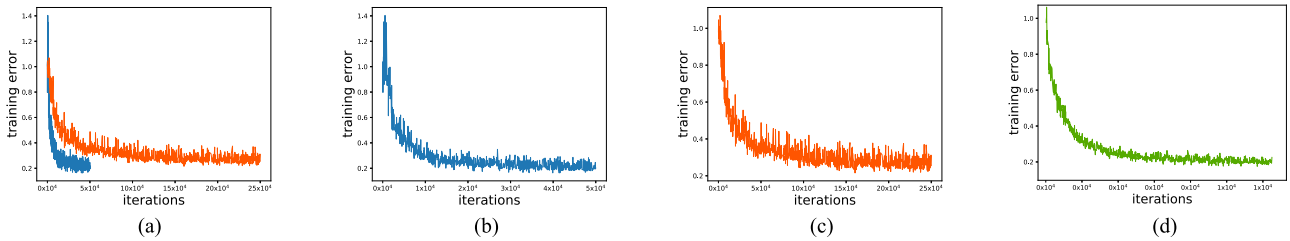



Fig. 5. Convergence performance presentation on SARptical data (the dataset will be introduced in Section IV-A). (a) Convergence performance comparison of 100-sized training set and 500-sized training set. (b) Convergence performance of 100-sized training set. (c) Convergence performance of 500-sized training set. (d) Convergence performance of 100-sized training set without data augmentation.

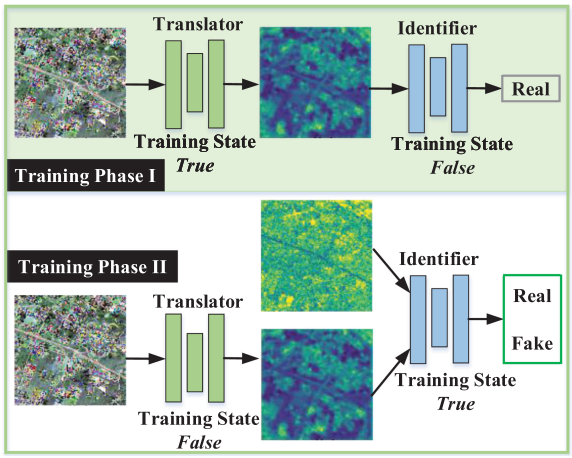



Fig. 6. Training a robust cross-domain translator to map $\mathbf{optical} \rightarrow \mathbf{SAR}$. The identifier learns to classify between fake and real SAR when its training state is activated. The translator learns to construct robust knowledge transfer between optical and SAR images.

of 0.001. The optimization hyperparameters for the identifier are fixed to $\beta_1 = 0.5, \beta_2 = 0.999$ with a learning rate of 0.0002. For the training procedure, DT-CNN is trained in 500 epoches.

Second, optical samples flow through DT-CNN to obtain features in SAR domain for generating inputs of GCA; therefore, GCA module models the second task based on previous domain translation. The weights and bias involved in GCA module are also initialized with glorot normalization. Finally, GCA module is optimized via the policy of Adam [43] with a learning rate of $1e-4$. In the training procedure, GCA module is trained over 100 epoches. For 500-sized training set, the training time is about 5.34 h.

## IV. Experiments and Analysis

In this section, public remote sensing datasets are used to evaluate the proposed DT-GCA. All the programs are implemented using Python language, and the networks are constructed using Keras[1] and Tensorflow.[2]

### A. Experimental Data

The SEN1-2 data [44], including 282,384 pairs of SAR and optical patches extracted from versatile Sentinel-1 and

[1][Online]. Available: https://github.com/fchollet/keras
[2][Online]. Available: http://tensorflow.org/

**Algorithm 1:** Training Strategy for DT-GCA.

1. Stage 1:
2. Initialize all weights of DT-CNN
3. **While** *epoch* $<$ *epochs* **do**
   - *Train the DT-CNN*
4. **End while**
5. Stage 2:
6. Initialize all weights of GCA and load weights of DT-CNN
7. **While** *epoch* $<$ *epochs* **do**
   - *Joint training of GCA and DT-CNN*
8. **End while**

TABLE I
Numbers of Training and Testing Samples for SEN1-2 Data

| # | Class | Training | Testing |
|---|---|---|---|
| 1 | *Positive* | 1246 | 1429 |
| 2 | *Negative* | 1246 | 1429 |
| - | *Total* | 2492 | 2858 |

Sentinel-2 scenes, were acquired from across the globe and throughout all meteorological seasons. Each image patch (both optical patch and SAR patch) consists of 256 $\times$ 256 pixels. For the optical data, only the red, green, and blue channels of sentinel-2 dataset are considered for generating realistically looking RGB images [44]. These data are intended to support the use of intelligence algorithms for multisource remote sensing collaborative observation and used for studies on image matching, image-to-image translation, or images correspondence identification. To investigate the identification performance of the proposed method with limited training samples, we only use the subset of the original dataset. Here, 2675 pairs are selected in the spring season, which have been matched, covering scenes of cities, mountains, woodlands, rivers, cultivated lands, and coastal features. All these patches are marked as positive samples (optical and SAR image-pairs were matched and aligned), and 1246 samples are divided as positive training samples while the remaining samples are served as positive testing samples. Additionally, 2675 negative pairs are randomly generated using the unmatched optical and SAR images. Similarly, 1246 negative samples are divided into training set and the remaining samples are placed into testing set. Therefore, as listed in Table I, there are 2492 pairs of training samples, including 1246 pairs of positive samples and 1246 negative samples, respectively. There are 2858 pairs of testing set samples, including 1429 pairs of positive

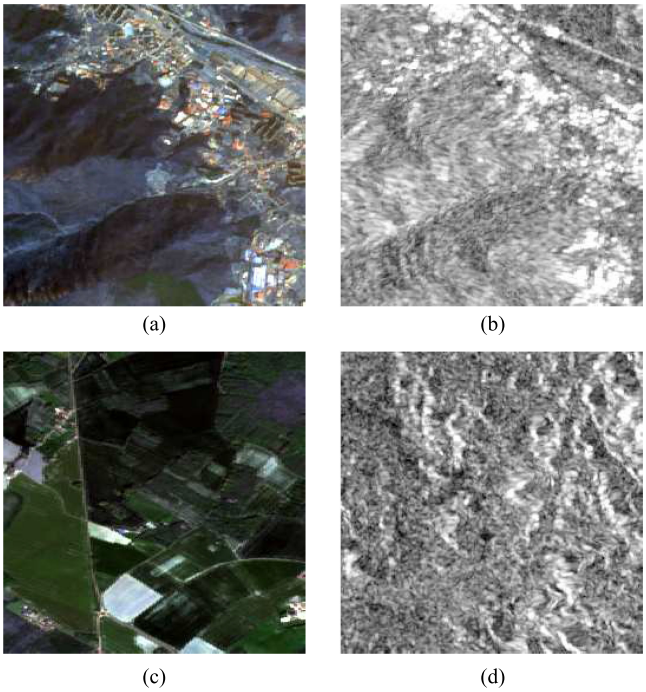



Fig. 7. SEN1-2 data: positive pair for (a) optical image, (b) SAR image; negative pair for (c) optical image, (d) SAR image.

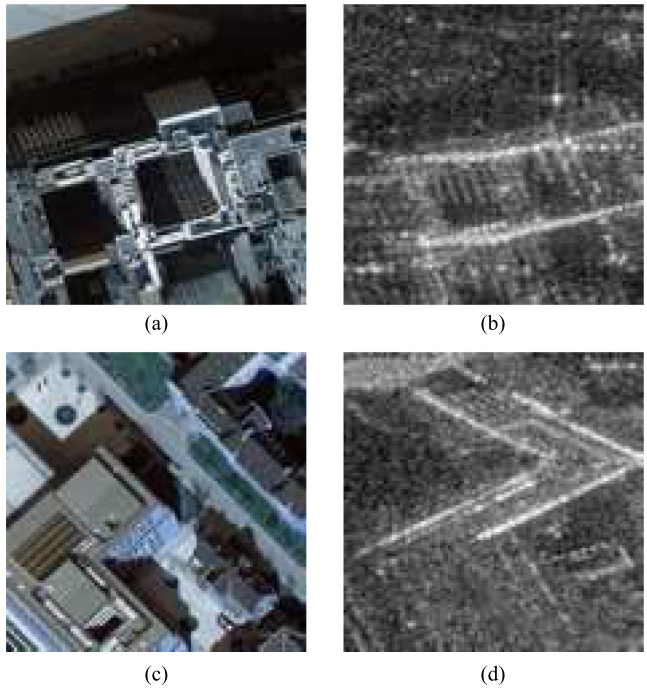



Fig. 8. SARptical data: positive pair for (a) optical image, (b) SAR image; negative pair for (c) optical image, (d) SAR image.

TABLE II
NUMBERS OF TRAINING AND TESTING SAMPLES FOR SARPTICAL DATA

| # | Class | Training | Testing |
|---|---|---|---|
| 1 | $Positive$ | 1500 | 1500 |
| 2 | $Negative$ | 1500 | 1500 |
| - | $Total$ | 3000 | 3000 |

samples and 1429 negative samples, respectively. Fig. 7 shows some samples of positive and negative pairs for SEN1-2 data. Using this lightweight dataset, the correspondence estimation performance of different strategies in case of small samples could be investigated.

The SARptical data [5], containing 10 108 pairs of corresponding high-resolution TerraSAR-X image and aerial UltraCAM optical image patches, were acquired in downtown Berlin. Each image patch consists of 112 $\times$ 112 pixels. The spatial resolution of TerraSAR-X image is about 1m. SARptical data are different from SEN1-2 data in that corresponding optical and SAR images may present the same object from different view angles. Therefore, these data are likely to be more challenging for multisource images correspondence identification, it is helpful to evaluate the geometric-differences processing capacity of the model. Similarly, we only use the subset of the original dataset for the performance investigation with training data scarcity. Here, we select a single correct optical correspondence for each SAR image patch for constructing positive pairs [20]. In addition, negative pairs are randomly generated using unmatched optical and SAR images. As listed in Table II, 3000 training pairs and 3000 testing pairs are eventually obtained. Fig. 8 shows some samples of positive and negative pairs for SARptical data. For SARptical data, there exists a serious overlapping phenomenon among patches, and even for the negative pairs, there is a degree of correlation among patches. Based on such data basis, it is far more challenging to train models for CA. Particularly, for correspondence identification training on GCA, *DT-CNN adopts a pre-trained network on the domain-transfer task of SEN1-2 dataset, which provides better generalization and more stable feature-translation without involving view angle interference*.

### B. Parameter Tuning

In order to validate the effectiveness of each module (DT-CNN and GCA) and the role of data augmentation, we compare the correspondence identification results using different network configurations based on SEN1-2 data. In GCA and the identifier of DT-CNN, the combination of fully connected layer and softmax activation function is regarded as the classifier.

*Performance Estimation for GCA*: GCA module is an independent identification framework that we designed for CA. Specifically, GCA keeps the correlation of different spatial scales, which is helpful for robust multisource image correspondence identification, and it can also be used directly for optical and SAR data correspondence identification. For the separate GCA module, detailed configurations are the same as that shown in Fig. 4 except the input feeding manner. The training strategy of GCA is also the same as that introduced in Section III-C. PS-CNN [20] is a useful identification strategy with concrete parameter setting introduction, and it is an original design for optical and SAR data correspondence identification. As such, the correspondence prediction performance of individual GCA module can be validated by comparing the identification performance with the effective baseline, i.e., PS-CNN. And the comparison results are listed in Table III. More specifically, in this part of comparison, both GCA and PS-CNN utilize original

TABLE III
CORRESPONDENCE IDENTIFICATION PERFORMANCE ESTIMATION FOR GCA ON SEN1-2 DATA

| | Accuracy (%) | Precision | Recall | F1-Score(%) |
|---|---|---|---|---|
| PS-CNN [20] | 94.05 | 0.9139 | 0.9727 | 94.24 |
| GCA | **95.52** | **0.9351** | **0.9783** | **95.62** |

Bold face value is represented for showing that the proposed pipeline has advantages on classification.

TABLE IV
CORRESPONDENCE IDENTIFICATION PERFORMANCE OF GCA WITH/WITHOUT DT-CNN ON SEN1-2 DATA

| | Accuracy (%) | Precision | Recall | F1-Score(%) |
|---|---|---|---|---|
| GCA | 95.52 | 0.9351 | 0.9783 | 95.62 |
| DT-GCA | **98.60** | **0.9826** | **0.9895** | **98.61** |

Bold face value is represented for showing that the proposed pipeline has advantages on classification.

TABLE V
CORRESPONDENCE IDENTIFICATION PERFORMANCE OF DT-GCA WITH/WITHOUT DATA AUGMENTATION ON SEN1-2 DATA

| | Accuracy (%) | Precision | Recall | F1-Score(%) |
|---|---|---|---|---|
| Without | 97.73 | 0.9716 | 0.9832 | 97.74 |
| With | **98.60** | **0.9826** | **0.9895** | **98.61** |

Bold face value is represented for showing that the proposed pipeline has advantages on classification.

TABLE VI
CORRESPONDENCE IDENTIFICATION PERFORMANCE COMPARISONS FOR ALL BASELINES ON SEN1-2 DATA

| | TP | FN | FP | TN | Accuracy(%) | Precision | Recall | F1-Score(%) |
|---|---|---|---|---|---|---|---|---|
| Two-Branch CNN [36] | 953 | 476 | 1113 | 316 | 44.40 | 0.3990 | 0.2211 | 28.46 |
| MATCHNet [9] | 544 | 885 | 677 | 752 | 45.35 | 0.4594 | 0.5262 | 49.05 |
| H-NET [45] | 874 | 555 | 671 | 758 | 57.10 | 0.5773 | 0.5304 | 55.29 |
| H-NET++ [45] | 619 | 810 | 904 | 525 | 40.03 | 0.3933 | 0.3674 | 37.99 |
| ReID-NET [35] | 1245 | 184 | 482 | 947 | 76.70 | 0.8373 | 0.6627 | 73.98 |
| PS-CNN [20] | 1298 | 131 | 39 | 1390 | 94.05 | 0.9139 | 0.9727 | 94.24 |
| DT-GCA | 1404 | 25 | 15 | 1414 | **98.60** | **0.9826** | **0.9895** | **98.61** |

Bold face value is represented for showing that the proposed pipeline has advantages on classification.

optical and SAR data as input, and the detailed configurations of PS-CNN are set as in [20].

*With/Without DT-CNN*: Since SAR and optical images lie on different manifolds, it is not advisable to compare them directly using existing descriptors. The optical-SAR domain translator aims for reducing the measurement distance between multi-source images of the same scene. With well-trained DT-CNN, transfer learning is effectively implemented for optical-SAR correspondence identification. It should be emphasized that we only focus on *wether DT-CNN can help improve the performance of GCA*, rather than improve the visual effect of generative SAR. Table IV shows the correspondence identification performance of GCA module with or without DT-CNN. Apparently, with DT-CNN, GCA module achieves much better correspondence identification performance for optical and SAR data.

*Learning Rate*: Learning rate determines the convergence speed of the training process, and also affects the training performance of correspondence identification. Here, the learning rate is set with an initial value with the policy of Adam in practical implementation. Different learning rates are tested for

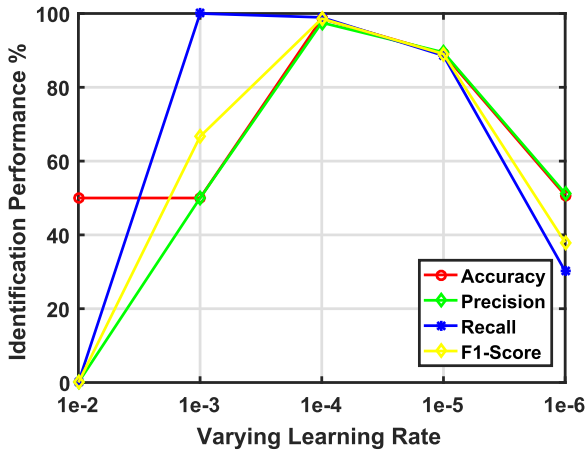
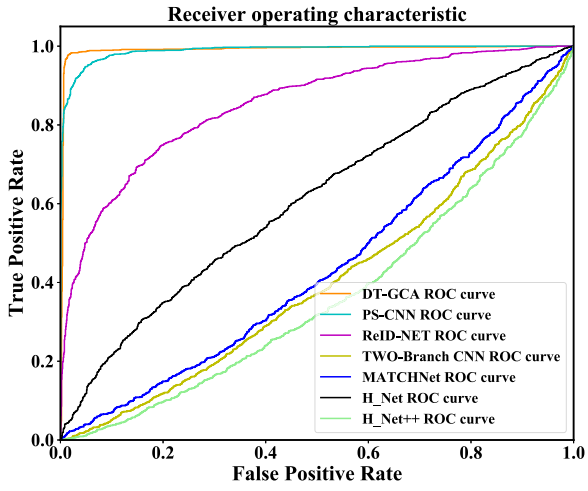
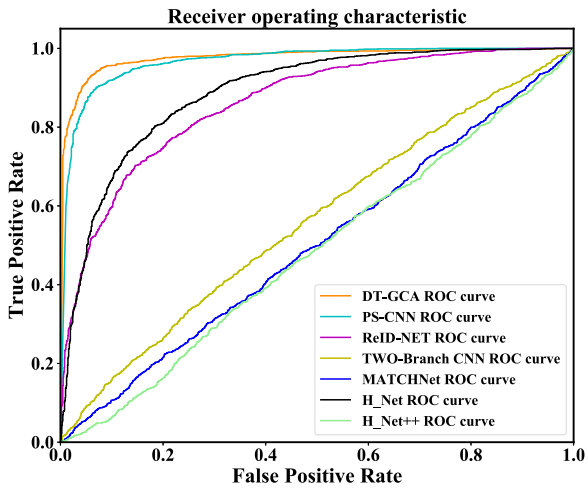



Fig. 9. Correspondence identification performance of DT-GCA under different learning rates on SEN1-2 data.




(a)




(b)

Fig. 10. ROC curves of different methods using (a) SEN1-2 data, (b) SARptical data.

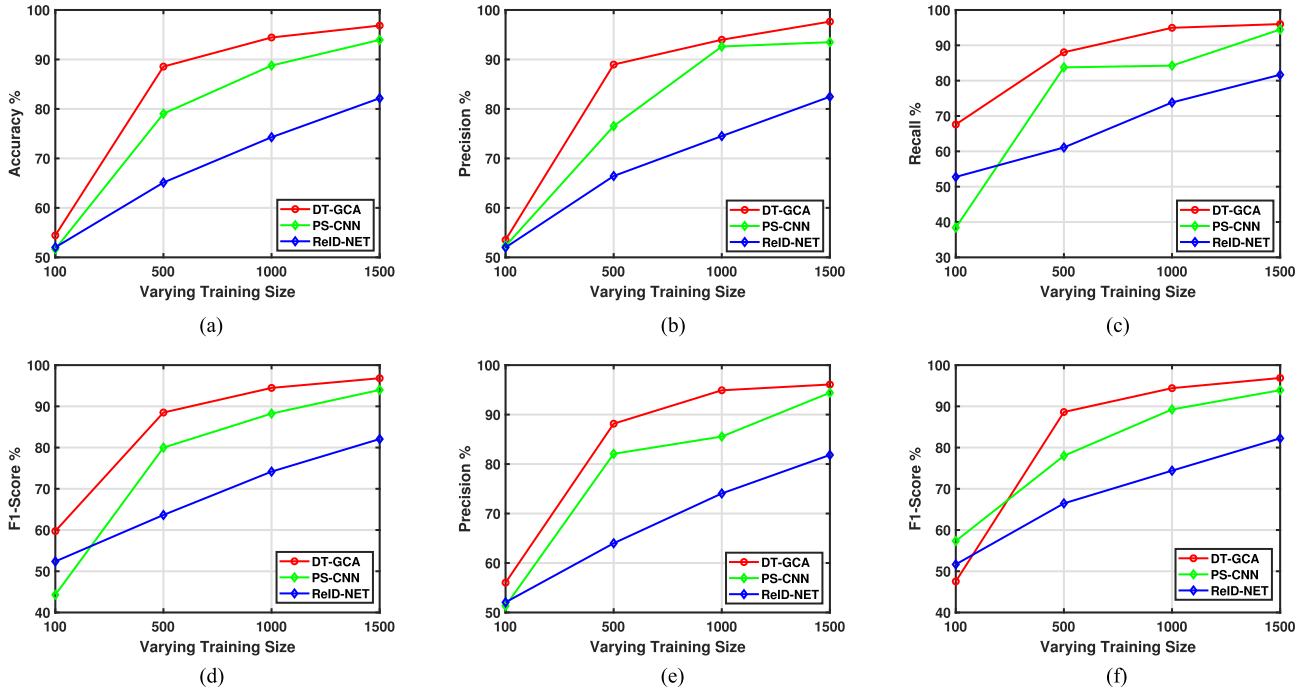



Fig. 11. Correspondence identification performance for DT-GCA, PS-CNN [20], and ReID-NET [35] versus different training data size on SEN1-2 data. (a) Accuracy. (b) Precision for positive class. (c) Recall for positive class. (d) F1-Score for positive class. (e) Precision for negative class. (f) F1-Score for negative class.

TABLE VII
CORRESPONDENCE IDENTIFICATION PERFORMANCE COMPARISONS FOR ALL BASELINES ON SARPTICAL DATA

| | TP | FN | FP | TN | Accuracy (%) | Precision | Recall | F1-Score(%) |
|---|---|---|---|---|---|---|---|---|
| Two-Branch CNN [36] | 898 | 602 | 767 | 733 | 54.37 | 0.5491 | 0.4887 | 51.71 |
| MATCHNet [9] | 876 | 624 | 861 | 639 | 50.50 | 0.5059 | 0.4260 | 46.25 |
| H-NET [45] | 1177 | 323 | 254 | 1246 | 80.77 | 0.7941 | 0.8307 | 81.20 |
| H-NET++ [45] | 733 | 767 | 749 | 751 | 49.47 | 0.4947 | 0.5007 | 49.77 |
| ReID-NET [35] | 1123 | 377 | 294 | 1206 | 77.63 | 0.7618 | 0.8040 | 78.24 |
| PS-CNN [20] | 1333 | 167 | 109 | 1391 | 90.80 | 0.8928 | 0.9273 | 90.97 |
| DT-GCA | 1384 | 116 | 81 | 1419 | **93.43** | **0.9244** | **0.9460** | **93.51** |

Bold face value is represented for showing that the proposed pipeline has advantages on classification.

the proposed DT-GCA, and corresponding identification results are listed in Fig. 9. From the results, the best learning rate is $1e-4$ for the proposed correspondence predictor.

*Data Augmentation*: With the augmented training data, the convergence fluctuation range of the optimization process can be enlarged, while the oscillating descent and the convergence process are still stable as mentioned in Section III-C. In small training set case, deep networks are inclined to learn a function with inappropriate variance such as to perfectly model available training data. And the imperfect, generated samples can act as a form of regularization, which can help networks learn more robust features, leading to a more discriminative model. Different training set configurations are tested for validating the effectiveness of data augmentation as listed in Table V. The comparison shows that the augmentation scheme can affect the CA result.

### C. Correspondence Identification Performance

To validate the effectiveness of the proposed DT-GCA, we compare it with several state-of-the-art methods, including the Two-Branch CNN [36], MATCHNet [9], H-NET [45], H-NET++ [45], ReID-NET [35], and PS-CNN [20]. Besides, for a fair comparison with other baselines, we use the same training and testing samples, as shown in Tables I and II. For a comprehensive comparison, different metrics are used for evaluating the experimental results, including true positive (TP), false positive (FP), true negative (TN), false negative (FN), accuracy, precision, recall, and F1-score. The quantitative results are shown in Tables VI and VII.

From the results, the proposed DT-GCA is obviously superior to the other methods. Taking the SEN1-2 data for example, the proposed DT-GCA yields accuracy of 98.60%, which is 54.2% and over 4% higher than that of the Two-Branch CNN and PS-CNN, respectively, and approximately 53% and 22% higher than that of MATCHNet and ReID-NET, respectively. Thus, it can be concluded that DT-GCA results in higher identification accuracy than other comparison baselines for optical and SAR images correlation estimation.

Fig. 10 further illustrates the effectiveness of the proposed DT-GCA. For evaluating all baselines, receiver operating characteristic (ROC) curves are generated by thresholding the distance between optical-SAR feature pairs in the descriptor space. From the results, DT-GCA is superior to all other baselines; specifically, taking SEN1-2 data for example, when compared with PS-NET, DT-GCA produces higher probability of identification when the FPR varies in a wide range as shown in Fig. 10. For example, when FPR is relatively small (e.g., 0.1), true positive rate of DT-GCA achieves approximately 98% while that of PS-NET is as small as around 70%. To facilitate comparison, Table VIII lists the area under ROC curve (AUC) values for all the predictors, and these AUC values are consistent to the

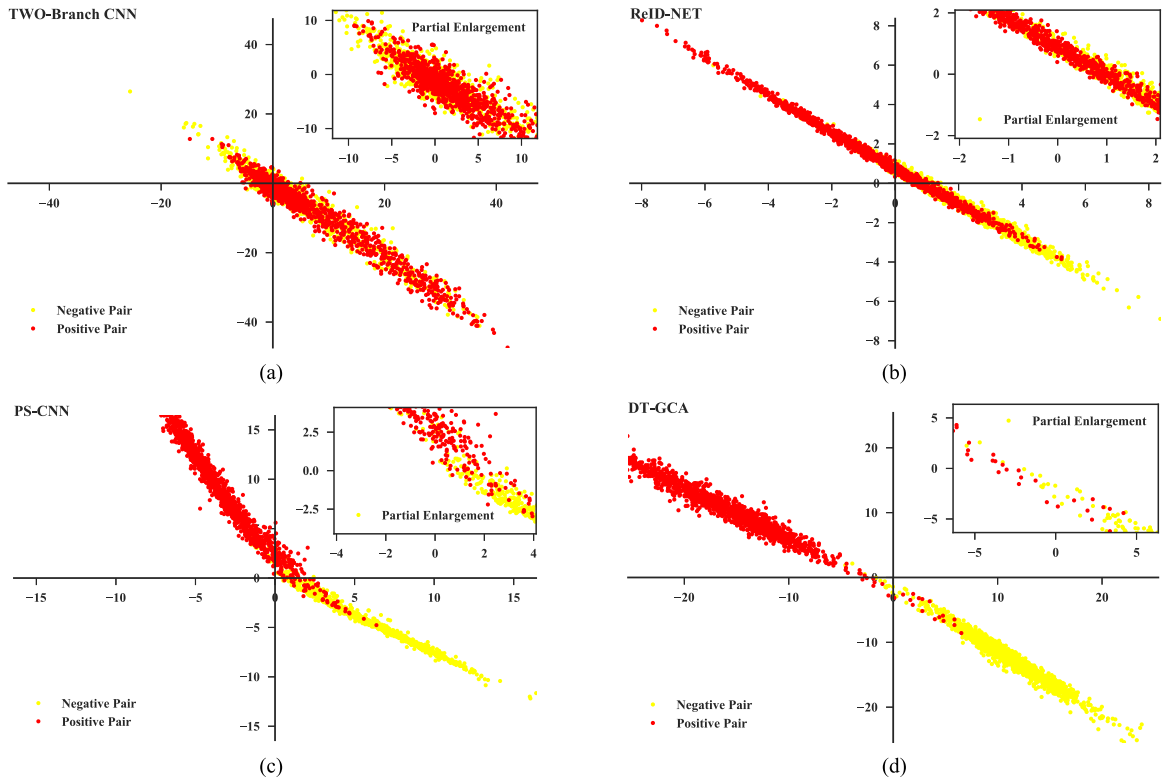



Fig. 12. Distribution of deeply-learned features under the supervision of softmax loss on SEN1-2 data: (a) Two-Branch CNN [36]. (b) ReID-NET [35]. (c) PS-CNN [20]. (d) DT-GCA.

TABLE VIII
AUC COMPARISONS FOR ALL AFOREMENTIONED METHODS

| Methods | AUC(%) | |
|---|---|---|
| | SEN1-2 Dataset | SARptical Dataset |
| Two-Branch CNN [36] | 40.06 | 56.01 |
| MATCHNet [9] | 43.49 | 50.11 |
| H-NET [45] | 60.79 | 88.75 |
| H-NET++ [45] | 36.50 | 48.38 |
| ReID-NET [35] | 85.78 | 86.11 |
| PS-CNN [20] | 98.18 | 96.73 |
| DT-GCA | **99.88** | **97.67** |

Bold face value is represented for showing that the proposed pipeline has advantages on classification.

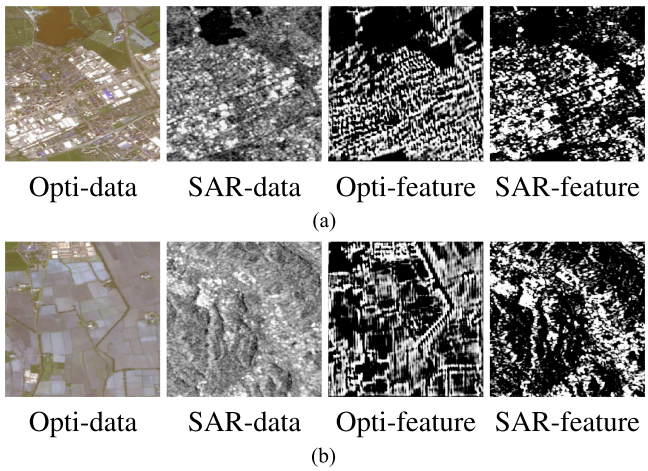



Fig. 13. Spatial features of the proposed DT-GCA on SEN1-2 data (optical is abbreviated as opti). (a) Positive pair. (b) Negative pair.

results shown in Fig. 10. Apparently, the proposed DT-GCA always provides the best performance.

Fig. 11 illustrates the correspondence identification performance on SEN1-2 data with different size of positive and negative training samples. The relationship between generalization performance and training set size can also be clearly represented through the comparison. Particularly, the training set is expanded to 3000 pairs for providing a detailed sensitivity comparison report with equal sample-interval. In Fig. 11, the value of 100 represents that the number of positive training pairs is 100, and the number of negative training pairs is also 100. Hence, total number of training samples is 200. When the size of training set is extremely small, all baselines run into the problem of bad generalization. And the situation occurs because corresponding model over-fits the training dataset yet has poor fitting ability on the unpredicted set. With the increase in the number of training samples, the identification performance of all baselines is increasing accordingly. The proposed DT-GCA consistently outperforms other methods in terms of different training data size. For example, in Fig. 11(a), with 1000 training samples, the accuracy of the proposed DT-GCA is 88.5% while the accuracies of other methods are all below 80%. Through the comparison of identification performance on different-sized training set, it can be demonstrated that the proposed framework is robust to the situations of small training sample sizes.

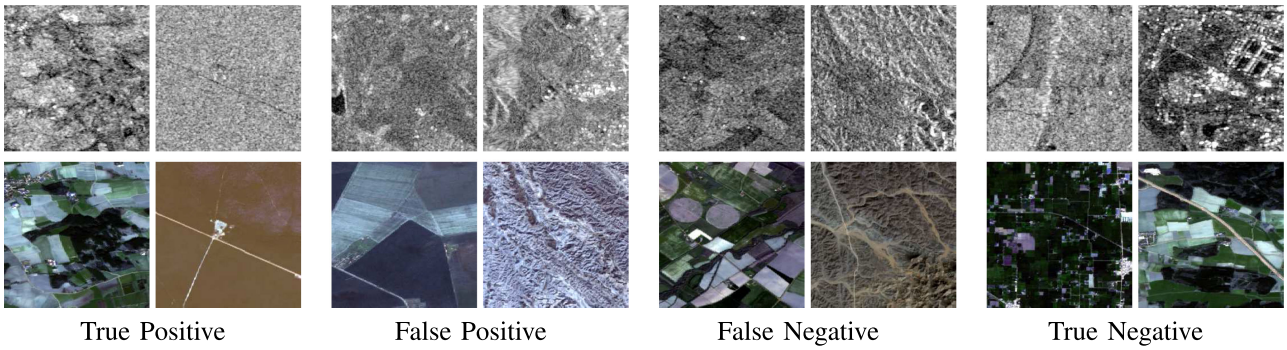



Fig. 14. Correspondence identification performance of the proposed DT-GCA on SEN1-2 data.

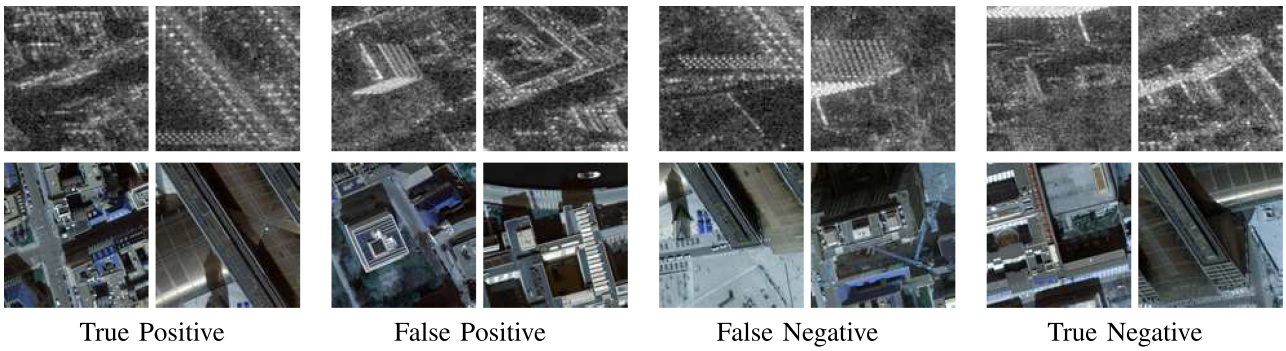



Fig. 15. Correspondence identification performance of the proposed DT-GCA on SARptical data.

For precise correspondence estimation, the deeply learned features need to be discriminative for label prediction. In other words, the deeply learned features dominate the correspondence identification performance of specific model. In this way, identification does not necessarily depend on the design of sophisticated classifiers, but can be well-classified by the last fully connected layer, which acts like a simple linear classifier. The resulting 2-D deeply learned features are illustrated in Fig. 12, which depict the feature distributions for different baselines. All features are obtained through the last fully connected layer, and these 2-D vectors are released before softmax activation function. Visual results indicate that deep features derived from the proposed DT-GCA are more separable and discriminative as shown in Fig. 12(d). Fig. 13 further depicts the original inputs and spatial features extracted from middle layer of DT-GCA (the first layer of GCA module). For multisource data correspondence identification, the difficulty lies in the heterogeneous information representation and measurement. As shown in Fig. 13, the heterogeneity of comparable content between cross-domain data can be effectively controlled based on the DT-GCA guided feature extraction. Taking the positive pair as an example, the structural similarity [46] of original optical and SAR data is 0.0986, while after feature extraction, the similarity is increased to 0.9175. Hence, the proposed method can effectively control the heterogeneous representation of multisource, leading to the superior discriminant results shown in Fig. 12. Moreover, visualization of partial identification results for the proposed method are illustrated in Figs. 14 and 15.

## V. Conclusion

In this article, a DT-GCA was proposed for accurate multisource images correspondence prediction in limited training sample environment. The proposed DT-GCA network combined two steps, i.e., DT-CNN and GCA, with advantages of knowledge transfer and the adaption of transfer learning on following correspondence estimation with the well designed GCA module. Validated on two multisource datasets, experimental results showed that the proposed DT-GCA could yield substantial improvements over several baselines, including the most recent techniques when the training set is limited. In future work, we will explore the possibility for improvement on domain adaption with different view angles to enhance the representation ability of domain-translator and reduce generalization error of the following correspondence estimation.

## References

[1] D. Quan, S. Wang, X. Liang, and R. Wang, "Deep generative matching network for optical and SAR image registration," in *Proc. Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, Jul. 2018, pp. 6215–6218.

[2] M. Schmitt and X. X. Zhu, "Data fusion and remote sensing: An ever-growing relationship," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 4, pp. 6–23, Dec. 2016.

[3] L. Hughes and M. Schmitt, "A semi-supervised approach to SAR-optical image matching," *Int. Soc. Photogrammetry Remote Sens. Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. IV-2/W7, pp. 71–78, Sep. 2019.

[4] R. Hong, L. Zhang, and D. Tao, "Unified photo enhancement by discovering aesthetic communities from flickr," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1124–1135, Jan. 2016.

[5] Y. Wang, X. X. Zhu, B. Zeisl, and M. Pollefeys, "Fusing meter-resolution 4-d InSAR point clouds and optical images for semantic urban infrastructure monitoring," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 14–26, Jan. 2017.
[6] Y. Li, Y. Zhang, X. Huang, and J. Ma, "Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6521–6536, Nov. 2018.
[7] T. Bürgmanna, W. Koppea, and M. Schmitt, "Matching of TerraSAR-X derived ground control points to optical image patches using deep learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 241–248, Nov. 2019.
[8] P. Fischer, P. Schuegraf, N. Merkle, and T. Storch, "An evolutionary algorithm for fast intensity based image matching between optical and SAR satellite imagery," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. IV-3, pp. 83–90, Apr. 2018.
[9] X. Han, T. Leung, Y. Jia, and R. Sukthankar, "Matchnet: Unifying feature and metric learning for patch-based matching," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 3279–3286.
[10] C. Qiu, M. Schmitt, and X. Zhu, "Towards automatic SAR-optical stereogrammetry over urban areas using very high resolution imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 138, pp. 218–231, Jan. 2017.
[11] B. Fan, C. Huo, C. Pan, and Q. Kong, "Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 657–661, Jul. 2013.
[12] Y. S. Kim, J. H. Lee, and J. B. Ra, "Multi-sensor image registration based on intensity and edge orientation information," *Pattern Recognit.*, vol. 41, no. 11, pp. 3356–3365, Nov. 2008.
[13] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2941–2958, May 2017.
[14] D. Quan, S. Wang, M. Ning, and T. Xiong, "Using deep neural networks for synthetic aperture radar image registration," in *Proc. Int. Geosci. Remote Sens. Symp.*, Beijing, China, Jul. 2016, pp. 2799–2802.
[15] V. Balntas, E. Johns, L. Tang, and K. Mikolajczyk, "PN-Net: Comjoined triple deep network for learning local image descriptors," 2016, *arXiv:1601.05030*.
[16] P. Fischer, A. Dosovitskiy and T. Brox, "Descriptor matching with convolutional neural networks: A comparison to SIFT," 2014, *arXiv:1405.5769*.
[17] J. Zbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 1592–1599.
[18] N. Merkle, W. Luo, S. Auer, R. Müller, and R. Urtasun, "Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images," *Remote Sens.*, vol. 9, no. 6, Jun. 2017, Art. no. 586.
[19] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 4353–4361.
[20] L. H. Hughes, M. Schmitt, L. Mou, and Y. Wang, "Identifying corresponding patches in SAR and optical images with a pseudo-siamese CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 784–788, May 2018.
[21] N. Merkle, S. Auer, R. Müller, and P. Reinartz, "Exploring the potential of conditional adversarial networks for optical and SAR image matching," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 6, pp. 1811–1820, Mar. 2018.
[22] L. H. Hughes, M. Schmitt, and X. Zhu, "Mining hard negative samples for SAR-optical image matching using generative adversarial networks," *Remote Sens.*, vol. 10, no. 10, Sep. 2018, Art. no. 1552.
[23] P. Isola, J. Zhu, and T. Zhou, "Image-to-image translation with conditional adversarial networks," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Honolulu, Hawaii, USA, Jul. 2017, pp. 1125–1134.
[24] V. Vielzeuf, A. Lechervy, S. Pateux, and F. Jurie, "Centralnet: A multilayer approach for multimodal fusion," in *Proc. Eur. Conf. Comput. Vis.*, Munich, Germany, pp. 575–589, Sep. 2018.
[25] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. Int. Conf. Comput. Vis.*, Kerkyra, Corfu, Greece, Sep. 1999, pp. 1150–1157.
[26] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image Vis. Comput.*, vol. 22, no. 10, pp. 761–767, Sep. 2004.
[27] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, New York, NY, USA, Jun. 2006, pp. 519–528.
[28] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, Jun. 2005, pp. 539–546.
[29] R. Salakhutdinov and G. Hinton, "Learning a nonlinear embedding by preserving class neighbourhood structure," in *Artif. Intell. Statist.*, San Juan, Puerto Rico, Mar. 2007, pp. 412–419.
[30] M. Brown, G. Hua, and S. Winder, "Discriminative learning of local image descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 43–57, Jan. 2011.
[31] T. Trzcinski, M. Christoudias, V. Lepetit, and P. Fua, "Learning image descriptors with the boosting-trick," in *NIPS*, Lake Tahoe, USA, Dec. 2012, pp. 269–277.
[32] K. Simonyan, A. Vedaldi, and A. Zisserman, "Learning local feature descriptors using convex optimisation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1573–1585, Aug. 2014.
[33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[34] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
[35] E. Ahmed, M. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 3908–3916.
[36] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multi-source remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.
[37] X. Liu, Z. Lei, Q. Yu, X. Zhang, Y. Shang, and W. Hou, "Multi-modal image matching based on local frequency information," *EURASIP J. Adv. Signal Process.*, vol. 1, no. 3, pp. 1–11, Jan. 2013.
[38] H. Qian, J. Yue, and M. Chen, "Research progress on feature matching of SAR and optical images," *ISPRS Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-3/W10, pp. 77–82, Feb. 2020.
[39] C. Chen, B. Jiang, Z. Cheng, and XinyuJin, "Joint domain matching and classification for cross-domain adaptation via ELM," *Neurocomputing*, vol. 349, pp. 314–325, Jan. 2019.
[40] I. Goodfellow, J. Pouget-Abadie, M. Mirza, and B. Xu, "Generative adversarial nets," in *Proc. NIPS*, Montrèal, Canada, Dec. 2014, pp. 2672–2680.
[41] Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by GAN improve the person re-identification baseline in vitro," in *Proc. Int. Conf. Comput. Vis.*, Venice, Italy, 2017, pp. 3774–3782.
[42] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, 2010.
[43] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computer Science*, 2014, *arXiv:1412.6980*.
[44] M. Schmitt, L. Hughes, and X. Zhu, "The SEN1-2 dataset for deep learning in SAR-optical data fusion," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. IV-1, pp. 141–146, Oct. 2018.
[45] W. Liu, X. Shen, C. Wang, Z. Zhang, C. Wen, and J. Li, "H-net: Neural network for cross-domain image patch matching," in *IJCAI*, Stockholm, Sweden, 2018, pp. 856–863.
[46] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

**Mengmeng Zhang** (Member, IEEE) received the B.S. degree in computer science and technology from the Qingdao University of Science and Technology, Qingdao, China, in 2014, and the Ph.D. degree in control science and engineering from the Beijing University of Chemical Technology, Beijing, China, in 2019.

She is currently a Post-Doctoral Researcher with the School of Information and Electronics, Beijing Institute of Technology, Beijing. Her research interests include remote sensing image processing and pattern recognition.

**Wei Li** (Senior Member, IEEE) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-Sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

He was a Postdoctoral Researcher with the University of California, Davis, CA, USA. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. His research interests include hyperspectral image analysis, pattern recognition, and data compression.

Dr. Li is an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

**Ran Tao** (Senior Member, IEEE) was born in 1964. He received the B.S. degree from the Electronic Engineering Institute of PLA, Hefei, China, in 1985, and the M.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1990 and 1993, respectively.

He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. He was a Distinguished Professor of the Changjiang Scholars Program, in 2009. He was a Chief Professor of the Program for Changjiang Scholars and Innovative Research Team in University from 2010 to 2012. He has been a Chief Professor of the Creative Research Groups.

Dr. Tao was the recipient of the National Science Foundation of China for Distinguished Young Scholars, in 2006, and the First Prize of Science and Technology Progress, in 2006 and 2007, and the First Prize of Natural Science, in 2013, both awarded by the Ministry of Education.

**Song Wang** (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign (UIUC), Champaign, IL, USA, in 2002.

He was a Research Assistant with the Image Formation and Processing Group, Beckman Institute, UIUC, from 1998 to 2002. In 2002, he joined the Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA, where he is currently a Professor. His current research interests include computer vision, image processing, and machine learning.

Dr. Wang is a member of the IEEE Computer Society. He is currently serving as the Publicity/Web Portal Chair for the Technical Committee of Pattern Analysis and Machine Intelligence of the IEEE Computer Society and as an Associate Editor for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, PATTERN RECOGNITION LETTERS, and ELECTRONICS LETTERS.