Learning Source-Invariant Deep Hashing Convolutional Neural Networks for Cross-Source Remote Sensing Image Retrieval

Yansheng Li, Yongjun Zhang, Xin Huang, *Senior Member, IEEE*, and Jiayi Ma

***Abstract*— Due to the urgent demand for remote sensing big data analysis, large-scale remote sensing image retrieval (LSRSIR) attracts increasing attention from researchers. Generally, LSRSIR can be divided into two categories as follows: uni-source LSRSIR (US-LSRSIR) and cross-source LSRSIR (CS-LSRSIR). More specifically, US-LSRSIR means the inquiry remote sensing image and images in the searching data set come from the same remote sensing data source, whereas CS-LSRSIR is designed to retrieve remote sensing images with a similar content to the inquiry remote sensing image that are from a different remote sensing data source. In the literature, US-LSRSIR has been widely exploited, but CS-LSRSIR is rarely discussed. In practical situations, remote sensing images from different kinds of remote sensing data sources are continually increasing, so there is a great motivation to exploit CS-LSRSIR. Therefore, this paper focuses on CS-LSRSIR. To cope with CS-LSRSIR, this paper proposes source-invariant deep hashing convolutional neural networks (SIDHCNNs), which can be optimized in an end-to-end manner using a series of well-designed optimization constraints. To quantitatively evaluate the proposed SIDHCNNs, we construct a dual-source remote sensing image data set that contains eight typical land-cover categories and 10 000 dual samples in each category. Extensive experiments show that the proposed SIDHCNNs can yield substantial improvements over several baselines involving the most recent techniques.**

***Index Terms*— Cross-source large-scale remote sensing image retrieval (CS-LSRSIR), dual-source remote sensing image data set (DSRSID), remote sensing big data (RSBD) management and mining, source-invariant deep hashing convolutional neural networks (SIDHCNNs).**

Manuscript received January 7, 2018; revised April 25, 2018; accepted May 20, 2018. Date of publication June 14, 2018; date of current version October 25, 2018. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0505003, in part by the National Natural Science Foundation of China under Grant 41601352, Grant 41522110, and Grant 61773295, in part by the China Postdoctoral Science Foundation under Grant 2016M590716 and Grant 2017T100581, and in part by the China Scholarship Council under Grant 201706275025. *(Corresponding authors: Yongjun Zhang; Xin Huang.)*

Y. Li and Y. Zhang are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: yansheng.li@whu.edu.cn; zhangyj@whu.edu.cn).

X. Huang is with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China, and also with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: xhuang@whu.edu.cn).

J. Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jiayima@whu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2839705

I. Introduction

With the rapid development of remote sensing observation technologies, our ability to acquire remote sensing data has increased to an unprecedented level. Remote sensing data owes its remarkable characteristics to the four V's (i.e., volume, variety, velocity, and veracity) of big data. We have entered an era of remote sensing big data (RSBD) [1]. Although RSBD provides a data-driven possibility for coping with various challenges, new theories and methods for addressing RSBD should be redeveloped as traditional methods with high computation and storage complexity that may not be applicable to RSBD [2]. As one of the most fundamental techniques for the management and mining of RSBD, content-based large-scale remote sensing image retrieval (LSRSIR) has many potential applications, such as for disaster rescue, and attracts increasing attention from the researchers [3]–[6].

Generally speaking, LSRSIR can be roughly divided into two categories: uni-source LSRSIR (US-LSRSIR) and cross-source LSRSIR (CS-LSRSIR). More specifically, US-LSRSIR is intended to retrieve remote sensing images with a similar content to the inquiry remote sensing image where all remote sensing images come from the same remote sensing data source, and CS-LSRSIR is intended to retrieve remote sensing images with a similar content to the inquiry remote sensing image where the inquiry image and the images in the searching data set come from different remote sensing data sources.

In the early stage, high-dimensional feature descriptors such as local invariant [7], morphological [8], and textural [9], [10] have been introduced to address the content-based remote sensing image retrieval task. As is well known, exhaustively comparing the high-dimensional feature descriptor of an inquiry remote sensing image with each image in a data set is computationally expensive and becomes impossible when the volume of a data set is very huge. To pursue the scalability, there exist two potential solutions: improving the feature search strategy and reducing the dimension of feature descriptors. The former solution can be implemented by tree-based methods [3], [11], [12], which split data spaces into subspaces and record the divisions via a tree structure. The tree-based methods indeed lift the search speed, but would significantly hurt the retrieval performance, especially when the dimension of the original feature descriptor is very high. In fact, the dimension of feature descriptors of remote sensing images is often very high. To avoid the drawback of tree-based methods, researchers turn to the alternative solution (i.e., feature reduction methods). Recently, hashing learning methods [4]–[6], [13] are introduced to address LSRSIR and show promising results. These hashing learning methods [4]–[5], [13] take the high-dimensional feature vector (HDFV) as the input, and project it to the low-dimensional binary feature vector (LDBFV). As the dimension of LDBFV is very low and each element in LDBFV is binary, the similarity between LDBFVs can be efficiently measured by a small number of bit operations (e.g., the hamming distance). Accordingly, the complexity of exhaustive searches using LDBFV is dramatically reduced relative to that of HDFV. To incorporate the powerful feature representation merit of deep learning, Li *et al.* [6] propose deep hashing convolutional neural networks (DHCNNs) to automatically extract the semantic feature from the raw image and map the semantic feature to LDBFV in one unified framework. Benefiting from the respective merits of deep learning and hashing learning, DHCNNs remarkably outperform the hand-crafted feature-based hashing learning methods [4], [5], [13]. As a whole, all of the aforementioned achievements work around the single source remote sensing image retrieval task, and tree-based methods [3], [11], [12] and hashing learning methods [4]–[6], [13] are potential candidates to address US-LSRSIR. In reality, remote sensing images from different sources are continually increasing, so there emerge more and more demands on CS-LSRSIR. Although kinds of methods [3]–[6], [11]–[13] have been proposed for US-LSRSIR, they cannot be readily extended to address CS-LSRSIR because of the source shift problem, which is also called the data shift problem [14]. To the best of our knowledge, there do not exist any effective methods to support CS-LSRSIR. With this consideration, this paper, for the first time, exploits the method for CS-LSRSIR.

In the artificial intelligence domain, the cross-modal retrieval technique [15]–[17] has been widely exploited. It first trains mapping functions to project the information from different modalities to a unified feature space, and then, the cross-modal retrieval can be implemented by the similarity measurement based on the unified feature space. At first glance, we may transfer the cross-modal retrieval technique employed in the machine-learning domain to CS-LSRSIR. In the literature, all existing cross-modal retrieval methods [15]–[17] work for cross-modal retrieval between natural images and documents. More specifically, based on the hand-crafted features of images and documents, unsupervised canonical correlation analysis (CCA) [15], and supervised semantic correlation maximum (SCM) [16] have been proposed. To fully incorporate the merits of deep learning [18]–[20], deep cross-modal hashing (DCMH) [17] could jointly learn the feature representation and projection way and significantly outperform cross-modal retrieval methods based on hand-crated features. In addition, DCMH was composed of deep convolutional neural networks (DCNNs) for images and deep fully connected networks (DFCNs) for documents. In DCMH, DCNNs were inherited from the deep networks that were pretrained on a similar natural image data set that can effectively decrease the training difficulty, and the input of DFCNs is still the hand-crafted feature (i.e., the word frequency feature). As remote sensing images differ considerably from natural images in terms of the spatial and spectral resolution, the deep networks pretrained on a natural image data set cannot be directly transferred to initialize the deep feature representation networks for remote sensing images, which significantly increases the training difficulty, especially in a cross-modal optimization circumstance. Hence, DCMH cannot be directly utilized to tackle with CS-LSRSIR. Due to the particular complexity of CS-LSRSIR, how to design and optimize cross-modal deep networks for CS-LSRSIR should be explored further.

With the aforementioned consideration, this paper proposes source-invariant deep hashing convolutional neural networks (SIDHCNNs) to cope with CS-LSRSIR. More specifically, SIDHCNNs are composed of two networks with different architectures, which are specifically designed based on the spatial–spectral resolution of the remote sensing images from two different data sources. To pursue the scalability, the networks in SIDHCNNs contain hashing layers, which makes the optimization of SIDHCNNs be a discrete optimization problem. Compared with US-LSRSIR [6], SIDHCNNs further suffers from the source shift problem as SIDHCNNs aim at measuring the similarity between remote sensing images from different data sources, and two hybrid networks need to be optimized simultaneously. Considering the aforementioned challenges, we propose a series of optimization constraints, including the intersource pairwise similarity constraint (IRSC), the intrasource pairwise similarity constraint (IASC), the binary quantization loss constraint, and the feature distribution constraint (FDC) to pursue a robust optimization of SIDHCNNs. In addition, the intuitive description and experimental validity of the advocated optimization constraints are given in Sections III-A and IV-B, respectively. Since there does not exist any publicly open multisource remote sensing image data set, this paper proposes a new dual-source remote sensing image data set (DSRSID), which contains eight typical land-cover categories and 10 000 dual samples in each category. Extensive experiments on the proposed DSRSID show that the proposed cross-source remote sensing image retrieval approach that is based on SIDHCNNs can significantly outperform several baselines, including the most recent technique. The main contributions of this paper can be summarized as follows.

1) To the best of our knowledge, this paper, for the first time, reveals the possibility of conducting CS-LSRSIR and shows the potential applications of CS-LSRSIR.
2) This paper proposes SIDHCNNs to cope with CS-LSRSIR where SIDHCNNs can be optimized from scratch in an end-to-end manner. In addition, a series of optimization constraints are advocated to pursue a stable optimization of SIDHCNNs.
3) This paper collects and releases a new DSRSID which is used to evaluate CS-LSRSIR in this paper and benefits promoting the multisource remote sensing image processing technology.

The remainder of this paper is organized as follows. Section II specifically introduces the collected DSRSID.

TABLE I
DESCRIPTION OF THE DUAL SAMPLE

Data source	Satellite sensor	Spatial resolution	Spectral channel	Image size
Panchromatic image	GF-1 panchromatic sensor	2 m	1 channel	256*256
Multi-spectral image	GF-1 multi-spectral sensor	8 m	4 channels	64*64

Section III presents the CS-LSRSIR approach based on SIDHCNNs. Section IV depicts the experimental results in detail. Finally, Section V gives the conclusion of this paper, the applications of our SIDHCNNs, and our future prospects.

II. Dual-Source Remote Sensing Image Data Set

Yang and Newsam [21], Basu *et al.* [22], and Cheng *et al.* [23] in the remote sensing community have proposed a large number of remote sensing image scene data sets, which have effectively promoted the development of remote sensing image scene understanding [24]–[27]. These existing data sets were constructed by only one kind of remote sensing data source and are called uni-source data sets in the following. Intuitively, these uni-source data sets would not be competent for evaluating CS-LSRSIR. To promote the multisource remote sensing image analysis techniques, including the discussed CS-LSRSIR in this paper, it is very urgent to construct a remote sensing image data set containing at least two kinds of remote sensing data sources. To this end, this paper collects a new DSRSID (DSRSID is available at https://pan.baidu.com/s/15ZWaZ2yArnvwcwtead_rpQ).

More specifically, the DSRSID is tiled from two kinds of remote sensing data sources (i.e., panchromatic images and multispectral images) and is manually annotated. The DSRSID is composed of large numbers of dual samples where each dual sample is a combination of one panchromatic image and one multispectral image covering the same ground region. It is noted that the panchromatic image and multispectral image in one dual sample belong to the same land-cover type, but reflect different aspects of the captured ground region because of the spatial and spectral variations. Table I gives a specific description of the dual sample. In Table I, GF-1 is the civil optical satellite that was launched by China in 2013.

In the following, the construction process of the DSRSID is given in detail. Based on the geographical correspondence, dual samples are randomly tiled from over 100 pairs of remote sensing images where each pair of remote sensing images are composed of one large panchromatic image and one large multispectral image that were shot at the same time. To make the constructed data set be universal, over 100 pairs of remote sensing images are randomly sampled from a large span between 116°4′E to 120°44′E and 35°23′N to 36°58′N. Furthermore, the DSRSID is generated by the manual annotation of the dual samples. As a first attempt, the DSRSID contains eight typical land-cover types including aquafarm, cloud, forest, high building, low building, farm land, river, and water. In addition, there are 10 000 dual samples in each land-cover type. Three dual-examples per land-cover type from the DSRSID are visually shown in Fig. 1. In our future work, we may further enrich the DSRSID in terms of the number of land-cover types and the volume of dual samples.

The DSRSID is formulated as $\mathbf{D} = \{(P_i, M_i, L_i)|i = 1, 2, \ldots, N\}$, where $\mathbf{D}$ denotes the set of dual samples, i denotes the index of the dual sample, N stands for the volume of the DSRSID (i.e., the number of dual samples), $P_i \in R^{256\times256}$ is the panchromatic image, $M_i \in R^{64\times64\times4}$ denotes the multispectral image, and L_i denotes the land-cover type.

In this paper, $\mathbf{D} = \{(P_i, M_i, L_i)|i = 1, 2, \ldots, N\}$ is randomly split into two nonoverlapped parts: a training data set $\mathbf{D}^U_{\text{Tr}} = \{(P_i, M_i, L_i)|i = 1, 2, \ldots, V\}$ and a testing data set $\mathbf{D}^U_{\text{Te}} = \{(P_i, M_i, L_i)|i = 1, 2, \ldots, Q\}$, where $N = V + Q$, V is the volume of the training data set, and Q is the volume of the testing data set.

III. Cross-Source Large-Scale Remote Sensing Image Retrieval

To facilitate understanding, we will specifically introduce the CS-LSRSIR method on the basis of the DSRSID that is introduced in Section II. As the DSRSID is a general case, conducting the CS-LSRSIR method on other data sets is straightforward.

As shown in Fig. 2, the proposed CS-LSRSIR approach is composed of two stages: the training stage and the testing stage. More specifically, the training stage is responsible for training SIDHCNNs and the testing stage presents the cross-source remote sensing image retrieval process based on SIDHCNNs. In the following, Section III-A introduces the training stage and Section III-B presents the testing stage.

A. Learning Source-Invariant Deep Hashing Convolutional Neural Networks

As an early attempt to cope with large-scale image retrieval, the hand-crafted feature-based hashing learning methods [4], [5], [13], [28] have been widely exploited as an effective way to reduce the dimension of the high-dimensional hand-crafted feature descriptor. Despite these hashing learning methods having achieved some degree of success, they still do not fulfill the practical application demand. To fully incorporate the merits of deep learning [20], [29]–[31] and hashing learning, deep hashing neural networks [6], [32], [33] have been proposed to automatically learn feature representation

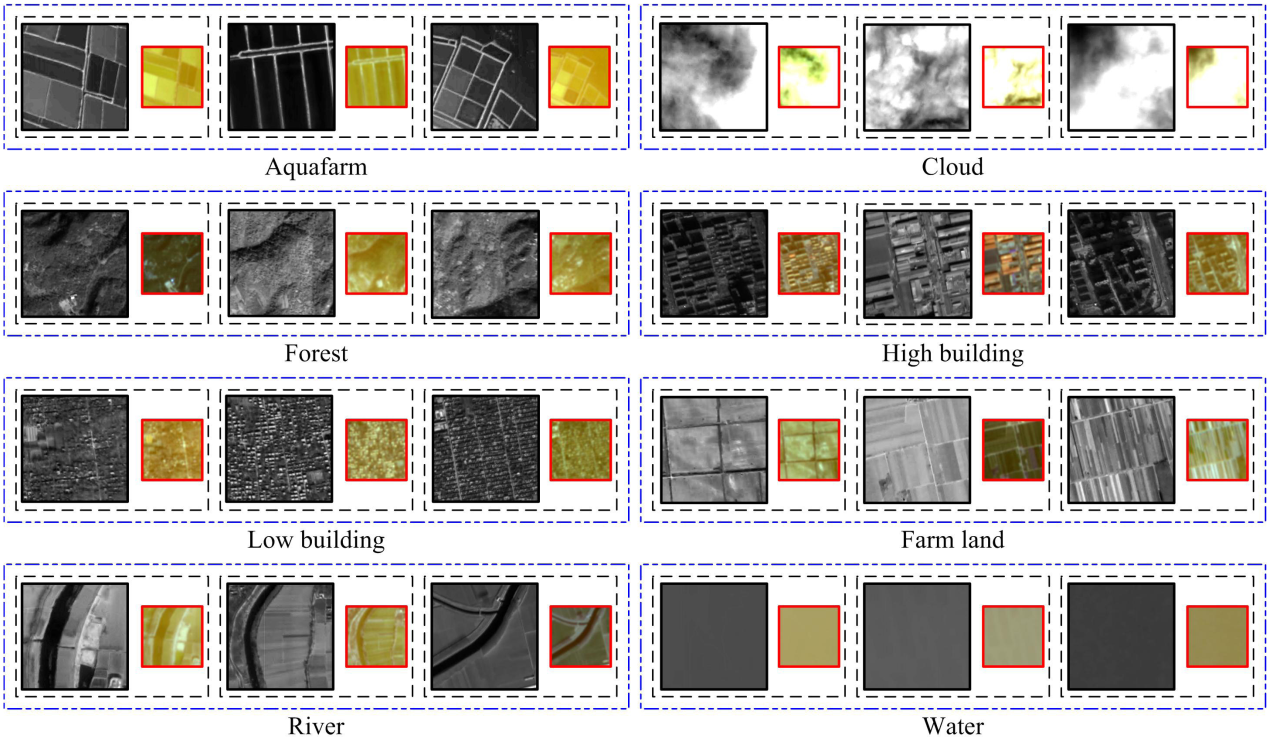


Fig. 1. Visual illustration of the collected DSRSID. In each land-cover type, three randomly sampled dual samples are shown. The images with black solid rectangles were shot by the panchromatic sensor, and the images surrounded by red solid rectangles were captured by the multispectral sensor.

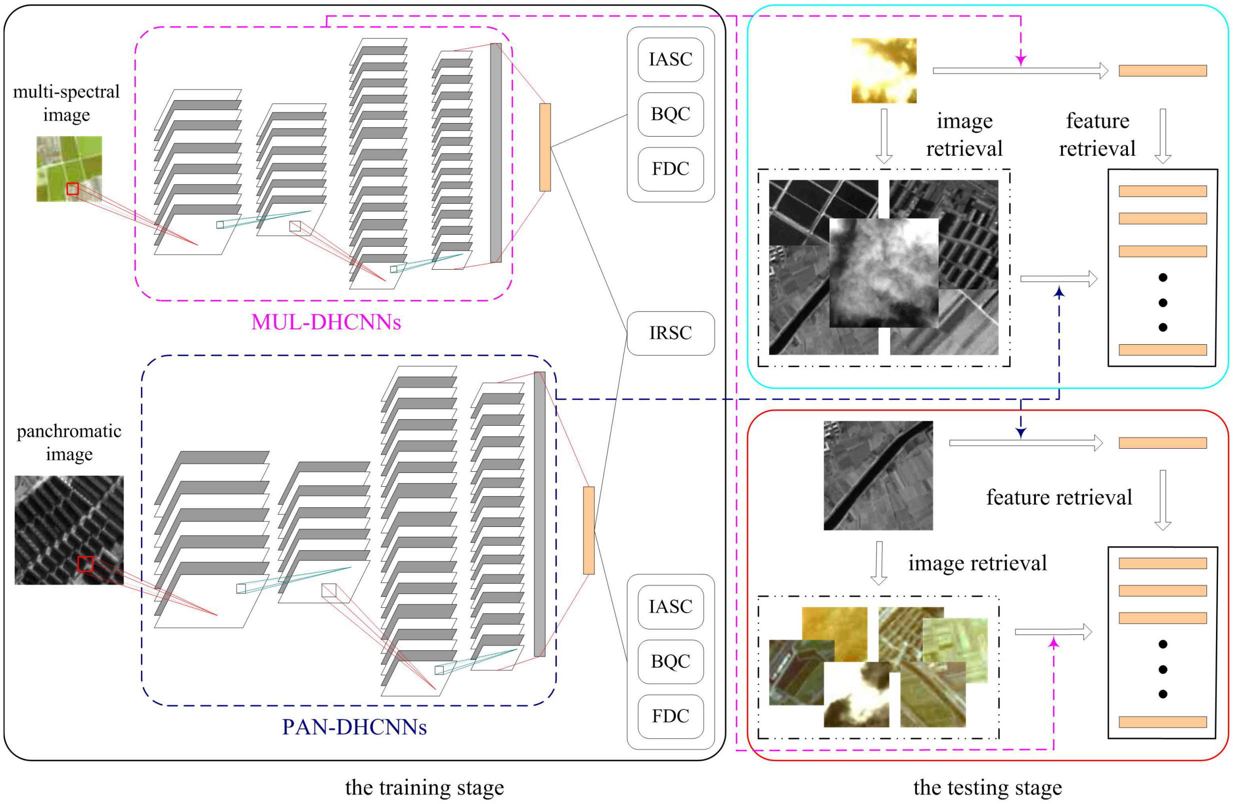


Fig. 2. Workflow of the proposed CS-LSRSIR approach via SIDHCNNs. The proposed CS-LSRSIR method includes two stages: the training stage works on training SIDHCNNs and the testing stage carries out the cross-source remote sensing image retrieval tasks based on the learned SIDHCNNs.

and reduction from data, and become the state-of-the-art technology in the large-scale image retrieval field. Due to the lack of the consideration of multisource data characteristics, these deep hashing neural networks cannot be directly utilized to cope with CS-LSRSIR. Inspired by the success of cross-modal retrieval between images and documents [17], this

TABLE II

ARCHITECTURE OF PAN-DHCNNS

Layer	Configuration
Conv1	filter: $32\times7\times7\times1$; stride1: 3×3 ; pooling: 3×3 ; stride2: 2×2
Conv2	filter: $64\times5\times5\times32$; stride1: 2×2 ; pooling: 2×2 ; stride2: 1×1
Conv3	filter: $128\times5\times5\times64$; stride1: 2×2
Conv4	filter: $256\times5\times5\times128$; stride1: 1×1 ; pooling: 3×3 ; stride2: 1×1
Full5	1024
Full6	1024
Full7	l

TABLE III

ARCHITECTURE OF MUL-DHCNNS

Layer	Configuration
Conv1	filter: $64\times5\times5\times4$; stride1: 2×2 ; pooling: 3×3 ; stride2: 2×2
Conv2	filter: $128\times3\times3\times64$; stride1: 2×2 ; pooling: 3×3 ; stride2: 1×1
Conv3	filter: $256\times3\times3\times128$; stride1: 1×1 ; pooling: 2×2 ; stride2: 1×1
Full4	1024
Full5	1024
Full6	l

paper proposes a series of cross-source constraints to train SIDHCNNs to address CS-LSRSIR.

In the following, we introduce the detail of the constructed SIDHCNNs, the formulated objective function for optimizing SIDHCNNs, and how to learn SIDHCNNs from scratch, respectively.

1) Architecture of SIDHCNNs: Instead of using one unified DHCNNs architecture in US-LSRSIR [6], [32], [33], we design two different DHCNNs architectures for different remote sensing image sources to fully mine the visual cues in images from different sources. Based on the specific remote sensing image types in the DSRSID, we craft two different DHCNNs architectures for the panchromatic and multispectral images. More specifically, the DHCNNs for panchromatic images are called PAN-DHCNNs, and the DHCNNs for multispectral images are called MUL-DHCNNs. The combination of PAN-DHCNNs and MUL-DHCNNs constitutes SIDHCNNs. The architectures of PAN-DHCNNs and MUL-DHCNNs are visually shown in Fig. 2.

The architectures of PAN-DHCNNs and MUL-DHCNNs are provided in Tables II and III, respectively. In Tables II and III, "filter" specifies the number of filters, the height of the field, the width of the field, and the channel number of the input data; "stride1" denotes the sliding step of the convolutional operation; "pooling" denotes the downsampling factor; "stride2" denotes the sliding step of the local pooling operation; and l stands for the feature length of the last fully connected layer (i.e., the hashing feature coding layer). Compared with MUL-DHCNNs, PAN-DHCNNs have larger fields to capture the high-resolution structures in the high-resolution panchromatic images. In contrast to PAN-DHCNNs, MUL-DHCNNs have more filters in the convolutional layers to mine the rich spectral information in the multispectral images. As a whole, PAN-DHCNNs and MUL-DHCNNs are specifically designed based on the characteristics of the remote sensing images from different sources. The architectures given in Tables II and III are just two of many candidates. This paper merely introduces a general solution for designing SIDHCNNs (i.e., the combination of PAN-DHCNNs and MUL-DHCNNs). More architecture can be explored and evaluated in the future works.

It is noted that the randomly initialized SIDHCNNs would suffer from the source shift problem which is verified in Section IV-D. To make SIDHCNNs possess the source-invariant characteristic, we learn SIDHCNNs by the following optimization method.

2) Objective Function for Optimizing SIDHCNNs: This section uses the training data set $\mathbf{D}_{\mathrm{Tr}}^{U} = \{(P_i, M_i, L_i)|i = 1, 2, \ldots, V\}$ which is introduced in Section II to formulate the objective function to learn SIDHCNNs. Let $S^U \in R^{V\times V\times 2}$ denotes the intersource pairwise similarity matrix, where $S_{i,j,1}^{U} = 1, S_{i,j,2}^{U} = 0, i = 1, 2, \ldots, V; j = 1, 2, \ldots, V$ if P_i and M_j belong to the same land-cover type and $S_{i,j,1}^{U} = 0, S_{i,j,2}^{U} = 1$ if P_i and M_j come from different land-cover types. $S^P \in R^{V\times V\times 2}$ stands for the intrasource pairwise similarity matrix on the panchromatic image data set, where $S_{i,j,1}^{P} = 1, S_{i,j,2}^{P} = 0$ if P_i and P_j belong to the same land-cover category and $S_{i,j,1}^{P} = 0, S_{i,j,2}^{P} = 1$ if P_i and P_j do not belong to the same category. Furthermore, $S^M \in R^{V\times V\times 2}$ denotes the intrasource pairwise similarity matrix on the multispectral image data set, where S^M can be generated by a similar generation process to that of S^P.

Let Λ^P and Λ^M stand for the network hyper parameters of PAN-DHCNNs and MUL-DHCNNs, respectively. $\Psi(P_i, \Lambda^P) \in R^l$ denotes the feature representation of the panchromatic image P_i by PAN-DHCNNs, and $\Upsilon(M_i, \Lambda^M) \in R^l$ stands for the feature output of the multispectral image M_i by MUL-DHCNNs. Furthermore, Λ^P and Λ^M can be learned in an end-to-end manner under a series of constraints including the IRSC, the IASC, the binary quantization constraint (BQC), and the FDC. More specifically, IRSC mainly serves to push the features of remote sensing images from different sources with the same category to be near each other and to separate the features of remote sensing images from different categories. IASC is designed to keep the feature similarity and dissimilarity of remote sensing images coherent with the category distribution in each individual source. BQC encourages the final feature representation to approach binary to accord with the goal of hashing learning, and ingeniously transforms the discrete optimization problem to the continuous one. In addition, FDC is designed to keep each element of the feature vector across the data set balanced, which means each element should have the same number of -1 and $+1$ across the data set. The contributions of these constraints will be further quantitatively discussed in Section IV-B. Moreover, the objective function for learning SIDHCNNs (i.e., Λ^P and Λ^M) can be formulated as (1), as shown at the bottom of this page where $\mathbf{F}^P \in R^{l\times V}$ with $\mathbf{F}^P_{*,i} = \Psi(P_i, \Lambda^P)$, $\mathbf{F}^M \in R^{l\times V}$ with $\mathbf{F}^M_{*,i} = \Upsilon(M_i, \Lambda^M)$ and $\mathbf{B} \in \{-1, +1\}^{l\times V}$, $p(\cdot|\mathbf{F}^P, \mathbf{F}^M)$ is the intersource-likelihood function, and $p(\cdot|\mathbf{F}^P)$ and $p(\cdot|\mathbf{F}^M)$ denote the intrasource-likelihood functions. In addition, α, β, and γ stand for the penalty weights of the constraints.

More specifically, the intersource-likelihood function is defined by the sigmoid function

$$\begin{cases} p\left(S^U_{i,j,1} = 1|\mathbf{F}^P, \mathbf{F}^M\right) = \sigma\left(\Omega^U_{i,j}\right) \\ p\left(S^U_{i,j,2} = 1|\mathbf{F}^P, \mathbf{F}^M\right) = 1 - \sigma\left(\Omega^U_{i,j}\right) \end{cases} \tag{2}$$

where $\Omega^U_{i,j} = \mathbf{F}^P_{*,i} \cdot \mathbf{F}^M_{*,j}/2$ and $\sigma(\Omega^U_{i,j}) = 1/(1 + e^{-\Omega^U_{i,j}})$.

For the panchromatic image source, the intrasource-likelihood function can be expressed by (3). In addition, the intrasource-likelihood function for the multispectral image source can be expressed by (4)

$$\begin{cases} p\left(S^P_{i,j,1} = 1\middle|\mathbf{F}^P\right) = \sigma\left(\Omega^P_{i,j}\right) \\ p\left(S^P_{i,j,2} = 1\middle|\mathbf{F}^P\right) = 1 - \sigma\left(\Omega^P_{i,j}\right) \end{cases} \tag{3}$$

$$\begin{cases} p\left(S^M_{i,j,1} = 1\middle|\mathbf{F}^M\right) = \sigma\left(\Omega^M_{i,j}\right) \\ p\left(S^M_{i,j,2} = 1\middle|\mathbf{F}^M\right) = 1 - \sigma\left(\Omega^M_{i,j}\right) \end{cases} \tag{4}$$

where $\Omega^P_{i,j} = \mathbf{F}^P_{*,i} \cdot \mathbf{F}^P_{*,j}/2$, $\sigma(\Omega^P_{i,j}) = 1/(1 + e^{-\Omega^P_{i,j}})$, and $\Omega^M_{i,j} = \mathbf{F}^M_{*,i} \cdot \mathbf{F}^M_{*,j}/2$.

If we plug the likelihood functions given in (2)–(4) into (1), the objective function for optimizing SIDHCNNs (i.e., learning the network hyper parameters Λ^P and Λ^M) can be rewritten as (5), as shown at the bottom of this page.

Despite us advocating various constraints to pursue the robust update of network hyper parameters, the final objective function given in (5) is still convex, which guarantees the efficiency of the optimization process. We give the specific learning algorithm in Section III-A3 the following section.

$$\begin{aligned} \min_{\Lambda^P, \Lambda^M, \mathbf{B}} \ E = & \overbrace{\sum_{i,j=1}^{V} \sum_{k=1}^{2} \left(-\mathbf{S}^U_{i,j,k} \log p\left(\mathbf{S}^U_{i,j,k} = 1\middle|\mathbf{F}^P, \mathbf{F}^M\right)\right)}^{\text{IRSC}} \\ & + \alpha \cdot \overbrace{\left(\sum_{i,j=1}^{V} \sum_{k=1}^{2} \left(-\mathbf{S}^P_{i,j,k} \log p\left(\mathbf{S}^P_{i,j,k} = 1\middle|\mathbf{F}^P\right)\right) + \sum_{i,j=1}^{V} \sum_{k=1}^{2} \left(-\mathbf{S}^M_{i,j,k} \log p\left(\mathbf{S}^M_{i,j,k} = 1\middle|\mathbf{F}^M\right)\right)\right)}^{\text{IASC}} \\ & + \beta \cdot \overbrace{\left(||\mathbf{F}^P - \mathbf{B}||^2_F + ||\mathbf{F}^M - \mathbf{B}||^2_F\right)}^{\text{BQC}} + \gamma \cdot \overbrace{\left(||\mathbf{F}^P \cdot \mathbf{1}||^2_F + ||\mathbf{F}^M \cdot \mathbf{1}||^2_F\right)}^{\text{FDC}} \end{aligned} \tag{1}$$

$$\begin{aligned} \min_{\Lambda^P, \Lambda^M, \mathbf{B}} \ E = & \overbrace{\sum_{i,j=1}^{V} \left(-\mathbf{S}^U_{i,j,1} \cdot \Omega^U_{i,j} + \log\left(1 + \Omega^U_{i,j}\right)\right)}^{\text{IRSC}} \\ & + \alpha \cdot \overbrace{\left(\sum_{i,j=1}^{V} \left(-\mathbf{S}^P_{i,j,1} \cdot \Omega^P_{i,j} + \log\left(1 + \Omega^P_{i,j}\right)\right) + \sum_{i,j=1}^{V} \left(-\mathbf{S}^M_{i,j,1} \cdot \Omega^M_{i,j} + \log\left(1 + \Omega^M_{i,j}\right)\right)\right)}^{\text{IASC}} \\ & + \beta \cdot \overbrace{\left(\|\mathbf{F}^P - \mathbf{B}\|^2_F + \|\mathbf{F}^M - \mathbf{B}\|^2_F\right)}^{\text{BQC}} + \gamma \cdot \overbrace{\left(\|\mathbf{F}^P \cdot \mathbf{1}\|^2_F + \|\mathbf{F}^M \cdot \mathbf{1}\|^2_F\right)}^{\text{FDC}} \end{aligned} \tag{5}$$

3) Learning SIDHCNNs: As multiple variants need to be optimized in the objective function in (5), we optimize them in an alternative learning strategy where one variant is optimized, while the others are fixed. Like [6] and [17], we adopt the mini-batch stochastic gradient descent (SGD) to learn the hyper parameters of SIDHCNNs including Λ^P and Λ^M using the following three steps iteratively until all image are processed over a fixed number of iterations. Let V_P and V_M denote the mini-batch sizes for panchromatic and multispectral images, and T stands for the number of iterations. In our implementation, both V_P and V_M are set to 128 based on the consideration of the memory space, and T is empirically set to 30 as the objective function in (5) generally converges to a stable state after dozens of iterations. We summarize the whole alternating learning procedure in Algorithm 1.

1) *Fix Λ^P and Λ^M, Optimize $\mathbf{B}$:* When Λ^P and Λ^M are fixed, the objective function in (5) can be transformed to

$$\max_{\mathbf{B}}\ \mathrm{tr}(\mathbf{B}^T(\beta(\mathbf{F}^P+\mathbf{F}^M))) = \mathrm{tr}(\mathbf{B}^T\mathbf{C}) = \sum_{i,j}\mathbf{B}_{i,j}\mathbf{C}_{i,j} \tag{6}$$

where $\mathbf{C} = \beta(\mathbf{F}^P+\mathbf{F}^M)$ and $\mathbf{B} \in \{-1,+1\}^{l\times V}$ is the binary feature matrix.

As shown in (6), it can be easily derived that the optimized $\mathbf{B}_{i,j}$ should have the same sign as $\mathbf{C}_{i,j}$. Hence, $\mathbf{B}$ can be updated to be

$$\mathbf{B} = \mathrm{sign}(\mathbf{C}) = \mathrm{sign}(\beta(\mathbf{F}^P+\mathbf{F}^M)). \tag{7}$$

2) *Fix $\mathbf{B}$ and Λ^M, Optimize Λ^P:* For each panchromatic image P_i in the sampled mini batch, we calculate its feature using $\mathbf{F}^P_{*,i} = \Psi(P_i, \Lambda^P)$ and update the feature matrix $\mathbf{F}^P \in R^{l\times V}$. With respect to the feature $\mathbf{F}^P_{*,i}$, we obtain the closed-form gradient of the objective function in (5), where the gradient can be expressed by (8). The gradient is further utilized to update Λ^P by SGD

$$\begin{aligned}\frac{\partial E}{\partial \mathbf{F}^P_{*,i}} &= \frac{1}{2}\sum_{j=1}^{V}\left(\sigma\left(\Omega^U_{i,j}\right)\mathbf{F}^M_{*,j} - \mathbf{S}^U_{i,j}\mathbf{F}^M_{*,j}\right)\\ &\quad + \alpha\cdot\sum_{j=1}^{V}\left(\sigma\left(\Omega^P_{i,j}\right)\mathbf{F}^P_{*,j} - \mathbf{S}^P_{i,j}\mathbf{F}^P_{*,j}\right)\\ &\quad + \beta\cdot\left(\mathbf{B}_{*,i} - \mathbf{F}^P_{*,i}\right) + \gamma\cdot(\mathbf{F}^P\cdot\mathbf{1}). \end{aligned} \tag{8}$$

3) *Fix $\mathbf{B}$ and Λ^P, Optimize Λ^M:* For each multispectral image M_j in the sampled mini batch, we calculate its feature using $\mathbf{F}^M_{*,j} = \Upsilon(M_j, \Lambda^M)$ and update the feature matrix $\mathbf{F}^M \in R^{l\times V}$. With respect to the feature $\mathbf{F}^M_{*,j}$, we can obtain the closed-form gradient of the objective function in (5), where the gradient is computed by (9) and further utilized to update Λ^M by SGD

$$\begin{aligned}\frac{\partial E}{\partial \mathbf{F}^M_{*,j}} &= \frac{1}{2}\sum_{i=1}^{V}\left(\sigma\left(\Omega^U_{i,j}\right)\mathbf{F}^P_{*,i} - \mathbf{S}^U_{i,j}\mathbf{F}^P_{*,i}\right)\\ &\quad + \alpha\cdot\sum_{i=1}^{V}\left(\sigma\left(\Omega^M_{i,j}\right)\mathbf{F}^M_{*,i} - \mathbf{S}^M_{i,j}\mathbf{F}^M_{*,i}\right)\\ &\quad + \beta\cdot\left(\mathbf{B}_{*,j} - \mathbf{F}^M_{*,j}\right) + \gamma\cdot(\mathbf{F}^M\cdot\mathbf{1}). \end{aligned} \tag{9}$$

Algorithm 1 Optimization Algorithm for Learning SIDHCNNs

```
Input: Dual-source training data set $\mathbf{D}^U_{Tr} = \{(P_i, M_i, L_i)|i = 1, 2, \cdots, V\}$; the pairwise similarity matrices $S^U$, $S^P$, and $S^M$; the constraint weights $\alpha$, $\beta$, and $\gamma$; the feature length of the last hashing layer $l$.
Output: Hyper-parameters $\Lambda^P$ and $\Lambda^M$ of PAN-DHCNNs and MUL-DHCNNs, and the subsidiary hashing features $\mathbf{B}$.
Initialization: Random hyper-parameters $\Lambda^P$ and $\Lambda^M$, the mini-batch size $V_P = V_M = 128$, the number of mini-batches $t_P = V/V_P$ and $t_M = V/V_M$, and the number of iterations $T = 30$.
for $t = 1, 2, \cdots, T$
  Update $\mathbf{B}$ according to Eq. (7).
  for $n = 1, 2, \cdots, t_P$
    • Randomly sample $V_P$ panchromatic images from $\mathbf{D}^U_{Tr}$ to construct a mini-batch;
    • Calculate the output $\mathbf{F}^P_{*,i} = \Psi(P_i, \Lambda^P)$ of each panchromatic image $P_i$ in the mini-batch and update the feature matrix $\mathbf{F}^P \in R^{l\times V}$;
    • Update the hyper-parameters $\Lambda^P$ based on the gradient which is calculated by Eq. (8).
  end
  for $n = 1, 2, \cdots, t_M$
    • Randomly sample $V_M$ multi-spectral images from $\mathbf{D}^U_{Tr}$ to construct a mini-batch;
    • Calculate the output $\mathbf{F}^M_{*,j} = \Upsilon(M_j, \Lambda^M)$ of each multi-spectral image $M_j$ in the mini-batch and update the feature matrix $\mathbf{F}^M \in R^{l\times V}$;
    • Update the hyper-parameters $\Lambda^M$ based on the gradient which is calculated by Eq. (9).
  end
end
```

Benefiting from the closed-form gradients, the hyper parameters Λ^P and Λ^M of SIDHCNNs can be efficiently learned. To reflect the dynamic convergence process of the advocated objective function, Section IV-D visually shows the features of images from different data sources using the hyper parameters which are optimized with different iteration numbers.

B. Cross-Source Large-Scale Remote Sensing Image Retrieval via SIDHCNNs

Cross-source remote sensing image retrieval includes two subtasks: the cross-source PAN->MUL retrieval task and the cross-source MUL->PAN retrieval task. The cross-source PAN->MUL retrieval task takes the panchromatic image as the inquiry image and aims at outputting the multispectral images with a similar content to the inquiry image. In addition, the cross-source MUL->PAN retrieval task takes the multispectral image as the inquiry image and works to output the similar panchromatic images with a similar content to the inquiry image. The cross-source PAN->MUL retrieval task is shown visually in the red rectangle in Fig. 2, and the cross-source MUL->PAN retrieval task is intuitively depicted in the blue rectangle in Fig. 2.

In the following, we take the cross-source PAN->MUL retrieval task as an example to depict the specific calculation process. Given one inquiry panchromatic image P_i, the cross-source PAN->MUL retrieval task works for retrieving the similar multispectral images from $\mathbf{D}_{\mathrm{Tr}}^{M} = \{(M_i, L_i)|i = 1, 2, \ldots, V\}$. Based on the learned SIDHCNNS, the cross-source PAN->MUL retrieval task is implemented using the following four steps. First, the feature representation $\mathbf{b} \in R^{l}$ of P_i is computed based on the hyper parameters Λ^{P} of SIDHCNNs. Second, the feature representations $\mathbf{B} \in R^{l\times V}$ of each M_i in $\mathbf{D}_{\mathrm{Tr}}^{M}$ are calculated based on the hyper parameters Λ^{M} of SIDHCNNs. Third, we calculate the hamming distances between $\mathbf{b}$ and $\mathbf{B}$. Finally, the similar multispectral images are automatically recommended by sorting the hamming distances. To facilitate understanding, we summarize the cross-source PAN->MUL retrieval task and the cross-source MUL->PAN retrieval task in Algorithm 2.

Algorithm 2 CS-LSRSIR Approach Based on SIDHCNNs

```
(1) PAN->MUL: Given one inquiry panchromatic image $P_i$ from the testing panchromatic data set $\mathbf{D}_{Te}^{P} = \{(P_i, L_i)|i = 1, 2, \cdots, Q\}$, we want to obtain the similar multi-spectral images from the searching multi-spectral data set $\mathbf{D}_{Tr}^{M} = \{(M_i, L_i)|i = 1, 2, \cdots, V\}$.
    • Calculate the feature representation $\mathbf{b} \in R^{l}$ of $P_i$ with $\mathbf{b} = \mathrm{sign}(\Psi(P_i, \Lambda^{P}))$;
    • Calculate the feature representations $\mathbf{B} \in R^{l\times V}$ of $\mathbf{D}_{Tr}^{M}$ with $\mathbf{B}_{*,i} = \mathrm{sign}(\Upsilon(M_i, \Lambda^{M}))$;
    • Calculate the hamming distances between $\mathbf{b}$ and $\mathbf{B}$;
    • Output the most similar images by sorting the hamming distances.
(2) MUL->PAN: Given one inquiry multi-spectral image $M_i$ from the testing multi-spectral data set $\mathbf{D}_{Te}^{M} = \{(M_i, L_i)|i = 1, 2, \cdots, Q\}$, we want to obtain the similar panchromatic images from the searching panchromatic data set $\mathbf{D}_{Tr}^{P} = \{(P_i, L_i)|i = 1, 2, \cdots, V\}$.
    • Calculate the feature representation $\mathbf{b} \in R^{l}$ of $M_i$ with $\mathbf{b} = \mathrm{sign}(\Upsilon(M_i, \Lambda^{M}))$;
    • Calculate the feature representations $\mathbf{B} \in R^{l\times V}$ of $\mathbf{D}_{Tr}^{P}$ with $\mathbf{B}_{*,i} = \mathrm{sign}(\Psi(P_i, \Lambda^{P}))$;
    • Calculate the hamming distances between $\mathbf{b}$ and $\mathbf{B}$;
    • Output the most similar images by sorting the hamming distances.
```

Benefiting from the well-designed constraints, including IRSC, IASC, BQC, and FDC, SIDHCNNs possess the source-invariant feature representation ability, which makes it possible to measure the content similarity of remote sensing images from different sources. As SIDHCNNs can represent remote sensing images by compact feature vectors (i.e., the dimension of the feature vector is low and each element in the feature vector is binary), the SIDHCNN-based image retrieval approach is qualified for the large-scale retrieval situation. Both of these merits make our SIDHCNNs competent at the challenging cross-source large-scale image retrieval task.

IV. Experimental Results

Section IV-A introduces the experimental setting and evaluation criteria. Section IV-B verifies the effectiveness of the advocated optimization constraints. Section IV-C analyzes the effect of the critical parameters in the objective function. Section IV-D gives the convergence analysis of the objective function. In addition, Section IV-E presents the comparison results with several competitive baselines.

A. Experimental Setting and Evaluation Criteria

1) Experimental Setting: $\mathbf{D} = \{(P_i, M_i, L_i)|i = 1, 2, \ldots, N\}$ which is specifically introduced in Section II is adopted to evaluate the cross-source remote sensing image retrieval task. In the following experiment, both of the training and testing data sets contain eight land-cover types, the training data set has 75 000 samples, and the testing data set has 5000 samples.

In the training stage, $\mathbf{D}_{\mathrm{Tr}}^{U} = \{(P_i, M_i, L_i)|i = 1, 2, \ldots, V\}$ is utilized to train the hyper parameters of SIDHCNNs. In the testing stage, $\mathbf{D}_{\mathrm{Te}}^{U} = \{(P_i, M_i, L_i)|i = 1, 2, \ldots, Q\}$ is taken as the inquiry image data set, and $\mathbf{D}_{\mathrm{Tr}}^{U} = \{(P_i, M_i, L_i)|i = 1, 2, \ldots, V\}$ is taken as the searching image data set.

All methods including our proposed method and other baselines are implemented by MATLAB and conducted on a Dell station with eight Intel Core i7-6700 processors, 32 GB of RAM, and the NVIDIA GeForce GTX 745.

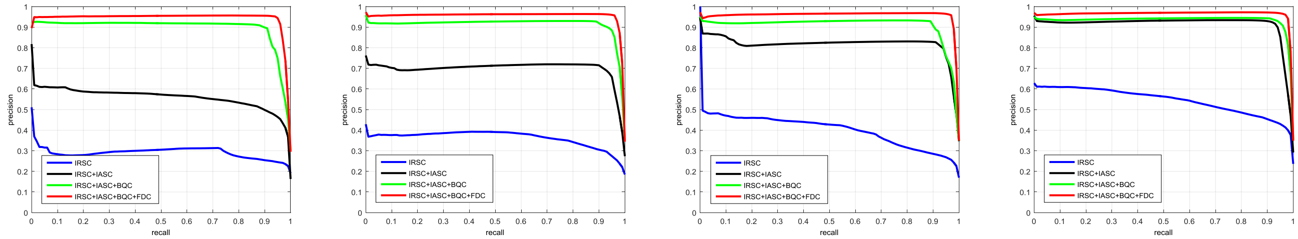
2) Evaluation Criteria: In this paper, the CS-LSRSIR performance is quantitatively evaluated by two widely adopted metrics [5], [32], including the mean average precision (MAP) and the precision-recall curve. More specifically, the MAP score can be computed by

$$\mathrm{MAP} = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{n_i}\sum_{j=1}^{n_i}\mathrm{precision}\left(R_i^{j}\right) \qquad (10)$$

where $q_i \in Q$ is the inquiry image, $|Q|$ denotes the volume of the inquiry image data set, and n_i is the number of images relevant to q_i in the searching image data set. Supposing that the relevant images are ordered as $\{r_1, r_2, \ldots, r_{n_i}\}$ across the images in the searching image data set, R_i^{j} is the set of ranked results from the first result to the r_jth result.

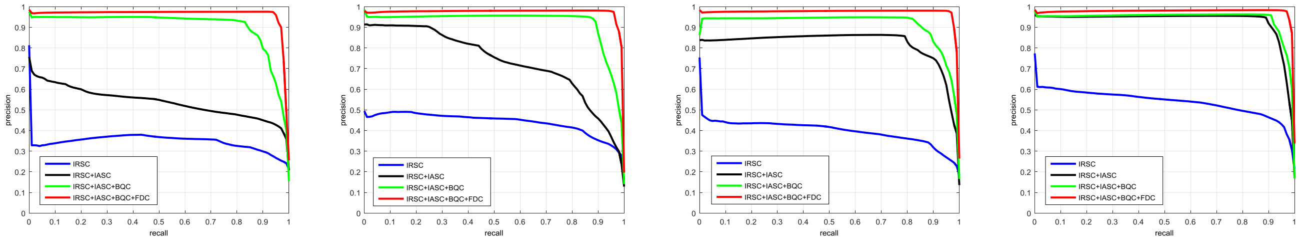
B. Validity Analysis of the Advocated Constraints

To demonstrate the validity of the advocated optimization constraints in (5), this section quantitatively evaluates the overall performance of our proposed SIDHCNNs under various optimization configurations. As this section intends to intuitively verify the effectiveness of the advocated constraints, we consider the following four major optimization configurations: only the adoption of IRSC, which is abbreviated as "IRSC"; the combination of IRSC and IASC, which is abbreviated as "IRSC + IASC"; the combination of IRSC, IASC, and BQC, which is abbreviated as "IRSC + IASC + BQC"; the combination of IRSC, IASC, BQC, and FDC, which is abbreviated as "IRSC + IASC + BQC + FDC." More specifically, "IRSC" means $\alpha = 0$,



(a) (b) (c) (d)

Fig. 3. Precision-recall curves on the cross-source PAN->MUL retrieval task under different optimization configurations and hashing feature coding lengths. (a) $l = 8$. (b) $l = 16$. (c) $l = 24$. (d) $l = 32$.



(a) (b) (c) (d)

Fig. 4. Precision-recall curves on the cross-source MUL->PAN retrieval task under different optimization configurations and hashing feature coding lengths. (a) $l = 8$. (b) $l = 16$. (c) $l = 24$. (d) $l = 32$.

TABLE IV

MAP Values of Our SIDHCNNs Under Different Optimization Configurations and Hashing Feature Coding Lengths

Cross-source retrieval tasks	Optimization configurations	$l = 8$	$l = 16$	$l = 24$	$l = 32$
PAN->MUL	IRSC	0.2903	0.3623	0.3982	0.5441
	IRSC+IASC	0.5587	0.6986	0.8164	0.9131
	IRSC+IASC+BQC	0.8967	0.9132	0.9079	0.9332
	IRSC+IASC+BQC+FDC	0.9433	0.9550	0.9577	0.9636
MUL->PAN	IRSC	0.3442	0.4400	0.3958	0.5378
	IRSC+IASC	0.5378	0.7340	0.8192	0.9298
	IRSC+IASC+BQC	0.9065	0.9225	0.9155	0.9436
	IRSC+IASC+BQC+FDC	0.9622	0.9726	0.9729	0.9760

$\beta = 0$, $\gamma = 0$ in (5), “IRSC + IASC” means $\alpha = 1$, $\beta = 0$, $\gamma = 0$, “IRSC + IASC + BQC” means $\alpha = 1$, $\beta = 1$, $\gamma = 0$, and “IRSC + IASC + BQC + FDC” means $\alpha = 1$, $\beta = 1$, $\gamma = 1$. In the following, Section IV-C will give a fine-grained analysis of α, β, γ.

Fixing the network architectures as given in Section III-A, we optimize the networks using the mentioned four optimization configurations and further evaluate the corresponding retrieval performance. Under different optimization configurations and hashing feature coding lengths, the precision-recall curves of our proposed SIDHCNNs are reported in Figs. 3 and 4. More specifically, Fig. 3 gives the evaluation results on the cross-source PAN->MUL retrieval task, and Fig. 4 reports the evaluation results on the cross-source MUL->PAN retrieval task. As depicted in Figs. 3 and 4, two cross-source retrieval tasks reflect the consistent fact that the more constraints we adopt, the better performance we achieve. In addition, the comprehensive constraints are more important for the pursuit of a short hashing feature coding length.

We also report the MAP values of our proposed SIDHCNNs under various optimization configurations and hashing feature coding lengths in Table IV. As depicted in Table IV, for two cross-source retrieval tasks, the full optimization configuration could achieve the best performance under the same hashing feature coding length. In addition, all advocated constraints could make the performance of the proposed SIDHCNNs stably grow along with the increase of the hashing feature coding length.

C. Sensitivity Analysis of the Critical Parameters

With the hashing feature coding length set to 32, this section mainly focuses on analyzing the effect of α, β, γ in the objective function in (5). As training deep networks is very time consuming, it is not possible to verify the whole parameter space. Hence, this section analyzes the sensitivity of each parameter with the other parameters fixed.

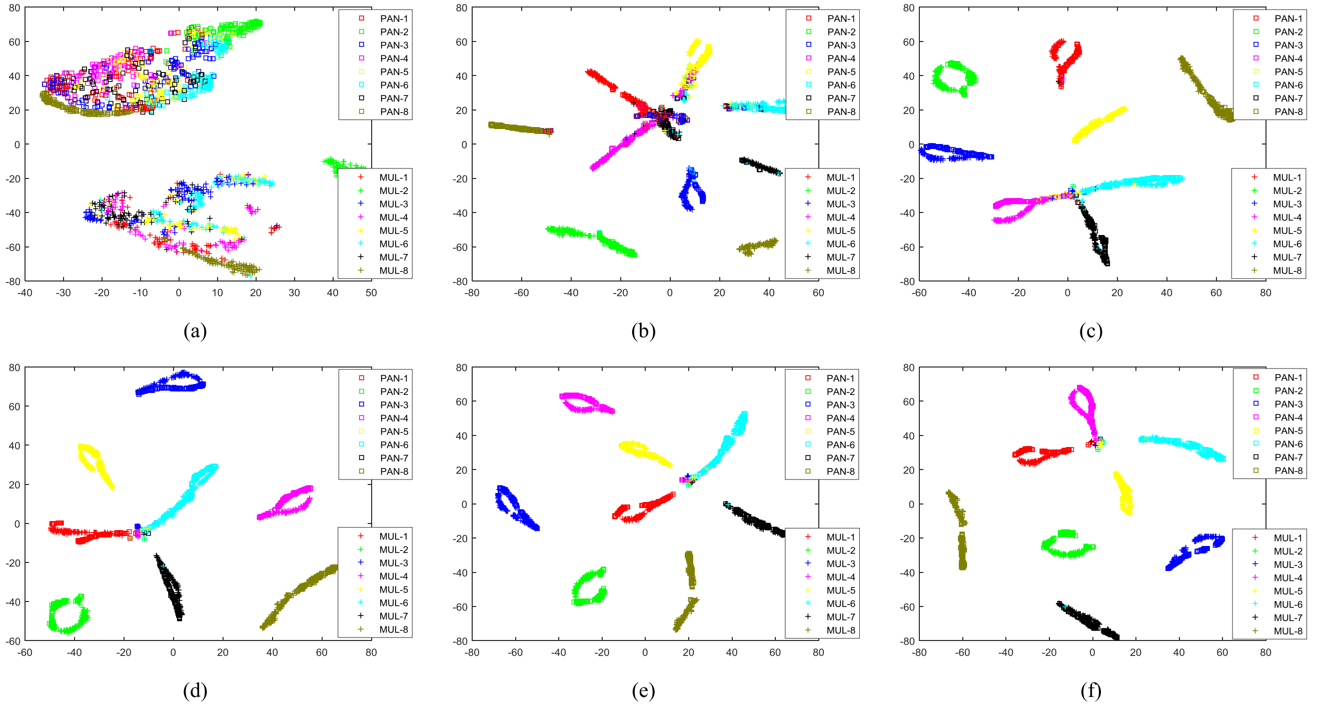
With both β and γ set to 1, Table V reports the MAP values of our SIDHCNNs by optimizing the objective function in (5) under different α. As shown in Table V, $\alpha = 0$ obviously hurts the performance of our SIDHCNNs. This is expected as $\alpha = 0$ means that the important constraint (i.e., IASC) is not adopted in the objective function. In addition, $\alpha = 1.0$

TABLE V
MAP Values of Our SIDHCNNs Under Different α

Cross-source retrieval tasks	α=0	α=0.1	α=1.0	α=10
PAN->MUL	0.8493	0.9071	0.9636	0.9134
MUL->PAN	0.8513	0.9160	0.9760	0.9291

TABLE VI
MAP Values of Our SIDHCNNs Under Different β

Cross-source retrieval tasks	β=0	β=0.1	β=1.0	β=10
PAN->MUL	0.9587	0.9643	0.9636	0.9616
MUL->PAN	0.9738	0.9789	0.9760	0.9745



(a) (b) (c)

(d) (e) (f)

Fig. 5. Feature visualization of panchromatic and multispectral images using our SIDHCNNs after our SIDHCNNs are optimized with different iteration numbers. T denotes the number of iterations. “PAN-1,” “PAN-2,” “PAN-3,”“PAN-4,” “PAN-5,” “PAN-6,” “PAN-7,” and “PAN-8” stand for panchromatic images from aquafarm, cloud, forest, high building, low building, farm land, river, and water, respectively. “MUL-1,” “MUL-2,” “MUL-3,” “MUL-4,” “MUL-5,” “MUL-6,” “MUL-7,” and “MUL-8” denote multispectral images from aquafarm, cloud, forest, high building, low building, farm land, river, and water, respectively. (a) $T = 0$. (b) $T = 10$. (c) $T = 20$. (d) $T = 30$. (e) $T = 40$. (f) $T = 50$.

makes our SIDHCNNs achieve the best performance. With both α and γ set to 1, Table VI gives the MAP values of our SIDHCNNs by optimizing the objective function in (5) under different β. As shown in Table VI, $\beta = 0.1$ and $\beta = 1.0$ make our SIDHCNNs achieve a very similar performance, and $\beta = 0.1$ makes our SIDHCNNs achieve a slightly better performance than $\beta = 1.0$. With α and β set to 1, and 0.1, Table VII reports the MAP values of our SIDHCNNs by optimizing the objective function under different γ. As shown in Table VII, our SIDHCNNs can achieve the best performance when $\gamma = 1.0$.

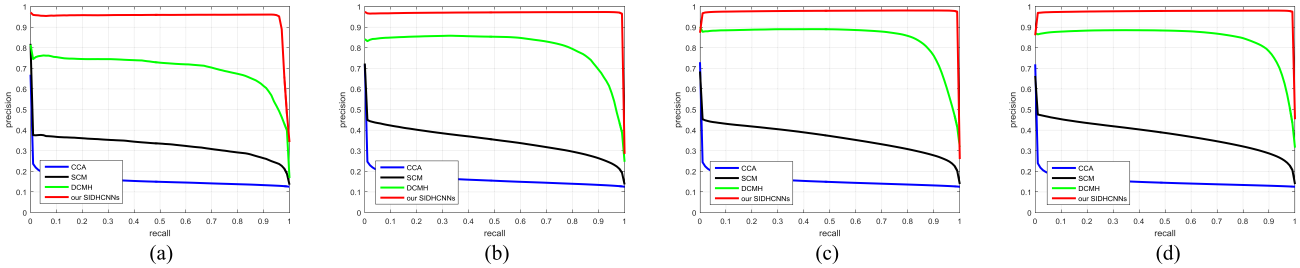
Based on the results in Tables V–VII, we can see that the performance of our SIDHCNNs is more sensitive to α compared with β and γ. Hence, researchers should pay more attention on the setting of α when they train SIDHCNNs for their tasks. To pursue the universality of our SIDHCNNs, α, β, and γ in the objective function in (5) are set to 1, 0.1, and 1 in the following experiments.

D. Convergence Analysis of the Objective Function

To show the convergence process of the objective function in (5), we update our SIDHCNNs with different iteration numbers and show the feature distributions of images using the learned SIDHCNNs to visually reflect the state of the objective function. In the visualization experiment, we consider all the land-cover classes in the training data set, and randomly select 100 dual samples for each class from the training data set where each dual sample is composed of one panchromatic image and one multispectral image. We calculate the hashing

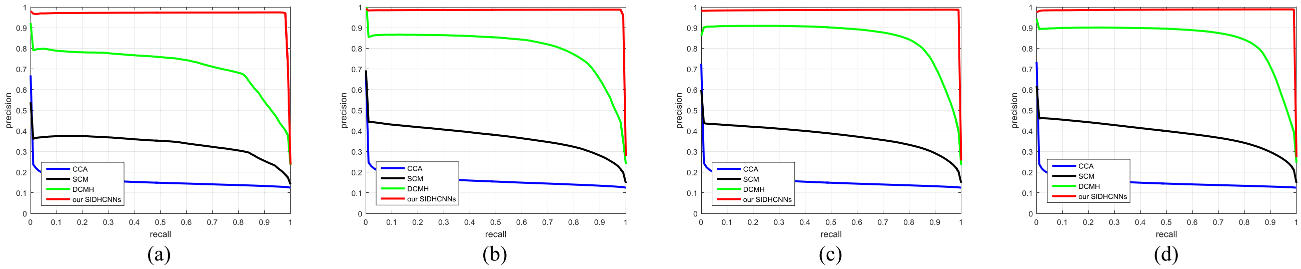
TABLE VII
MAP VALUES OF OUR SIDHCNNs UNDER DIFFERENT γ

Cross-source retrieval tasks	γ=0	γ=0.1	γ=1.0	γ=10
PAN->MUL	0.9142	0.9607	0.9643	0.9534
MUL->PAN	0.9274	0.9733	0.9789	0.9671



(a) (b) (c) (d)

Fig. 6. Precision-recall curves on the cross-source PAN->MUL retrieval task under various methods and hashing feature coding lengths. (a) $l = 8$. (b) $l = 16$. (c) $l = 24$. (d) $l = 32$.



(a) (b) (c) (d)

Fig. 7. Precision-recall curves on the cross-source MUL->PAN retrieval task under various methods and hashing feature coding lengths. (a) $l = 8$. (b) $l = 16$. (c) $l = 24$. (d) $l = 32$.

features of the panchromatic images using the PAN-DHCNNs in our SIDHCNNs, and compute the hashing features of the multispectral images using the MUL-DHCNNs in our SIDHCNNs. Furthermore, the hashing features are projected into the 2-D feature space by the t-distributed stochastic neighbor embedding method [34]. Fig. 5 intuitively shows the feature distributions in the 2-D feature space. In Fig. 5(a), $T = 0$ means that our SIDHCNNS are randomly initialized without any optimization, and the source shift problem is very significant. In addition, Fig. 5(b)–(f) shows the feature distributions of panchromatic and multispectral images using our SIDHCNNs after our SIDHCNNs are optimized with different iteration numbers. Along with the increase of iteration numbers, the source shift problem gradually minishes. After our SIDHCNNs are optimized with 20 iterations or more, the objective function converges to a stable state that the features of panchromatic images are aligned with multispectral images and the features of images are distributed on well-separated clusters for each class. To achieve a balance between the method performance and the training time, the number of iterations T in Algorithm 1 is set to 30.

E. Comparison With Several Baselines

To demonstrate the superiority of our proposed SIDHCNNs, we reimplement several baselines for comparison since there does not exist any cross-source remote sensing image retrieval work in the literature. The reimplemented baselines include two cross-modal hashing retrieval methods [15], [16] based on hand-crafted features and the most recent DCMH retrieval method [17].

Here, we first give a brief introduction of the implemented baselines. As a popular hand-crafted feature descriptor, GIST [35] is adopted and taken as the input of CCA [15] and SCM [16]. In addition, DCMH [17] does not depend on any hand-crafted features. More specifically, DCMH adopts the same network architectures as our method, but has its own optimization model. In the optimization model, DCMH ignores the IASC, but our method considers it and recommends a series of low-cost constraints, which do not obviously increase the training complexity. Under the same experimental setting of the training and testing data sets, we report the cross-source remote sensing image retrieval performance of various methods.

The precision-recall curves of various methods have been illustrated in Figs. 6 and 7. More specifically, Fig. 6 denotes the comparison results on the cross-source PAN->MUL retrieval task, and Fig. 7 stands for the comparison results on the cross-source MUL->PAN retrieval task. As depicted in Figs. 6 and 7, CCA achieves the worst performance because it works in an unsupervised way. With benefits from supervision, SCM can outperform CCA, but its performance is still

TABLE VIII
MAP VALUES UNDER DIFFERENT METHODS AND HASHING FEATURE CODING LENGTHS

Cross-source retrieval tasks	Methods	$l = 8$	$l = 16$	$l = 24$	$l = 32$
PAN->MUL	CCA [15]	0.1540	0.1593	0.1543	0.1502
	SCM [16]	0.3240	0.3472	0.3618	0.3767
	DCMH [17]	0.7009	0.8076	0.8488	0.8509
	Our SIDHCNNs	0.9473	0.9552	0.9641	0.9643
MUL->PAN	CCA [15]	0.1538	0.1594	0.1546	0.1505
	SCM [16]	0.3330	0.3671	0.3725	0.3871
	DCMH [17]	0.7142	0.8023	0.8527	0.8445
	Our SIDHCNNs	0.9668	0.9725	0.9730	0.9789

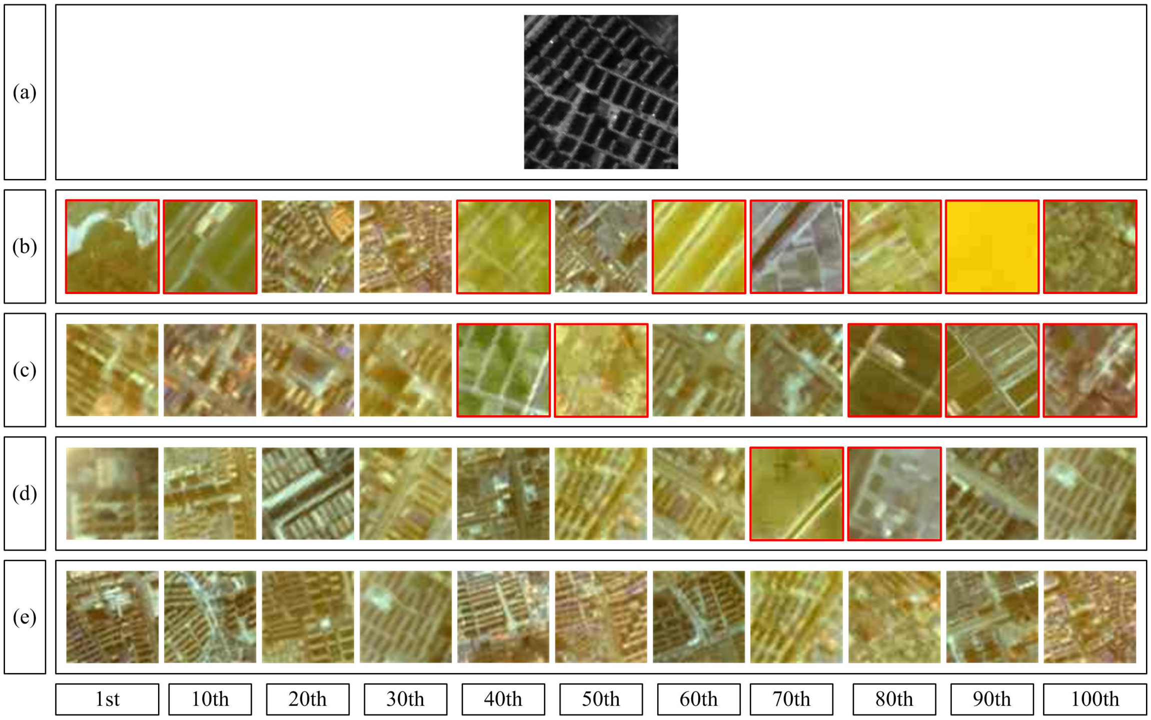


Fig. 8. Visual retrieval results on the cross-source PAN->MUL retrieval task under various methods when the hashing feature coding length is set to 32. (a) Inquiry panchromatic image from the high building category. (b) Similar multispectral images output by CCA [15]. (c) Similar multispectral images output by SCM [16]. (d) Similar multispectral images output by DCMH [17]. (e) Similar multispectral images output by our SIDHCNNs. The red rectangles stand for false retrieval results that are irrelevant to the inquiry image.

unsatisfactory due to the dependence of the hand-crafted feature. With the benefits from the adoption of deep learning, DCMH can achieve better performance than CCA and SCM. As a whole, our proposed SIDHCNNs that are optimized by the well-designed optimization constraints can achieve a notable performance improvement compared with these baselines.

We also summarize the MAP values of various methods in Table VIII. As depicted in Table VIII, our SIDHCNNs could achieve the best retrieval performance under various situations. The superiority of our SIDHCNNs is more remarkable when the hashing feature coding length is small.

To intuitively show the superiority of our SIDHCNNs, we show the visual retrieval results of various methods. With the hashing feature coding length set to 32, Fig. 8 visually shows the retrieval results of our SIDHCNNs and three baselines on the cross-source PAN->MUL retrieval task, and Fig. 9 intuitively reports the retrieval results of our SIDHCNNs and three baselines on the cross-source MUL->PAN retrieval task.

As depicted in Fig. 8, given one inquiry image, which is captured by the panchromatic sensor and belongs to the high building category, the most similar multispectral images output by various methods including the three baselines and our SIDHCNNs are visually shown. Benefiting from the usage of deep networks, our SIDHCNNs and DCMH [17] obviously perform better than CCA [15] and SCM [16]. In addition, our SIDHCNNs outperform DCMH. As a whole, we can easily

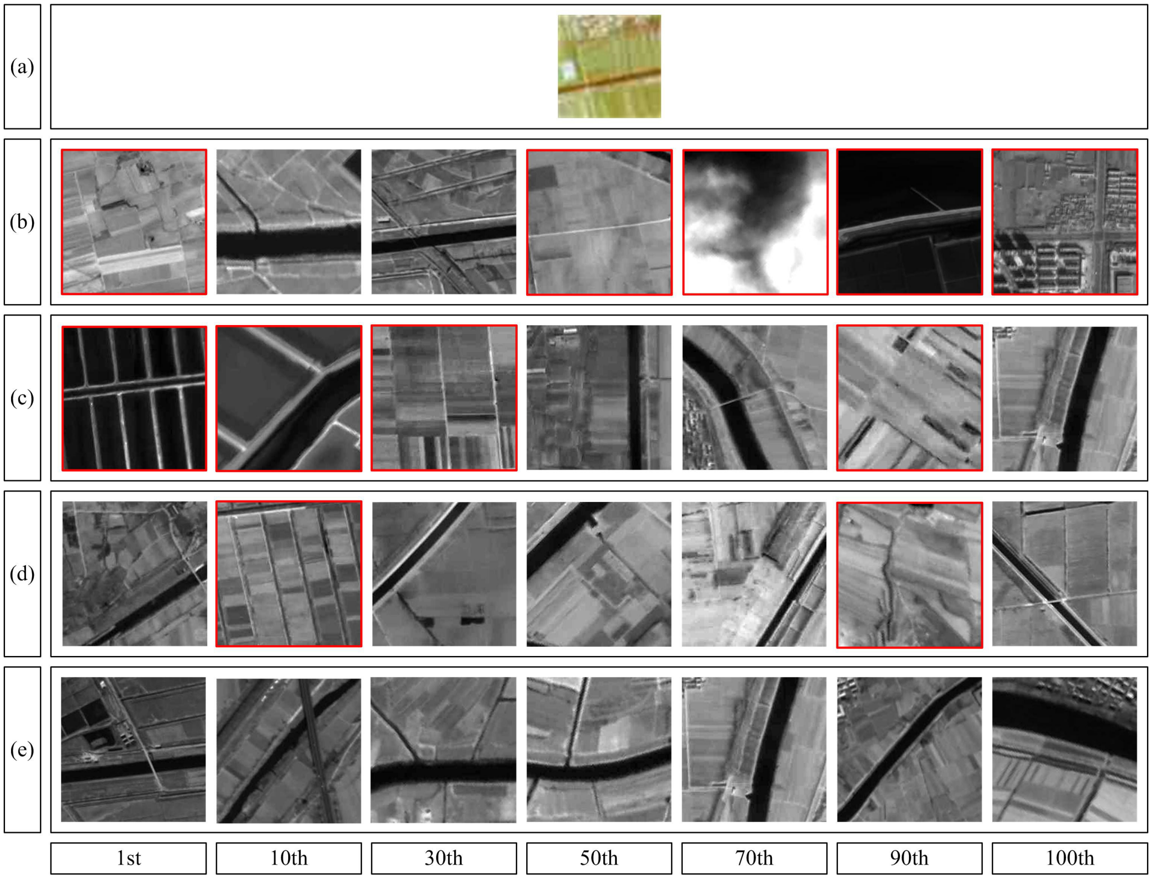


Fig. 9. Visual retrieval results on the cross-source MUL->PAN retrieval task under various methods when the hashing feature coding length is set to 32. (a) Inquiry multispectral image from the river category. (b) Similar panchromatic images output by CCA [15]. (c) Similar panchromatic images output by SCM [16]. (d) Similar panchromatic images output by DCMH [17]. (e) Similar panchromatic images output by our SIDHCNNs. The red rectangles stand for false retrieval results that are irrelevant to the inquiry image.

TABLE IX
TRAINING AND TESTING TIME OF OUR SIDHCNNS

Cross-source image retrieval tasks	Calculation steps	Time
PAN->MUL	Training	6.5 (hours)
	Testing (1): calculate the hashing feature of the inquiry image	1.5e-3 (seconds)
	Testing (2): calculate the hashing features of the searching image dataset	2.9e+1 (seconds)
	Testing (3): calculate the feature distances between the inquiry image and the searching image dataset	3.1e-4 (seconds)
	Testing (4): sort the feature distances	4.2e-3 (seconds)
MUL->PAN	Training	6.5 (hours)
	Testing (1): calculate the hashing feature of the inquiry image	3.9e-4 (seconds)
	Testing (2): calculate the hashing features of the searching image dataset	1.1e+2 (seconds)
	Testing (3): calculate the feature distances between the inquiry image and the searching image dataset	3.1e-4 (seconds)
	Testing (4): sort the feature distances	4.2e-3 (seconds)

draw the conclusion that our SIDHCNNs can achieve better performance than various potential techniques on the cross-source PAN->MUL retrieval task.

In Fig. 9, given one inquiry image, which is captured by the multispectral sensor and comes from the river category, the most similar panchromatic images output by different methods are visually shown. As depicted in Fig. 9, our SIDHCNNs and DCMH [17] can output more accurate retrieval results compared with the hand-crafted feature-based methods, including CCA [15] and SCM [16]. In the given case, the river scene may contain certain farm land elements and seems to be confused with the farm land category at first glance. Actually, the river scene has been taken as an independent category in the literature because of its special land-use type [21]–[23]. DCMH [17] cannot perfectly distinguish the difference between the river scenes and the farm land scenes, and its retrieval results include some false farm land results as depicted. Fortunately, our SIDHCNNs can robustly perceive these slight differences and show excellent retrieval results on the cross-source MUL->PAN retrieval task.

With the hashing feature coding length set to 32, we report the training and testing time of our SIDHCNNs in Table IX. Due to the usage of GPU, our SIDHCNNs can be trained from scratch in 6.5 h. As depicted in Table IX, we report the time of the testing stage in four substeps. Among the four substeps, the time for calculating the hashing features of the searching image data set is dramatically longer than the other substeps as the volume of the searching image data set is very large. As pointed out in [4], we can calculate the hashing features of the searching image data set in advance and store them in an auxiliary archive as the hashing features needs a very small storage space compared with the real-value features. Accordingly, the online testing stage will skip this substep, and could be finished in real time. This merit makes our SIDHCNNs qualified for the large-scale image retrieval case.

V. Conclusion

In this section, we first conclude our work in Section V-A; we briefly show some applications of our SIDHCNNs in Section V-B. We depict the future prospects of our work in Section V-C.

A. Conclusion

Driven by the demand of RSBD mining, this paper reveals the urgency and possibility of CS-LSRSIR. To promote the CS-LSRSIR technique, we propose a new DSRSID, which can be utilized to more multisource remote sensing image analysis techniques. To cope with CS-LSRSIR, this paper proposes SIDHCNNs, which can be optimized in an end-to-end manner under a series of well-designed constraints. It is noted that our SIDHCNNs can learn the source-invariant feature representation and reduction mapping from scratch without the requirement of any pretrained models. As a consequence, our proposed SIDHCNNs can be flexibly designed based on the special remote sensing data characteristics, and can be easily extended to more applications. To fairly demonstrate the superiority of our proposed SIDHCNNs, we reimplement three representative baselines including CCA [15] and SCM [16] and DCMH [17]. Under the same experimental environment, our SIDHCNNs can significantly outperform these baselines in terms of the quantitative and qualitative performances.

B. Real-Life Applications

In the literature, multisource remote sensing image matching [36]–[39] is a fundamental task for wide applications such as information fusion and change detection. In addition, source-invariant feature descriptors are the key module of this task. Without much expertise or effort in designing descriptors, our SIDHCNNs can automatically learn suitable source-invariant feature descriptors from data, which can be used in multisource remote sensing image matching.

Another application of SIDHCNNs would be visual navigation [40]–[42], which aims at recovering the geographical location of the imaging sensor based on scene matching between the real-time image, and the reference images. Our SIDHCNNs could extend visual navigation to a more general case that the real-time image and reference images were captured by different remote sensors.

C. Future Prospects

As a first attempt, the proposed DSRSID contains a limited number of land-cover types, and its volume is relatively small. To approach the real problem in remote sensing image big data analysis, we may enrich the data set in terms of the sample volume and the category number with the aid of crowdsourcing [43]. In the future, we may work on more challenging cases, such as cross-source retrieval between optical images and synthetic aperture radar (SAR) images.

In our future work, we will try to extend our SIDHCNNs to more cross-domain knowledge transfer problems such as the cross-domain remote sensing image scene classification task [14], and the zero-shot remote sensing image scene classification task [44].

References

[1] Y. Ma *et al.*, "Remote sensing big data computing: Challenges and opportunities," *Future Gener. Comput. Syst.*, vol. 51, pp. 47–60, Oct. 2015.

[2] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big data for remote sensing: Challenges and opportunities," *Proc. IEEE*, vol. 104, no. 11, pp. 2207–2219, Nov. 2016.

[3] G. J. Scott, M. N. Klaric, C. H. Davis, and C. R. Shyu, "Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1603–1616, May 2011.

[4] B. Demir and L. Bruzzone, "Hashing-based scalable remote sensing image search and retrieval in large archives," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 892–904, Sep. 2016.

[5] P. Li and P. Ren, "Partial randomness hashing for large-scale remote sensing image retrieval," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 464–468, Mar. 2017.

[6] Y. Li, Y. Zhang, X. Huang, H. Zhu, and J. Ma, "Large-scale remote sensing image retrieval by deep hashing neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 950–965, Feb. 2018.

[7] Y. Yang and S. Newsam, "Geographic image retrieval using local invariant features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 818–832, Feb. 2013.

[8] E. Aptoula, "Remote sensing image retrieval with global morphological texture descriptors," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 3023–3034, May 2014.

[9] B. Luo, J. F. Aujol, Y. Gousseau, and S. Ladjal, "Indexing of satellite images with different resolutions by wavelet features," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1465–1472, Aug. 2008.

[10] R. Rosu, M. Donias, L. Bombrun, S. Said, O. Regniers, and J.-P. Da Costa, "Structure tensor Riemannian statistical models for CBIR and classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 248–260, Jan. 2017.

[11] K. W. Tobin *et al.*, "Automated feature generation in large-scale geospatial libraries for content-based indexing," *Photogramm. Eng. Remote Sens.*, vol. 72, pp. 531–540, May 2006.

[12] C.-R. Shyu, M. Klaric, G. J. Scott, A. S. Barb, C. H. Davis, and K. Palaniappan, "GeoIRIS: Geospatial information retrieval and indexing system—Content mining, semantics modeling, and complex queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, Apr. 2007.

[13] D. Ye, Y. Li, C. Tao, X. Xie, and X. Wang, "Multiple feature hashing learning for large-scale remote sensing image retrieval," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 11, p. 364, 2017.

[14] E. Othman, Y. Bazi, F. Melgani, H. Alhichri, N. Alajlan, and M. Zuair, "Domain adaptation network for cross-scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4441–4456, Aug. 2017.

[15] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2916–2929, Dec. 2013.

[16] D. Zhang and W. Li, "Large-scale supervised multimodal hashing with semantic correlation maximization," in *Proc. 28th AAAI Conf. Artif. Intell.*, Quebec City, QC, Canada, 2014, pp. 2177–2183.

[17] Q. Jiang and W. Li, "Deep cross-modal hashing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jun. 2017, pp. 3270–3278.

[18] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[19] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst* Lake Tahoe, NV, USA: Harrahs and Harveys, 2012, pp. 1097–1105.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[21] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, San Jose, CA, USA, 2010, pp. 270–279.

[22] S. Basu, S. Ganguly, S. Mukhopadhyay, R. Dibiano, M. Karki, and R. Nemani, "DeepSat: A learning framework for satellite imagery," in *Proc. 23rd SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, New York, NY, USA, 2015, Art. no. 37.

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

[24] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS J. Photogramm. Remote Sens.*, vol. 98, pp. 119–132, Dec. 2014.

[25] Y. Li, C. Tao, Y. Tan, K. Shang, and J. Tian, "Unsupervised multilayer feature learning for satellite image scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 157–161, Feb. 2016.

[26] Y. Li, Y. Zhang, C. Tao, and H. Zhu, "Content-based high-resolution remote sensing image retrieval via unsupervised feature learning and collaborative affinity metric fusion," *Remote Sens.*, vol. 8, pp. 709–723, Aug. 2016.

[27] B. Zhao, Y. Zhong, and L. Zhang, "A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 73–85, Jun. 2016.

[28] W.-C. Kang, W.-J. Li, and Z.-H. Zhou, "Column sampling based discrete supervised hashing," in *Proc. 30th AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, 2016, pp. 1230–1236.

[29] R. Alshehhi, P. R. Marpu, W. L. Woon, and M. Dalla Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 139–149, Aug. 2017.

[30] W. Zhao, S. Du, Q. Wang, and W. J. Emery, "Contextually guided very-high-resolution imagery classification with semantic segments," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 48–60, Oct. 2017.

[31] Y. Li, X. Huang, and H. Liu, "Unsupervised deep feature learning for urban village detection from high-resolution remote sensing images," *Photogramm. Eng. Remote Sens.*, vol. 83, pp. 567–579, Aug. 2017.

[32] H. Liu, R. Wang, S. Shan, and X. Chen, "Deep supervised hashing for fast image retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 2064–2072.

[33] H. Zhu, M. Long, J. Wang, and Y. Cao, "Deep hashing network for efficient similarity retrieval," in *Proc. 30th AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, 2016, pp. 2415–2421.

[34] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[35] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[36] Y. Liu, F. Mo, and P. Tao, "Matching multi-source optical satellite imagery exploiting a multi-stage approach," *Remote Sens.*, vol. 9, p. 1249, Nov. 2017.

[37] M. Chen, A. Habib, H. He, Q. Zhu, and W. Zhang, "Robust feature matching method for SAR and optical images by using Gaussian-gamma-shaped bi-windows-based descriptor and geometric constraint," *Remote Sens.*, vol. 9, p. 882, Aug. 2017.

[38] 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, Mar. 2017.

[39] J. Li, C. Li, T. Yang, and Z. Lu, "Cross-domain co-occurring feature for visible-infrared image matching," *IEEE Access*, vol. 6, pp. 17681–17698, Mar. 2018.

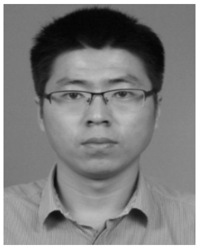
[40] C. Ivancsits and M.-F. R. Lee, "Visual navigation system for small unmanned aerial vehicles," *Sensor Rev.*, vol. 33, pp. 267–291, Jun. 2013.

[41] Q. Yu *et al.*, "Full-parameter vision navigation based on scene matching for aircrafts," *Sci. China Inf. Sci.*, vol. 57, pp. 1–10, May 2014.

[42] F. Andert and S. Krause, "Optical aircraft navigation with multi-sensor SLAM and infinite depth features," in *Proc. Int. Conf. Unmanned Aircr. Syst.*, Orlando, FL, USA, 2017, pp. 1030–1036.

[43] P. Welinder and P. Perona, "Online crowdsourcing: Rating annotators and obtaining cost-effective labels," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, San Francisco, CA, USA, Jun. 2010, pp. 25–32.

[44] A. Li, Z. Lu, L. Wang, T. Xiang, and J.-R. Wen, "Zero-shot scene classification for high spatial resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 4157–4167, Jul. 2017.



Yansheng Li received the B.S. degree from the School of Mathematics and Statistics, Shandong University, Weihai, China, in 2010, and the Ph.D. degree from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2015.

Since 2015, he has been an Assistant Professor with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan. He is currently a Visiting Assistant Professor with the Department of Computer Science, Johns Hopkins University, Baltimore, MD, USA, until 2018. He has authored over 20 peer-reviewed articles in international journals from multiple domains such as remote sensing and computer vision. His research interests include computer vision, machine learning, deep learning, and their applications in remote sensing.

Dr. Li has been serving as a Reviewer for multiple international journals including the IEEE TRANSACTIONS ON GEOSCIENCES AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the *Photogrammetric Engineering and Remote Sensing*, the *Remote Sensing*, *International Journal of Digital Earth*, and the *ISPRS International Journal of Geo-Information*. He is also a communication evaluation expert for the National Natural Science Foundation of China.

Yongjun Zhang received the B.S., M.S., and Ph.D. degrees from Wuhan University (WHU), Wuhan, China, in 1997, 2000, and 2002, respectively.

He is currently a Professor of photogrammetry and remote sensing with the School of Remote Sensing and Information Engineering, WHU. His research interests include space, aerial, and low-attitude photogrammetry; image matching, combined bundle adjustment with multisource data sets, and 3-D city reconstruction.

Xin Huang (M'13–SM'14) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009.

He was with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, where he is currently a Luojia Distinguished Professor, and teaches remote sensing, photogrammetry, and image interpretation. He is the Founder and Director of the Institute of Remote Sensing Information Processing, School of Remote Sensing and Information Engineering, Wuhan University. He has authored over 100 peer-reviewed articles (SCI papers) in the international journals. His research interests include remote sensing image processing methods and applications.

Prof. Huang was a recipient of the Boeing Award for the Best Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing in 2010, and the National Excellent Doctoral Dissertation Award of China in 2012. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society (GRSS) as the Best Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was the winner of the IEEE GRSS 2014 Data Fusion Contest. He was the lead Guest Editor of the special issue on information extraction from high-spatial-resolution optical remotely sensed imagery for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (vol. 8, no. 5, May 2015), and the special issue on *Sparsity-Driven High Dimensional Remote Sensing Image Processing and Analysis* for the *Journal of Applied Remote Sensing* (vol. 10, no. 4, Oct. 2016). Since 2016, he serves as an Associate Editor of the *Photogrammetric Engineering and Remote Sensing*. Since 2014, he has been serving as an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was supported by the Youth Talent Support Program of China in 2017, the China National Science Fund for Excellent Young Scholars in 2015, and the New Century Excellent Talents in University from the Ministry of Education of China in 2011.

Jiayi Ma received the B.S. degree from the Department of Mathematics and the Ph.D. degree from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. From 2014 to 2015, he was a Post-Doctoral Researcher with the Electronic Information School, Wuhan University, Wuhan, where he is currently an Associate Professor. He has authored or co-authored over 70 scientific articles. His research interests include computer vision, machine learning, and pattern recognition.