Remote Sensing Image Retrieval by Deep Attention Hashing With Distance-Adaptive Ranking

Yichao Zhang, Xiangtao Zheng, *Senior Member, IEEE*, and Xiaoqiang Lu, *Senior Member, IEEE*

***Abstract*—With the joint advancement of numerous related fields of remote sensing, the amount of remote sensing data is growing exponentially. As an essential remote sensing Big Data management technique, content-based remote sensing image retrieval has attracted more and more attention. A novel deep attention hashing with distance-adaptive ranking (DAH) is proposed for remote sensing image retrieval in this article. First, a channel-spatial joint attention mechanism is employed for feature extraction of remote sensing images to make the proposed DAH method focus more on the critical details of the remote sensing images and suppress irrelevant regional responses. Second, a novel balanced pairwise weighted loss function is proposed to enable discrete hash codes to participate in neural network training, which contains pairwise weighted similarity loss, classification loss, and quantization loss. The pairwise weighted similarity loss is designed to decrease the impact of the imbalance of positive and negative sample pairs. The classification loss and quantization loss are added to the loss function to decrease background interference and information loss during the quantization phase, respectively. Finally, a distance-adaptive ranking strategy with category-weighted Hamming distance is presented in the retrieval phase to utilize the category probability information fully. Experiments on benchmark datasets compared with state-of-the-art methods demonstrate the effectiveness of the proposed DAH method.**

***Index Terms*—Channel-spatial joint attention, deep hashing, distance-adaptive ranking, remote sensing image retrieval.**

I. Introduction

With the rapid development of technologies related to Earth observation, there has been a meteoric rise in the volume of remote sensing data [1]. To manage large amounts of remote sensing image data, content-based remote sensing image retrieval (CBRSIR) have attracted considerable attention [2]. The main purpose of CBRSIR is to seek the desiring remote sensing images from the massive remote sensing images, which is convenient for users to collect and manage specific remote sensing images [3], [4].

Recently, hashing methods are widely used in CBRSIR tasks due to low storage requirements and efficient retrieval computation [5]. In general, hashing method for remote sensing image retrieval consists of three main parts: feature extraction, hash code learning, and retrieval ranking.

1) Remote sensing image feature extraction is to extract discriminative features of remote sensing images with semantic information. In the initial periods of remote sensing image feature extraction, the hand-crafted features were the primary descriptions built by texture, color, or shape of remote sensing images. Li et al. [6] proposed an improved context-sensitive Bayesian network for remote sensing image retrieval, which uses the surrounding features in addition to its own relevant features to explore the semantic information of the image. Sebai et al. [7] proposed a dual-tree complex wavelet transform (DT-CWT) for improving remote sensing image retrieval's performance, which processes different types of features at the same time. Recently, feature extraction has progressed from hand-crafted feature extraction to deep feature extraction, which has improved discrimination [8], [9]. Liu et al. [10] proposed a similarity-based Siamese convolutional neural networks (SBS-CNN) for remote sensing image retrieval, which implements unsupervised training by deep transfer learning. Roy et al. [11] proposed a deep metric-learning-based hashing for remote sensing image retrieval, which utilized a pretrained convolutional neural networks (CNNs) for intermediate feature extraction. Although remote sensing image feature extraction makes rapid progress, it is still susceptible to background information interference, and the extracted deep features are difficult to focus on discriminative visual information. This problem was not addressed very effectively in most methods.
2) Hash code learning maps remote sensing images into binary hash codes for efficient retrieval and convenient storage. Pairwise hash code learning has been demonstrated to be an efficient method for learning hash codes [12], [13]. Li et al. [12] proposed a pairwise hash code learning framework for image retrieval, which uses image pairs as input and learns hash code by pairwise label and loss function. On the basis of [12], Li et al. [13] introduced discrete hash code learning into the pairwise hashing framework, which enables the discrete hash codes to

Manuscript received 2 March 2023; revised 13 April 2023; accepted 21 April 2023. Date of publication 28 April 2023; date of current version 15 May 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62271484, in part by the National Science Fund for Distinguished Young Scholars under Grant 61925112, and in part by the Key Research and Development Program of Shaanxi under Grant 2023-YBGY-225. *(Corresponding author: Xiangtao Zheng.)*

Yichao Zhang is with the Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: yczhang0819@gmail.com).

Xiangtao Zheng and Xiaoqiang Lu are with the College of Physics and Information Engineering, Fuzhou University, Fuzhou 350002, China, and also with the Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, Shaanxi 710119, China (e-mail: xiangtaoz@gmail.com; luxq666666@gmail.com).

Digital Object Identifier 10.1109/JSTARS.2023.3271303

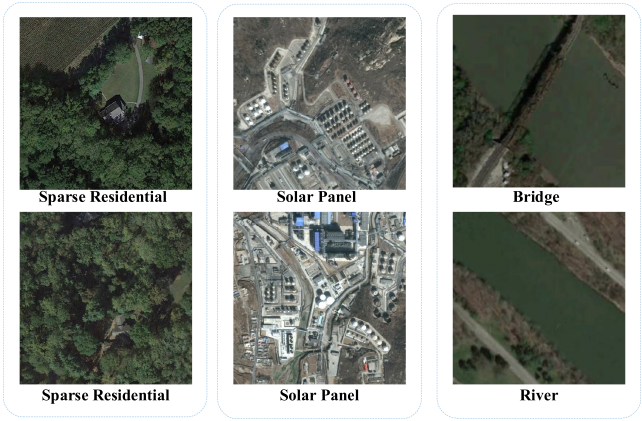

Fig. 1. Problems faced by content-based remote sensing image retrieval. (a) Main objects representing the semantic information of the images are not prominent. (b) Locations of the objects are scattered. (c) Differences between certain categories are small.

participate in training directly for reducing quantization loss. Li et al. [14] proposed a supervised hashing network for remote sensing image retrieval, which introduces pairwise similarity measurement and binary quantization loss in hash code learning of remote sensing images. However, because of the randomness of sampling, the number of positive and negative sample pairs is prone to imbalance throughout the training process. This imbalance that impacts the precision of this class of methods is always ignored.

3) Retrieval ranking stage uses appropriate distance measurements to access the similarity between the query image and database, such as Hamming distance [15]. Bao and Guo [16] implemented comparative experiments of eight different similarity measurements for remote sensing image retrieval. The hashing method retrieves the image with the lowest Hamming distance relative to the query image, which takes linear time [17]. The differences between features are not only reflected in the differences in the values of the feature vectors, but also in the overall category differences. However, in the retrieval ranking phase of most methods, the learned category information is not fully utilized.

Moreover, for developing a deep hashing network that fulfills the characteristics of remote sensing images, it is required to address the issues that now plague content-based remote sensing image retrieval.

1) As shown in Fig. 1(a), the spatial proportion of the object is small in some remote sensing images, and the object features that represent the semantic information of remote sensing images are not prominent and easily influenced by irrelevant background information.
2) The positions of objects in some remote sensing images are scattered, and it is difficult to concentrate on information from several local regions simultaneously during the feature extraction process, which weakens the semantic expression of core features and hence affects retrieval accuracy. This situation can be observed in Fig. 1(b).
3) In the examples shown in Fig. 1(c), the remote sensing image of "bridge" contains a vast region of the river, which is readily affected by the remote sensing image of "river" during retrieval. This interference also impairs retrieval performance.

To address the problems and fit the characteristics of remote sensing images mentioned previously, a novel deep attention hashing with distance-adaptive ranking (DAH) is proposed for remote sensing image retrieval. Attention mechanism can effectively guide feature extraction to pay more attention to valuable information [18]. To suppress unnecessary regional responses, a channel-spatial joint attention mechanism is employed for representative feature extraction from the complex remote sensing images, which makes the proposed DAH focus more on the important features of the remote sensing images and suppress unnecessary regional responses. A balanced pairwise weighted loss function is proposed for high-quality hash code learning. For the problem of imbalance in the number of positive and negative sample pairs in pairwise hash code learning, a pairwise weighted similarity loss is employed in the loss function. The classification loss and quantization loss are added to the loss function to decrease background interference and information loss in the quantization process. Finally, a distance-adaptive ranking strategy is presented to utilize the category probability information with Hamming distance, which further improves the performance of the proposed DAH method in the retrieval phase. Experiments on benchmark datasets compared with state-of-the-art methods demonstrate the effectiveness of the proposed method.

The remaining sections of this article are organized as follows. Section II briefly reviews and summarizes related work. The proposed DAH method is outlined in full in Section III. Section IV shows the performance of the proposed DAH method on two benchmark datasets and the analysis of the experimental results. Finally, Section V concludes this article.

II. Related Work

The objective of remote sensing image retrieval is to find the needed remote sensing images amid huge quantities of remote sensing image data, which are grouped as real-valued and hashing method. The difference between processes of the typical real-valued and hashing remote sensing image retrieval method is shown in Fig. 2. The main difference is the feature used to calculate the similarity between images. The real-valued method uses real-valued features for retrieval. The hashing method incorporates hash code learning into the workflow and employs the learned hash codes for retrieval [19]. The benefit of the hashing method is that the hash code requires less space for storage, and it is more efficient to evaluate image similarity using hash codes. This section will discuss the related work in the aforementioned two categories: real-valued and hashing method.

A. Real-Valued Method for Remote Sensing Image Retrieval

As mentioned before, real-valued method maps images to real-valued features and then uses certain similarity measurement to perform retrieval. Depending on whether or not the

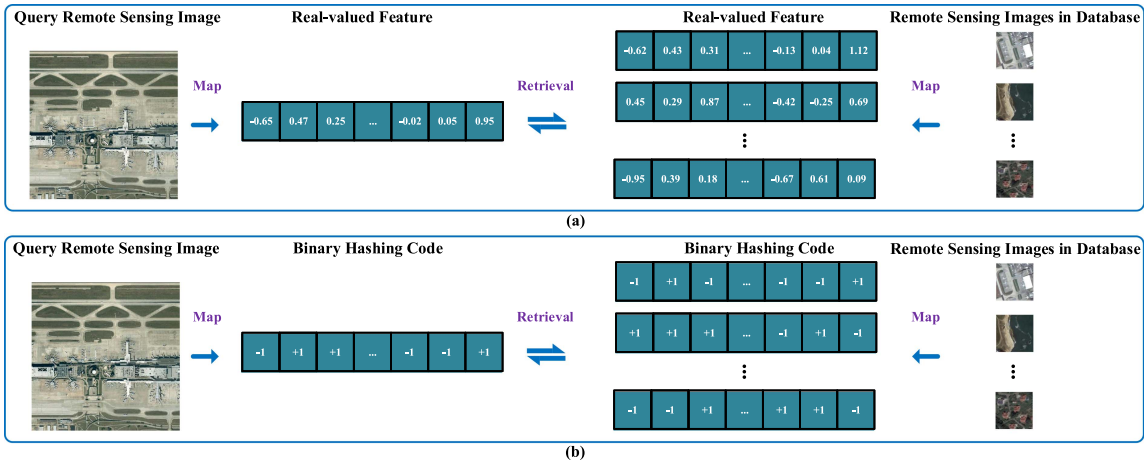

Fig. 2. Difference between processes of the typical real-valued and hashing method. (a) Real-valued method for remote sensing image retrieval. (b) Hashing method for remote sensing image retrieval.

feature extraction method has a deep network structure, real-valued methods can be further divided into two categories: real-valued method without deep learning and real-valued method with deep learning.

1) Real-Valued Method Without Deep Learning: As hand-crafted features widely used in computer vision, texture, color, and shape features are also used for low-level feature representation of remote sensing images. Wang et al. [20] proposed an image scene semantic matching model for remote sensing image retrieval, which utilizes multiple low-level information from remote sensing images. Ma et al. [21] designed a novel ensemble neural networks (ENNs) with low-level feature extraction and subfeature selection, which takes advantage of artificial neural networks to combine low-level features better. Wang et al. [22] proposed a graph-based method with a three-layer structure, which combines the advantages of query expansion and the integration of holistic and local information. Sukhia et al. [23] proposed a remote sensing image retrieval method with a novel local ternary pattern descriptor, which generates upper and lower texture patches for the histogram representation. Byju et al. [24] proposed a content-based remote sensing image retrieval system utilizing a unique coarse-to-fine retrieval technique, which is unsupervised and does not need complete image decoding.

2) Real-Valued Method With Deep Learning: Deep learning models represented by CNNs are able to capture more fundamental features of images [25]. There have been many methods to represent the visual content of remote sensing image images with high-level features output from deep learning models. Imbriaco et al. [26] designed a global descriptor for remote sensing image retrieval, which combines local convolutional features via the vector of locally aggregated descriptors (VLAD). Chaudhuri et al. [27] proposed a Siamese graph convolution network (SGCN) to better retrieve remote sensing image, which learns features by measuring the pairwise similarity of graphs. Fan et al. [28] proposed a novel distribution consistency loss for remote sensing image retrieval method based on deep metric learning, which leads the network to extract more meaningful information quickly. Liu et al. [29] proposed an eagle-eyed multitask CNN for center-metric learning, similarity distribution learning and aerial scene classification, which has ability to distinguish subtle differences between remote sensing images.

B. Hashing Method for Remote Sensing Image Retrieval

Hashing method maps remote sensing images to hash code and accesses the similarity between hash codes to perform retrieval [30]. As the review of real-valued methods, hashing methods are also divided into two categories: hashing method without deep learning and hashing method with deep learning.

1) Hashing Method Without Deep Learning: In the initial periods of hashing methods for remote sensing image retrieval, hand-crafted features were widely utilized to extract features. Demir and Bruzzone [31] leveraged two kernel-based hashing methods on remote sensing image retrieval, which uses labeled and unlabeled images for encoding hash code in kernel space, respectively. Li and Ren [32] proposed a partial randomness hashing (PRH) for remote sensing image retrieval, which makes the learning of hash functions more efficient due to the random parameters. Fernandez-Beltran et al. [33] proposed a novel probabilistic latent Semantic hashing (pLSH) for hash function construction of remote sensing images, which generates hash code with hidden semantic information by probabilistic topic model. Ye et al. [34] proposed a hashing retrieval framework for remote sensing images, which maps multiple feature descriptors into compact binary hash codes. Reato et al. [35] proposed a primitive cluster sensitive hashing for unsupervised remote sensing image retrieval, which employs multiple hash codes in hash function construction and matching phase.

2) Hashing Method With Deep Learning: Due to the powerful feature extraction capabilities demonstrated by deep learning, more and more work applies it to hashing methods for remote sensing image retrieval. Song et al. [36] proposed a deep hashing CNN (DHCNN), which employs the CNN to extract high-level features for remote sensing image retrieval. Han et al. [37] proposed a cohesion intensive deep hashing (CIDH) method with a residual hash net, which makes hash codes more similar to each other within the same class. Liu et al. [38] proposed a deep hashing model with generative

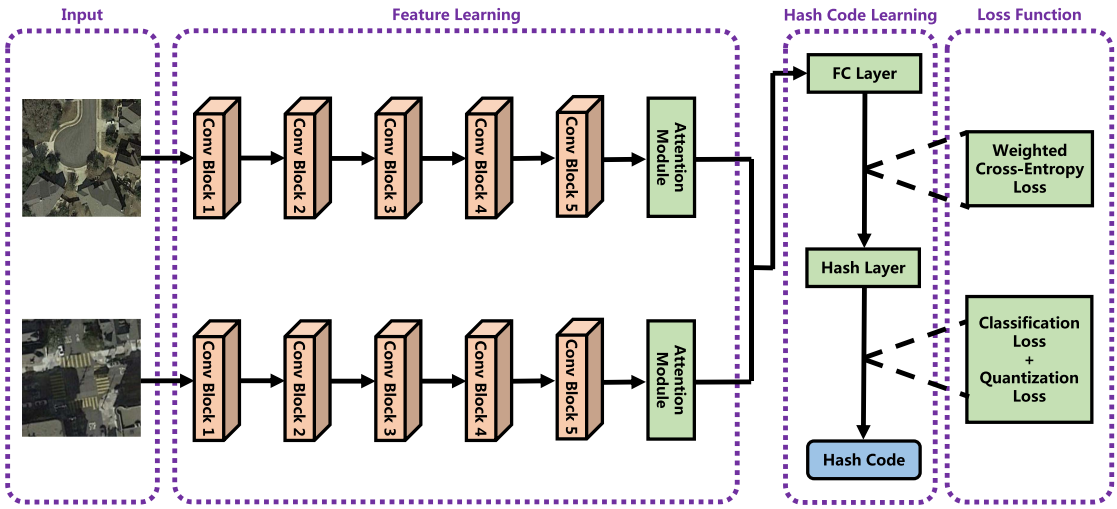

Fig. 3. Framework of the proposed DAH method. First, deep features are extracted from pairwise randomly sampled remote sensing images via five convolution blocks and channel-spatial joint attention. Second, after reducing the dimension of the deep feature with the full-connected layer, the obtained continuous features are quantized to the hash codes with the values of +1 and −1 via the hash layer. Finally, the loss function proposed in this article is employed to optimize the discrete hash code.

adversarial networks (GAN), which designed a unique loss function for the generator. Li et al. [39] proposed a quantized deep learning to hash (QDLH) model, which reduces the computational burden of deep hash code learning. Tang et al. [40] proposed a novel deep hashing network, which implements the hash code learning of remote sensing in a few-shot learning way.

III. Proposed Method

The proposed DAH method is divided into the following three parts: 1) feature extraction with channel-spatial joint attention, 2) hash code learning with balanced pairwise weighted loss, and 3) distance-adaptive ranking. The overall framework of the proposed DAH method is shown in Fig. 3.

A. Channel-Spatial Joint Attention

The proposed DAH method employs a channel-spatial joint attention mechanism to extract representative features from the complicated remote sensing image information [41]. Both channel attention and spatial attention are leveraged in a united framework. The channel attention module produces the channel attention map by compressing the spatial dimension of the input features, which can explore the relative importance among different channels. The spatial attention module produces the spatial attention map by compressing the channel dimension of the input features, which can explore the relative importance among different spatial positions of the images [42]. Experiments designed in [41] demonstrate that the connection with channel attention first and then spatial attention works better.

Given a convolutional feature $g \in \mathbb{R}^{H\times W\times C}$ as input, the maximum pooling and average pooling are performed on each channel separately for the input feature block to obtain the maximum pooling features $g_{\text{ch}}^{\text{Max}}$ and average pooling features $g_{\text{ch}}^{\text{Avg}}$ with a constant number of channels. The summed features of $g_{\text{ch}}^{\text{Max}}$ and $g_{\text{ch}}^{\text{Avg}}$ are scaled using a multilayer perceptron, and the activation output is performed using the Sigmoid function to obtain the channel attention feature map A_{ch}. The channel attention map A_{ch} is computed as

$$\begin{aligned} g_{\text{ch}}^{\text{Max}} &= \text{Max}^{\text{ch}}(g) \\ g_{\text{ch}}^{\text{Avg}} &= \text{Max}^{\text{ch}}(g) \\ A_{\text{ch}} &= \sigma(W_2(W_1(g_{\text{ch}}^{\text{Max}} + g_{\text{ch}}^{\text{Avg}})) \end{aligned} \tag{1}$$

where $\text{Max}^c h$ signifies the maximum pooling operation on each channel, $\text{Avg}^c h$ denotes the average pooling operation on each channel, W_1 and W_2 signify the weight parameters in the multilayer perceptron, and $\sigma(\cdot)$ means Sigmoid function.

Then, the channel attention map A_{ch} is done the dot product with the input features g to obtain the intermediate features g_{ch} as

$$g_{\text{ch}} = A_{\text{ch}} \cdot g. \tag{2}$$

The spatial attention module performs maximum pooling and average pooling for all channel values at each spatial position of the intermediate feature g_{ch} to obtain maximum pooling feature $g_{\text{sp}}^{\text{Max}}$ and average pooling feature $g_{\text{sp}}^{\text{Avg}}$. Then, the maximum pooling feature $g_{\text{sp}}^{\text{Max}}$ and average pooling feature $g_{\text{sp}}^{\text{Avg}}$ are concatenated in the channel dimension. A convolutional layer with a convolutional kernel size of 1 is used for the dimensionality reduction of the concatenated feature. The reduced-dimensional features are calculated by the Sigmoid function to obtain the spatial attention map A_{sp}. The calculation of A_{sp} is

$$\begin{aligned} g_{\text{sp}}^{\text{Max}} &= \text{Max}^{\text{sp}}(g_{\text{ch}}) \\ g_{\text{sp}}^{\text{Avg}} &= \text{Avg}^{\text{sp}}(g_{\text{ch}}) \\ A_{\text{sp}} &= \sigma(\text{Conv}(\text{Cat}(g_{\text{sp}}^{\text{Max}}, g_{\text{sp}}^{\text{Avg}}))) \end{aligned} \tag{3}$$

where Max^{sp} indicates the maximum pooling at each spatial location, Avg^{sp} means the mean pooling operation at each spatial location, $\text{Conv}(\cdot)$ denotes the convolution operation, and $\text{Cat}(\cdot)$ is the concatenate operation.

Finally, the spatial attention map A_{sp} is done with the dot product with the intermediate feature g_{ch} to obtain the output

features g_{sp}:

$$g_{\text{sp}} = A_{\text{sp}} \cdot g_{\text{ch}}. \tag{4}$$

B. Hash Code Learning With Balanced Pairwise Weighted Loss

After acquiring the feature g_{sp} from the attention block, the hash feature $h \in \mathbb{R}^K$ can be obtained by fully connected layers, where K denotes the hash code length. The hash feature $h \in \mathbb{R}^K$ can be expressed as follows:

$$h = \tanh\left(F_c\left(g_{\text{sp}}\right)\right) \tag{5}$$

where F_c represents the fully connected layer, the $\tanh$ activation function maps the vector value between -1 and 1.

Given a pair of remote sensing images x_i and x_j, s_{ij} represents the similarity label between x_i and x_j. If the images x_i and x_j belong to the same category, $s_{ij} = 1$; otherwise, $s_{ij} = 0$. If two remote sensing images are from the same category, the similarity between hash features h_i and h_j should be raised consistently during the network training process, and vice versa. The inner product can be used to measure the similarity between two different hash features, which is written as $I_{ij} = \langle h_i, h_j \rangle$. The relationship between similarity labels and hash feature similarity can be defined using the logistic regression model in binary classification as follows:

$$\begin{cases} P(s_{ij} = 1 \mid I_{ij}) = \sigma(I_{ij}) \\ P(s_{ij} = 0 \mid I_{ij}) = 1 - \sigma(I_{ij}) \end{cases} \tag{6}$$

where $\sigma(I_{ij}) = \frac{1}{1+e^{-I_{ij}}}$. The more similar the hash features, the more likely they belong to the same category, and vice versa. Equation (6) can be transformed as the cross-entropy loss with maximum likelihood estimation as follows:

$$\begin{aligned} L_{\text{cross}} &= -\sum_{s_{ij} \in \Omega} \left[s_{ij} \log\left(\sigma\left(I_{ij}\right)\right) + \left(1 - s_{ij}\right) \log\left(1 - \sigma\left(I_{ij}\right)\right) \right] \\ &= \sum_{s_{ij} \in \Omega} \left(\log\left(1 + e^{I_{ij}}\right) - s_{ij} I_{ij} \right) \end{aligned} \tag{7}$$

where Ω represents the set of similarity labels, and the image pairs are randomly sampled. However, the similar and dissimilar pairs may be imbalanced, which will cause the neglect of information from quantitatively inferior sample pairs. Inspired by the motivation of Focal Loss [43], a weight coefficient w_{ij} is proposed to reduce this imbalance.

$$w_{ij} = \begin{cases} |\Omega| / |\Omega_1|, & s_{ij} = 1 \\ |\Omega| / |\Omega_2|, & s_{ij} = 0 \end{cases} \tag{8}$$

where Ω_1 represents the set of labels of similar samples, Ω_2 represents the set of labels of dissimilar samples, and $|\cdot|$ represents the number of elements in the set.

Equation (7) can be transformed as

$$L_{\text{w-cross}} = \sum_{s_{ij} \in \Omega} w_{ij} \left(\log\left(1 + e^{I_{ij}}\right) - s_{ij} I_{ij} \right). \tag{9}$$

The weighted cross-entropy loss is utilized to optimize the similarity distance between hash features. The smaller the distance of the hash codes between similar images, the greater the distance of the hash codes between dissimilar images.

The purpose of the hash layer is to make discrete hash codes participate in the training process, directly train the classification loss for the hash code, and further reduce the information loss between the hash feature and the hash code. The hash features of all remote sensing images in the training set are quantified into hash codes using the Sign function

$$B = \operatorname{sign}(H) \tag{10}$$

where B denotes the hash code, and $\operatorname{sign}(\cdot)$ represents the Sign function that quantizes continuous features into a hash code with values of +1 or -1. The quantization loss is constructed to make the continuous features fit the hash code and reduce the information loss in the quantization process.

$$L_q = \frac{1}{N} \sum_{i=1}^{N} \left\| h_i - b_i \right\|_3^3 \tag{11}$$

where N represents the number of images in the training set.

The classification loss can be used to judge whether the generated hash code can better distinguish different types of data, which is constructed using the linear Softmax classification function as

$$L_{\text{class}} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{j=1}^{C} 1\left\{y^n = j\right\} \log \frac{e^{\theta_j^T b_n}}{\sum_{i=1}^{C} e^{\theta_i^T b_n}} \tag{12}$$

where θ denotes the parameter in the linear classifier and C denotes the number of categories in the remote sensing image. When the values of y^n and j are equal, the value of $1\{y^n = j\}$ is 1, otherwise it is 0.

The overall loss function is

$$L_{\text{total}} = L_{w-\text{cross}} + \alpha L_q + \beta L_{\text{class}}. \tag{13}$$

C. Distance-Adaptive Ranking

A query set of remote sensing images is denoted as $Q = \{q_j\}_{j=1}^{N_q}$, where N_q represents the number of query images and the corresponding hash code is denoted as $B_q = \{b_j^q\}_{j=1}^{N_q}$. In the retrieval stage, the original Hamming distance is generally used to measure the similarity between the images from the database and the query image, and the retrieved images are ranked according to the corresponding Hamming distance. The original Hamming distance formula between hash codes is

$$D_h\left(b_j^q\right) = \frac{1}{2}\left(K - \left\langle b_j^q, b_i \right\rangle\right) \tag{14}$$

where K denotes the hash code length.

To make better use of the model's learned category information and to reduce the distance between images with the same potential category, a distance-adaptive ranking strategy is proposed in the retrieval stage. In the distance-adaptive ranking strategy, the Hamming distance is replaced with a novel category-weighted distance. The calculation of the category-weighted distance is shown in Fig. 4. After the prediction of the linear Softmax classifier, the most probable category C_q of the query image q_j is obtained. The predicted probability of the image x_i from the retrieval database on the most probable category C_q of the query image is denoted as p_i. The category

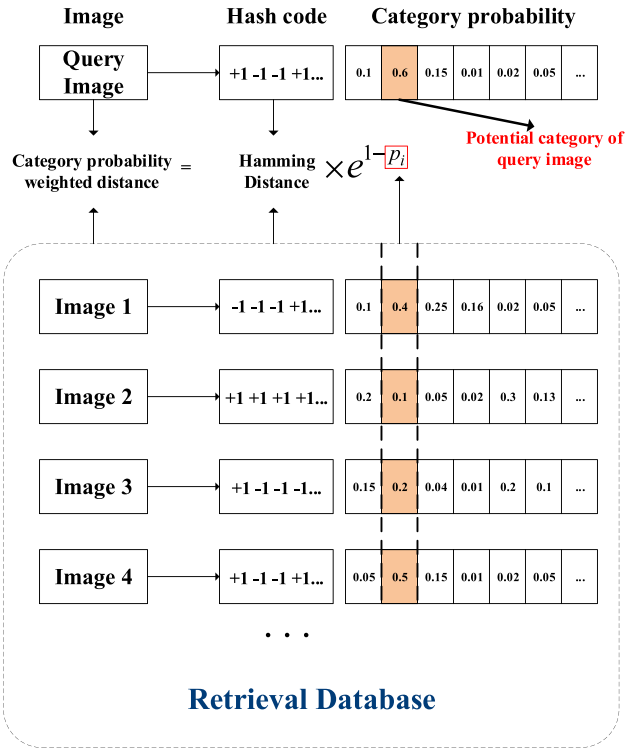

Fig. 4. Calculation of the category-weighted Hamming distance.

probability weight W_c can be described as

$$W_c = e^{1-p_i}. \tag{15}$$

When image x_i in the retrieval database has a larger probability value in the potential category C_q of the query image q_j, the adaptive weight for the Hamming distance is reduced, and vice versa. The Hamming distance with the category probability weight is

$$D_a\left(b_j^q, b_i\right) = W_c \times D_h\left(b_j^q, b_i\right). \tag{16}$$

It makes images with the same potential category as the query image closer together, thus ranking higher.

IV. Experiments

A. Implementation Settings

The proposed DAH method is implemented by the PyTorch deep learning framework with GeForce GTX TITAN X. Stochastic Gradient Descent is employed to update the parameters throughout the training phase. The learning rate is set to 0.01, the momentum to 0.9, and the weight decay to 0.0005. The size of each training batch is set to 36, and the total number of epochs is set to 30.

B. Datasets

To evaluate the performance of the proposed DAH method and other comparative methods, experiments are conducted on two popular remote sensing image datasets: AID and PatternNet.

1) AID Remote Sensing Image Dataset [44]: AID remote sensing image dataset contains 30 categories, including airports, farms, roads, rivers, etc. The number of images in each category ranges from 220 to 420, with 10 000 remote sensing images.

2) PatternNet Remote Sensing Image Dataset [45]: PatternNet remote sensing image dataset contains 38 categories, and each category contains 800 remote sensing images, a total of 30 400 images. The spatial resolution of the images is between 0.062 and 4.693 m, and the categories cover football fields, golf courses, airports, parks, etc.

For the aforementioned datasets, 80% of the remote sensing images of each category are randomly sampled as the training set, and the reminder 20% are used as the testing set.

C. Quantitative Evaluation Metrics

Two quantitative evaluation metrics are utilized in the experiments to evaluate the performance of remote sensing image retrieval.

1) Mean Average Precision (mAP): mAP is used to assess overall retrieval performance. The greater the mAP value, the higher the retrieval performance.

2) Average Normalized Modified Retrieval Rank: ANMRR is to measure the ranking of the correct images in the retrieval results. The lower the ANMRR, the higher the correct image ranks and the better the retrieval performance.

D. Ablation Analysis

To verify the effectiveness of each part, the ablation analysis is implemented on the AID and PatternNet datasets. DAH-A represents a variation of DAH without the attention mechanism module, and DAH-D is another version without distance-adaptive ranking. The ablation experimental results on the AID dataset are shown in Table I. The bolded font indicates the optimal results of the two metrics for different hash code lengths.

As can be seen from Table I, when the length of the hash codes is short, the retrieval performance by using the attention module is relatively improved. For example, with the 16-bit hash code, using the attention module improves the mAP by nearly 6% compared with the version without the attention module. However, with the increase of hash code length, the mAP will decrease by about 1% to 2%, and the value of ANMRR gets bigger when the attention module is not used. In the retrieval phase, if only the original Hamming distance is used to measure the image similarity, the mAP will decrease about 2% to 3%.

The ablation experimental results on the PatternNet dataset are shown in Table II. The bolded font indicates the optimal results of the two metrics for different hash code lengths. Similar to the results of the ablation experiments conducted on the AID dataset, a more significant improvement in the mAP can be achieved when using the attention module in the case of lower length hash codes. The mAP decreases by about 2% to 4% without using the attention module. In the retrieval phase, if the category information is not used as the weight of Hamming distance, the mAP decreases by about 1%, and the value of ANMRR gets bigger. The aforementioned ablation experiments demonstrate the effectiveness of each module of the proposed DAH model.

E. Comparative Experimental Results on Benchmark Datasets

To verify the effectiveness of the proposed DAH method, five deep hashing retrieval models are selected for comparative

TABLE I
ABLATION EXPERIMENTS ON AID DATASET

Methods	MAP(%)				ANMRR(%)			
	16bit	32bit	48bit	64bit	16bit	32bit	48bit	64bit
DAH-A	76.47	85.31	86.95	86.83	12.69	7.61	6.72	6.72
DAH-D	80.81	83.87	85.58	85.64	9.90	8.07	6.86	6.82
DAH	**82.26**	**86.10**	**87.38**	**87.80**	**9.28**	**7.35**	**6.41**	**6.24**

The significance of bold values is optimal experimental result.

TABLE II
ABLATION EXPERIMENTS ON PATTERNNET DATASET

Methods	MAP(%)				ANMRR(%)			
	16bit	32bit	48bit	64bit	16bit	32bit	48bit	64bit
DAH-A	91.47	92.78	96.17	95.67	5.37	4.82	2.58	2.99
DAH-D	95.54	93.41	96.79	97.64	2.77	4.19	2.04	1.52
DAH	**96.12**	**94.09**	**97.28**	**97.96**	**2.47**	**3.93**	**1.81**	**1.39**

The significance of bold values is optimal experimental result.

TABLE III
COMPARATIVE EXPERIMENTAL RESULTS ON THE AID DATASET

Methods	mAP(%)				ANMRR(%)			
	16bit	32bit	48bit	64bit	16bit	32bit	48bit	64bit
DHN [46]	59.98	66.81	67.48	66.82	22.75	19.04	18.72	19.03
DPSH [12]	57.73	72.23	72.80	73.22	25.80	17.44	17.08	16.96
HashNet [47]	65.94	76.37	77.65	78.13	19.92	14.01	13.30	12.99
DSDH [13]	55.47	66.42	69.77	71.37	28.93	22.12	19.37	18.47
GreedyHash [48]	75.19	81.73	83.26	84.61	15.43	10.56	3.56	7.92
DHNNs-L2 [14]	78.28	82.61	83.36	84.21	10.57	9.17	8.92	8.73
DAH	**82.26**	**86.10**	**87.38**	**87.80**	**9.28**	**7.35**	**6.41**	**6.24**

The significance of bold values is optimal experimental result.

experiments: deep hashing network (DHN) [46], deep pairwise supervised hashing (DPSH) [12], HashNet [47], deep supervised discrete hashing (DSDH) [13], GreedyHash [48], and deep hashing neural networks (DHNNs) [14]. Both the proposed DAH method and comparative methods use pairwise similarity to extract deep features and learn hash code. And they are both hashing-based image retrieval methods. However, the proposed DAH method employs a channel-spatial joint attention module in the process of remote sensing image feature extraction, and incorporates the category probabilistic information into the calculation of the Hamming distance by distance-adaptive ranking to further enhance the retrieval performance. In addition, the proposed DAH method considers the imbalance of sample pairs in the pairwise hash code learning phase and balanced pairwise weighted loss is designed to reduce this imbalance. The open-source codes of DHN, DPSH, HashNet, and DSDH methods are used for comparative experiments on the AID and PatternNet datasets. The reproduction code of DHNNs-L2 method is also implemented on the AID and PatternNet datasets.

1) Comparative Experimental Results on AID Dataset: The comparative experimental results on the AID dataset are shown in Table III. The bolded font indicates the optimal results of the two metrics for different hash code lengths. As can be seen from Table III, the proposed DAH method outperforms other hashing image retrieval methods (DHN, DPSH, HashNet, DSDH, and DHNNS-L2) in terms of retrieval performance. For example, the proposed DAH method improves the mAP from DHN (66.82%), DPSH (73.22%), HashNet (78.13%), DSDH (71.37%), GreedyHash (84.61%) and DHNNs-L2 (84.21%) to 87.80% with 64-bit hash code. Moreover, the ANMRR of the proposed DAH method is reduced from 7.92% to 6.24% compared to the DHNNs-L2 method, which was the best performer among comparative methods.

To comprehensively reflect the performance of the proposed DAH method, some other comparative results on AID dataset are shown in Fig. 5. Fig. 5(a) represents the precision of the retrieval results within 2 with Hamming distance using hash codes of different lengths. It can be observed from Fig. 5(a) that the proposed DAH method performs optimally in this evaluation metric for hash code lengths of 16, 32, 48, and 64. When 64-bit hash codes are used, the samples are retrieved that their distance from the query image are less than the Hamming distance thresholds (0–64). These retrieved samples are used to calculate recall and mAP. The recall curve with different Hamming distance thresholds (0–64) is shown in Fig. 5(b). The mAP with different Hamming distance thresholds (0–64) is shown in Fig. 5(c). The recall of DAH is not the highest at the short Hamming distance

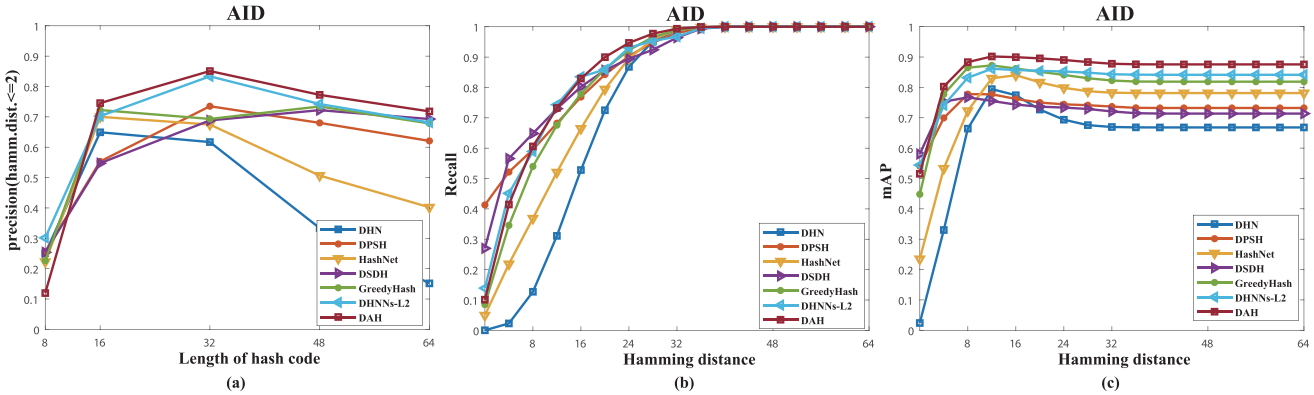

Fig. 5. Comparative experimental results on AID dataset. (a) Precision of the retrieval results within 2 with Hamming distance using hash codes of different lengths. (b) Recall curve with different Hamming distance thresholds (0–64) when 64-bit hash codes are used for retrieval. (c) mAP with different Hamming distance thresholds (0–64) when 64-bit hash codes are used for retrieval.

TABLE IV
COMPARATIVE EXPERIMENTAL RESULTS ON THE PATTERNNET DATASET

Methods	mAP(%)				ANMRR(%)			
	16bit	32bit	48bit	64bit	16bit	32bit	48bit	64bit
DHN [46]	78.02	86.38	87.55	87.93	13.72	8.30	7.48	7.28
DPSH [12]	86.48	87.28	96.45	96.46	7.78	8.26	1.99	1.99
HashNet [47]	73.41	91.88	93.60	94.02	16.11	5.08	4.04	3.78
DSDH [13]	44.88	62.80	71.54	73.27	38.52	24.21	18.14	17.14
GreedyHash [48]	85.36	92.84	93.75	94.78	8.83	4.72	4.04	3.56
DHNNs-L2 [14]	89.35	91.62	92.73	93.21	5.29	4.52	4.39	4.18
DAH	**96.12**	**94.09**	**97.28**	**97.96**	**2.47**	**3.93**	**1.81**	**1.39**

The significance of bold values is optimal experimental result.

threshold, as shown in Fig. 5(b). However, the DAH outperforms other methods during the increase of the Hamming distance threshold, indicating that the DAH is more sensitive to hard positive samples. According to the comparative results on the AID dataset, the proposed method can better extract the hash code with discrimination and filter out worthless background and noise information in remote sensing images with greater noise and fewer prominent objects.

2) Comparative Experimental Results on PatternNet Dataset: Table IV provides the mAP and ANMRR results of the comparative results on the PatternNet dataset. The bolded font indicates the optimal results of the two metrics for different hash code lengths. Similar to the experiments on AID dataset, the proposed DAH method leads other comparative hashing methods on the retrieval performance, as can be observed in Table IV. For example, using 64-bit hash code, the proposed DAH method increases the mAP from DHN (87.93%), DPSH (96.46%), HashNet (94.02%), DSDH (73.27%), GreedyHash (94.78%), and DHNNs-L2 (93.21%) to 97.96%. Furthermore, compared to the DHNNs-L2 method, which was the top performer among the comparable methods, the ANMRR of the proposed DAH method is lowered from 1.99% to 1.39%. The performance improvement of the proposed DAH method on the PatternNet dataset is more significant than that on the AID dataset.

The comparative results on PatternNet dataset are shown in Fig. 6. Fig. 6(a) represents the precision of the retrieval results within 2 with Hamming distance using hash codes of different lengths. Fig. 6(a) shows that the proposed DAH method performs best in this evaluation metric for hash code lengths of 16, 32, and 48. The recall curve with different Hamming distance thresholds (0–64) is shown in Fig. 6(b). The mAP with different Hamming distance thresholds (0–64) is shown in Fig. 6(c). The recall of DAH is not the highest at the short Hamming distance threshold, as shown in Fig. 6(b). However, in combination with Fig. 6(c), the proposed DAH method is almost leading in the corresponding mAP with all Hamming distance thresholds. Overall, the proposed DAH method performs the best compared to other hashing-based retrieval methods. This is because the proposed DAH method implements a channel-spatial joint attention module in the process of remote sensing image feature extraction, and includes category probabilistic information into the computation of Hamming distance via the distance-adaptive ranking to improve model's understanding of images and retrieval performance. The comparative experimental results on the AID and PatternNet datasets demonstrate the effectiveness and advancement of the proposed DAH method even further.

F. Analysis of Retrieval Results

Fig. 7 shows the top ten retrieval results by the proposed DAH method with 64-bit hash code on AID dataset. The five query examples shown in Fig. 7 are belong to five categories, namely beach, playground, airport, forest, and desert. The images on

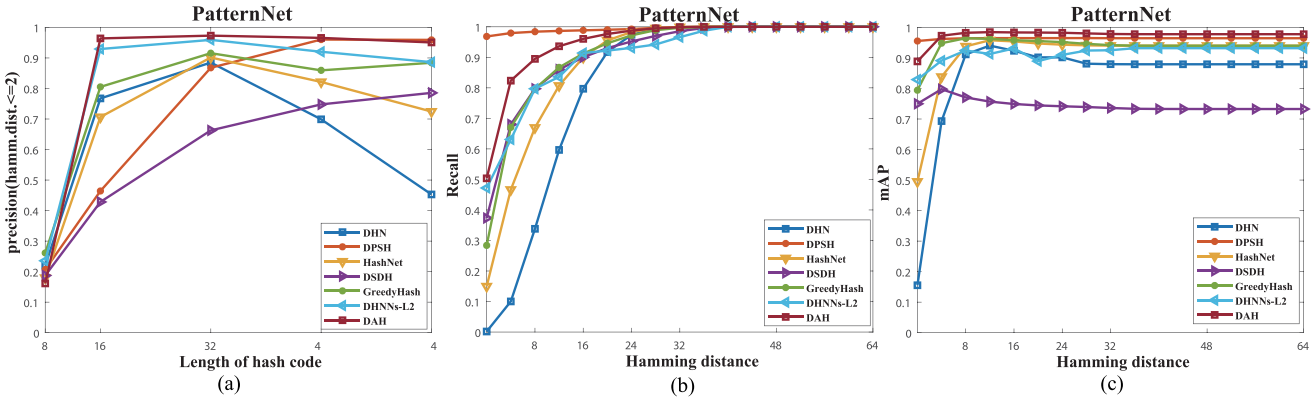

Fig. 6. Comparative experimental results on PatternNet dataset. (a) Precision of the retrieval results within 2 with Hamming distance using hash codes of different lengths. (b) Recall curve with different Hamming distance thresholds (0–64) when 64-bit hash codes are used for retrieval. (c) mAP with different Hamming distance thresholds (0–64) when 64-bit hash codes are used for retrieval.

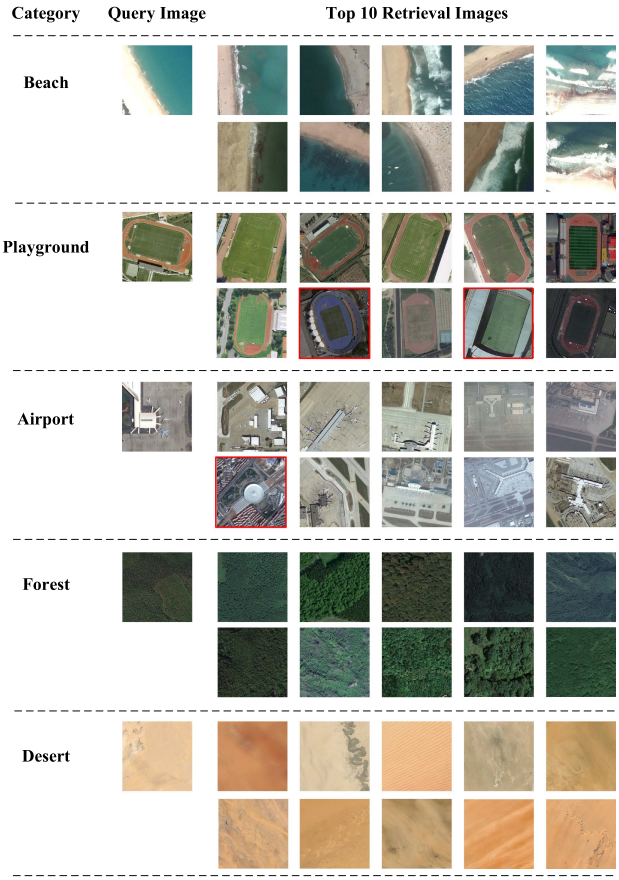

Fig. 7. Top ten retrieval results by the proposed DAH method on AID dataset. Incorrect retrieval results are marked with red boxes.

the left are the query images, the corresponding top ten retrieval results are shown on the right. The retrieval results whose boundaries are colored red indicate incorrect results. There are no inaccurate results among the top ten image retrieval results for the "beach," "forest," and "desert" categories. The retrieval results of the query image from "airport" have one error result from the category "center." The objects of "center" and "airport" have high similarity in color and shape. The retrieval results of the query image from "playground" have two incorrect results from the category "stadium." The playground is also included in the remote sensing images of the "stadium," which cause the model to make mistakes. However, this demonstrates that the proposed DAH method retrieves remote sensing images based on the essential visual information in images. Overall, the proposed DAH method has a satisfactory retrieval performance.

G. Further Analysis

1) Selection of Attention Module: In order to discuss the performance of different attention modules, the Squeeze-and-Excitation (SE) attention module [49] based on channel attention was used to compare with the attention mechanism used in this article on PatternNet Dataset. Table V shows the performance of different attention modules on PatternNet Dataset. DAH+SE represents a variant of DAH that the attention mechanism is replaced with SE. As can be seen from Table V, the channel-spatial joint attention used in this article has a certain advantage over the SE attention mechanism in the retrieval performance.

2) Analysis of Hyperparameter: To analyze the selection of weights for the loss terms, a hyperparameter analysis experiment is implemented. Two hyperparameters α and β in the loss function are utilized for hash code learning. $L_{w-\text{cross}}$ serves as the core of the loss function and is used to generate hash codes in a pairwise hash code learning manner. The hyperparameter α represents the contribution of quantization loss L_q. The hyperparameter β determines the contribution of classification loss L_{class}. This experiment modifies one of the two hyperparameters while leaving the other alone. The value of α or β is fixed to 1, and the value of β or α is set to 0, 0.1, 1, and 2, respectively. The hyperparameter experiments results are shown in Fig. 8. From the experimental results, it can be seen that the retrieval performance is optimal when hyperparameters α and β are both set to 1. α or β is set to 0 corresponds to removing the quantization loss or classification loss term from the loss function, respectively. It can be seen that the contributions of the two loss terms are similar, with the classification loss slightly higher than the quantization loss in terms of contribution.

3) Analysis of Running Efficiency and Time Complexity: The theoretical inference time complexity of the proposed framework can be approximated as $O(N^2 \cdot K)$, where N is the

TABLE V
PERFORMANCE OF DIFFERENT ATTENTION MODULES ON PATTERNNET DATASET

Methods	MAP(%)				ANMRR(%)			
	16bit	32bit	48bit	64bit	16bit	32bit	48bit	64bit
DAH+SE	94.26	95.31	95.66	96.72	3.08	2.75	2.24	1.98
DAH	**96.12**	**94.09**	**97.28**	**97.96**	**2.47**	**3.93**	**1.81**	**1.39**

The significance of bold values is optimal experimental result.

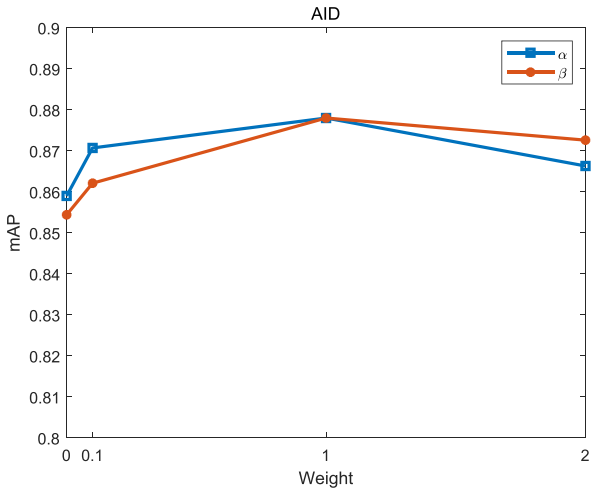

Fig. 8. Analysis of hyperparameter.

TABLE VI
RUNNING EFFICIENCY AND TIME COMPLEXITY OF THE PROPOSED DAH METHOD

	Inference time complexity	Running time
DAH	$O(N^2 \cdot K)$	0.932 ms

number of images in the remote sensing dataset and K is the length of the hash code [50]. To test the running efficiency, the proposed DAH method is implemented on a server with GeForce GTX TITAN X and measured the single query runtime. The single query runtime refers to the time it takes to compare a single query to all images in the test set of the PatternNet dataset. The results of single query runtime and the inference time complexity of the proposed DAH method are reported in Table VI.

V. CONCLUSION

To address the problems of inadequate feature expression and interference by background information, a novel DAH is proposed for remote sensing image retrieval. First, a channel-spatial joint attention module is exploited to direct the network to pay greater attention to meaningful visual information, which reduces interference from irrelevant information. Second, a novel balanced pairwise weighted loss function is proposed by combing pairwise weighted similarity loss, classification loss, and quantization loss. Finally, a distance-adaptive ranking method is proposed further to enhance the retrieval precision in the retrieval phase. The experimental results prove the effectiveness and advancement of the proposed DAH method. In the future, more efficient methods for learning hash codes will be explored to enable faster and more accurate retrieval of remote sensing images. Additionally, multilabel remote sensing image retrieval methods will be studied that are tailored to remote sensing images with multiple semantic information.

REFERENCES

[1] B. Zhang et al., "Remotely sensed Big Data: Evolution in model development for information extraction [point of view]," *Proc. IEEE*, vol. 107, no. 12, pp. 2294–2301, Dec. 2019.
[2] O. E. Dai, B. Demir, B. Sankur, and L. Bruzzone, "A novel system for content-based retrieval of single and multi-label high-dimensional remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2473–2490, Jul. 2018.
[3] X.-Y. Tong, G.-S. Xia, F. Hu, Y. Zhong, M. Datcu, and L. Zhang, "Exploiting deep features for remote sensing image retrieval: A systematic investigation," *IEEE Trans. Big Data*, vol. 6, no. 3, pp. 507–521, Sep. 2020.
[4] Y. Li, J. Ma, and Y. Zhang, "Image retrieval from remote sensing Big Data: A survey," *Inf. Fusion*, vol. 67, pp. 94–115, 2021.
[5] Y. Chen, D. Zhao, X. Lu, S. Xiong, and H. Wang, "Unsupervised balanced hash codes learning with multichannel feature fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, no. 3, pp. 2816–2825, Mar. 2022.
[6] Y. Li, S. Yang, T. Liu, and X. Dong, "Comparative assessment of semantic-sensitive satellite image retrieval: Simple and context-sensitive Bayesian networks," *Int. J. Geographical Inf. Sci.*, vol. 26, no. 2, pp. 247–263, 2012.
[7] H. Sebai, A. Kourgli, and A. Serir, "Dual-tree complex wavelet transform applied on color descriptors for remote-sensed images retrieval," *J. Appl. Remote Sens.*, vol. 9, no. 1, pp. 095994–095994, 2015.
[8] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 5, pp. 3735–3756, Jun. 2020.
[9] X. Zheng, Y. Zhang, and X. Lu, "Deep balanced discrete hashing for image retrieval," *Neurocomputing*, vol. 403, pp. 224–236, 2020.
[10] Y. Liu, L. Ding, C. Chen, and Y. Liu, "Similarity-based unsupervised deep transfer learning for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7872–7889, Nov. 2020.
[11] S. Roy, E. Sangineto, B. Demir, and N. Sebe, "Metric-learning-based deep hashing network for content-based retrieval of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 2, pp. 226–230, Feb. 2021.
[12] W. Li, S. Wang, and W. Kang, "Feature learning based deep supervised hashing with pairwise labels," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 1711–1717.
[13] Q. Li, Z. Sun, R. He, and T. Tan, "A general framework for deep supervised discrete hashing," *Int. J. Comput. Vis.*, vol. 128, no. 8, pp. 2204–2222, 2020.
[14] 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.
[15] W. Hu, L. Wu, M. Jian, Y. Chen, and H. Yu, "Cosine metric supervised deep hashing with balanced similarity," *Neurocomputing*, vol. 448, pp. 94–105, 2021.
[16] Q. Bao and P. Guo, "Comparative studies on similarity measures for remote sensing image retrieval," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, 2004, vol. 1, pp. 1112–1116.
[17] Y. Zhang, X. Zheng, and X. Lu, "Remote sensing cross-modal retrieval by deep image-voice hashing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, no. 9, pp. 9327–9338, Oct. 2022.
[18] H. Zhang, J. Yao, L. Ni, L. Gao, and M. Huang, "Multimodal attention-aware convolutional neural networks for classification of hyperspectral and LiDAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 3635–3644, Jul. 2023.

[19] J. Kang, R. Fernandez-Beltran, Z. Ye, X. Tong, and A. Plaza, "Deep hashing based on class-discriminated neighborhood embedding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 8, pp. 5998–6007, Sep. 2020.
[20] M. Wang and T. Song, "Remote sensing image retrieval by scene semantic matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2874–2886, May 2013.
[21] C. Ma, F. Chen, J. Yang, J. Liu, W. Xia, and X. Li, "A remote-sensing image-retrieval model based on an ensemble neural networks," *Big Earth Data*, vol. 2, no. 4, pp. 351–367, 2018.
[22] Y. Wang et al., "A three-layered graph-based learning approach for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6020–6034, Oct. 2016.
[23] K. N. Sukhia, M. M. Riaz, A. Ghafoor, and S. S. Ali, "Content-based remote sensing image retrieval using multi-scale local ternary pattern," *Digit. Signal Process.*, vol. 104, 2020, Art. no. 102765.
[24] A. P. Byju, B. Demir, and L. Bruzzone, "A progressive content-based image retrieval in JPEG 2000 compressed remote sensing archives," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5739–5751, Aug. 2020.
[25] F. Ahmad, "Deep image retrieval using artificial neural network interpolation and indexing based on similarity measurement," *CAAI Trans. Intell. Technol.*, vol. 7, no. 2, pp. 200–218, 2022.
[26] R. Imbriaco et al., "Aggregated deep local features for remote sensing image retrieval," *Remote Sens.*, vol. 11, no. 5, 2019, Art. no. 493.
[27] U. Chaudhuri, B. Banerjee, and A. Bhattacharya, "Siamese graph convolutional network for content based remote sensing image retrieval," *Comput. Vis. Image Understanding*, vol. 184, pp. 22–30, 2019.
[28] L. Fan, H. Zhao, and H. Zhao, "Distribution consistency loss for large-scale remote sensing image retrieval," *Remote Sens.*, vol. 12, no. 1, 2020, Art. no. 175.
[29] Y. Liu, Z. Han, C. Chen, L. Ding, and Y. Liu, "Eagle-eyed multitask CNNs for aerial image retrieval and scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6699–6721, Sep. 2020.
[30] M. Abu-Hashem and A. Gutub, "Efficient computation of Hash Hirschberg protein alignment utilizing hyper threading multi-core sharing technology," *CAAI Trans. Intell. Technol.*, vol. 7, no. 2, pp. 278–291, 2022.
[31] 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, Feb. 2016.
[32] 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.
[33] R. Fernandez-Beltran, B. Demir, F. Pla, and A. Plaza, "Unsupervised remote sensing image retrieval using probabilistic latent semantic hashing," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 2, pp. 256–260, Feb. 2021.
[34] 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, 2017, Art. no. 364.
[35] T. Reato, B. Demir, and L. Bruzzone, "Primitive cluster sensitive hashing for scalable content-based image retrieval in remote sensing archives," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 2199–2202.
[36] W. Song, S. Li, and J. A. Benediktsson, "Deep hashing learning for visual and semantic retrieval of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9661–9672, Nov. 2021.
[37] L. Han, P. Li, X. Bai, C. Grecos, X. Zhang, and P. Ren, "Cohesion intensive deep hashing for remote sensing image retrieval," *Remote Sens.*, vol. 12, no. 1, 2019, Art. no. 101.
[38] C. Liu, J. Ma, X. Tang, X. Zhang, and L. Jiao, "Adversarial hash-code learning for remote sensing image retrieval," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 4324–4327.
[39] P. Li et al., "Hashing nets for hashing: A quantized deep learning to hash framework for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7331–7345, Oct. 2020.
[40] X. Tang et al., "Meta-hashing for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 12, pp. 1–19, Dec. 2021.
[41] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
[42] J. Yao, D. Hong, J. Chanussot, D. Meng, X. Zhu, and Z. Xu, "Cross-attention in coupled unmixing Nets for unsupervised hyperspectral super-resolution," in *Proc. 16th Eur. Conf. Comput. Vis.*, Glasgow, U.K., 2020, pp. 208–224.
[43] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
[44] G.-S. Xia et al., "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.
[45] W. Zhou, S. Newsam, C. Li, and Z. Shao, "PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 197–209, 2018.
[46] H. Zhu, M. Long, J. Wang, and Y. Cao, "Deep hashing network for efficient similarity retrieval," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 2415–2421.
[47] Z. Cao, M. Long, J. Wang, and P. S. Yu, "HashNet: Deep learning to hash by continuation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5608–5617.
[48] S. Su, C. Zhang, K. Han, and Y. Tian, "Greedy hash: Towards fast optimization for accurate hash coding in CNN," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 806–815.
[49] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[50] Z. Wang et al., "Camp: Cross-modal adaptive message passing for text-image retrieval," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 5764–5773.

Yichao Zhang is currently working toward the Ph.D. degree with the Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, and also with the University of Chinese Academy of Sciences, Beijing, China.

His current research interests include pattern recognition, computer vision, and machine learning.

Xiangtao Zheng (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China, in 2014 and 2017, respectively.

He is currently an Associate Professor with the Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China. His main research interests include computer vision and pattern recognition.

Xiaoqiang Lu (Senior Member, IEEE) received the Ph.D. degree in signal and information processing from Dalian University of Technology, Dalian, China, in 2010.

He is currently a Full Professor with the Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China. His current research interests include pattern recognition, machine learning, hyperspectral image analysis, cellular automata, and medical imaging.