Deep Unsupervised Weighted Hashing for Remote Sensing Image Retrieval

Weipeng Jing, Northeast Forestry University, China

Zekun Xu, Northeast Forestry University, China*

🆔 https://orcid.org/0000-0003-0849-4650

Linhui Li, Northeast Forestry University, China

Jian Wang, Aerospace Information Research Institute, China

Yue He, Northeast Forestry University, China

Guangsheng Chen, Northeast Forestry University, China

ABSTRACT

Deep unsupervised hashing methods are gaining attention in the field of remote sensing (RS) image retrieval due to the rapid growth in the volume of unlabeled RS data. Most previous unsupervised hashing research used only natural image-based pre-trained models to generate label matrices; however, this method cannot capture the semantic information of RS images well and limits the accuracy of retrieval. To solve this problem, the authors propose a deep unsupervised weighted hashing (DUWH) model that uses a similarity matrix updating strategy based on a weighted similarity structure to achieve the mutual optimization of the similarity matrix and hash network. The authors devise a novel combinatorial loss function to improve the hash performance that can be used to obtain higher quality hash codes by assigning different weights to the sample pairs with different difficulties. Experiments were conducted on two RS datasets to verify the excellent performance of the proposed method.

KEYWORDS

Adaptive Weight, Deep Unsupervised Learning, Hash Learning, Remote Sensing Image Retrieval, Similarity Matrix Update

INTRODUCTION

In recent years, remote sensing (RS) observation technology has developed rapidly, harvesting vast amounts of unprocessed RS images in the global region. Therefore, how to retrieve the required images accurately and efficiently in a large-scale RS image database has become an increasingly noteworthy direction (Tang et al., 2018). Although content-based image retrieval (CBIR) methods have made good research progress in the fields of medical images, traditional images, and video images, there are fewer techniques for content-based remote sensing image retrieval (CBRSIR). CBRSIR mainly includes the extraction of effective features and metrics of inter-image similarity (Demir & Bruzzone, 2014; Lu & Man, 2013). The representative features of RS images are extracted validly,

and all similar images are retrieved by measuring the similarity between the query image and the target image. Among them, the hash-based approximate nearest neighbor search method has been widely used because of its high query efficiency and low storage cost (Indyk & Motwani, 1998). The hash algorithm in image retrieval represents the useful features of RS images as a series of binary hash codes, so as to come to reduce the search cost.

Generally, there are broadly two types of image-based hashing methods: supervised hashing and unsupervised hashing. The supervised method mainly combines the feature vector information of the data and the labeled similarity information between data to learn the hash function. Currently, most of the work still revolves around supervised hash learning, and representative methods include supervised discrete hashing (SDH) (Shen et al., 2015), fast supervised hashing (FastH) (Lin et al., 2014), ranking-based supervised hashing (RSH) (Wang et al., 2013), and deep supervised hashing. However, the production of high-quality label information needs significant cost and manual labor in practical application. The unsupervised hashing method does not utilize any supervised information but uses the internal information of all images alone for hash learning. Representative methods include spectral hashing (SH) (Weiss et al., 2008), shift-invariant kernelized locality-sensitive hashing (SKLSH) (Raginsky & Lazebnik, 2009), iterative quantization hashing (ITQ) (Gong et al., 2012), and unsupervised hashing based on deep learning. Unsupervised hashing is simple and easy to implement in the learning process because the label information is not used, which saves the cost of data annotation. In reality, most RS data have no semantic labels, so unsupervised hashing is more effective in future practical applications, which is also the focus of this work.

At present the depth network's strong fitting ability has become an effective means to extract influence features from the unsupervised hash. Since data has no label information, how to accurately construct the semantic similarity structure between data is still an open problem. Some methods use machine learning methods to construct pseudo-labels as a way to learn similarities, however, the semantic information is weak in pseudo-labels and cannot reach satisfactory performance results (Dong et al., 2020; Hu et al., 2017; Song et al., 2015). Shen et al. (2018) adopted the model and pseudo-label alternate optimization method to update the data similarity graph for the first time; the construction method has a large localization, which affects the search performance. In the traditional image domain, Lin et al. (2021) used a pre-trained CNN model to obtain the initial label matrix, and excavated the potential neighbor relations behind the features during the training process, then updated the similarity pairs in the labels.

In the existing research, the loss function of most methods is constructed on the sample pair or the sample triple (Chen et al., 2020; Li et al., 2020; Liu et al., 2018), and the same weight is used for the sample pair. The purpose is to widen the distance between non-similar samples and at the same time close the distance between similar samples. However, RS image retrieval has correspondingly difficult samples in both positive and negative sample pairs due to the problem of large intra-class variation and high inter-class similarity, and if the difficult sample pairs are incorrectly determined (e.g., near a false positive sample pair or far from a false negative sample pair) during the training process, both will amplify the hash learning error. At the same time, the loss functions of many methods ignore the problem of imbalance in the data during training. In the input images of each batch during the training process, the amount of similar image pairs is only a small fraction of the total, resulting in excessive attention to negative samples during the training process and causing overfitting of negative samples.

In order to respond to the above issues, the authors propose a novel deep unsupervised hashing framework, called Deep Unsupervised Weighted Hashing (DUWH), which is used to mine semantic information in RS images and learn binary codes that maintain similarity by combining discriminant information between different difficulty sample pairs. The main contributions of this work are as follows:

- In the fine-tuning process of the pre-trained Swin Transformer, the output new feature representation is used to regenerate the similarity matrix and update it by capturing the information of neighboring images to better characterize the relationship of image pairs, which greatly improves the optimization of the subsequent binary hash code.
- The authors design an adaptive weight-based loss function to assign different weights to positive and negative example image pairs with different degrees of difficulty to utilize the discriminative information of the image pairs more effectively and further improve the performance of the model.
- Through comprehensive evaluation on two benchmark RS datasets, it is concluded that the DUWH can achieve the most advanced results.

BACKGROUND

From the perspective of whether their supervised information is involved in the learning process of hash codes, the existing hash methods can be roughly classified into supervised methods and unsupervised methods (Li et al., 2021). Generally, supervised hashing methods utilize label information for hash code learning. Research in recent years has used many kinds of loss functions to base CNN hash models. Xiong et al. (2019) adopted a central loss function to learn the features of RS images through intra-class compression and inter-class dispersion. Cao et al. (2020) proposed a depth measurement learning CNN model based on Triplet loss for CBRSIR. Roy et al. (2018) combined Triplet loss, representation penalty, and balance loss to mine a metric space containing semantic information. However, due to the problem of local optimization in Triplet loss, replacing the global optimal structural loss of the hinge function with the softmax function effectively solved the problem of Triplet loss (Liu et al., 2020).To cope with the lack of performance of the standard CNN model for partial RS image retrieval, graph neural networks were introduced to optimize the contrast loss and used to better acquire local interactions within image regions (Chaudhuri et al., 2019). Fan et al. (2020) put forward a new result ranking loss to control the loss reduction direction to be consistent with the optimization direction to achieve end-to-end hash learning optimization.

Unsupervised hashing methods are beginning to receive more attention for wider, unmarked data. Unsupervised hashing algorithms usually need to maintain the importance of training data points and train the model by using distance functions to learn the underlying deep semantic relationships in the data (Lin et al., 2022). In the field of natural image retrieval, Deng et al. (2019) obtained the initial similarity matrix by performing a k-nearest neighbor search according to the cosine distances of image features; this monitoring of the semantic similarity matrix was integrated into an antagonistic learning framework, which came to retain the semantic information. Tu et al. (2020) adopted a local manifold structure to build a similar structure for local types and optimized the network using a log-cash loss function. On the CBRSIR side, Li et al. (2016) proposed a synergistic affinity merger to measure the similarity between RS images. Zhang et al. (2020) constructed a hybrid similarity matrix by combining image pairs with high and normal confidence levels. Sun et al. (2021) presented soft pseudo-labels to learn the similarity of image pairs. Most of the existing unsupervised hash algorithms only generate the initial semantic similarity matrix as pseudo-labels. However, pre-trained network models on the natural image datasets cannot generate the feature representation of RS images very well. Therefore, we consider that through the similarity matrix to help the model fine-tuning, and the matrix itself can be updated with the new feature representation.

On the other hand, the complex and diverse content of high-resolution RS images leads to the limitation of CNN by the local constraint of convolution operation, so the migration to high-resolution RS images cannot achieve the best retrieval performance. In contrast, the naturally self-contained long range property of the Transformer made it more capable of utilizing the global effective information from shallow to deep layers (Dosovitskiy et al., 2020; Khan et al., 2021; Vaswani et al., 2017). El-Nouby et al. (2021) used Vision Transformers to generate feature representations and optimized the model through differential entropy regularization and contrast loss. Swin Transformer introduced the

hierarchical construction method commonly used in CNN to construct the hierarchical Transformer, and introduced the locality idea to perform self-attention calculation in the non-overlapping areas, which substantially reduced the computational complexity (Liu et al., 2021). Therefore, in this paper, the Swin Transformer network model pre-trained on ImageNet is used for unsupervised hash learning, which is the first time RS image retrieval has been implemented based on Swin Transformer. In addition, an adaptive weight similar loss function is designed for positive and negative sample pairs with different degrees of difficulty, which is combined with quantitative loss to optimize hash learning.

DEEP UNSUPERVISED WEIGHTED HASHING

Notation and Problem Definition

Bold lowercase like $\mathbf{m}$ is used to represent the vector. Rough capital letters such as $\mathbf{M}$ represent the matrix. The transpose of $\mathbf{M}$ is expressed as $\boldsymbol{M}^T$. $\left\| \cdot \right\|_F$ denotes the Frobenius norm of the matrix, $\left\| \cdot \right\|$ denotes the Euclidean norm of the vector. $\mathrm{sgn}\left(\cdot \right)$ denotes the sign function; the element is regular return 1, otherwise, return -1. $\tanh\left(\cdot \right)$ is a hyperbolic tangent function with a range of $\left\{ -1,1 \right\}$.

In the training set with n RS images $\boldsymbol{X} = \left\{ \boldsymbol{x_i} \right\}_{i=1}^{n}$, $\boldsymbol{x_i}$ represents the i th image. The goal of the hash algorithm is to map the RS image into a compact k -bit hash code through the hash function F, i.e., $F: \boldsymbol{X} \rightarrow \boldsymbol{B}, \boldsymbol{B} \in \left\{ -1,1 \right\}^{n \times k}$. The algorithm requires that similar images are brought closer together and dissimilar images are pulled further apart.

Network Design

As shown in Figure 1, in the preprocessing stage, the feature representation extracted by the pre-trained model is used to construct the initial similarity matrix by double-layer K-NN as the pseudo-label of the model training. During multiple rounds of training, the similarity matrix is regenerated and optimally updated by a weighted similarity structure. A loss function consisting of adaptive weight similarity loss and quantization loss is designed to achieve assigning different weights to pairs of image samples with different degrees of similarity.

Feature Extraction

The backbone of DUWH adopts the tiny version of Swin Transformer pre-trained on ImageNet-1K. The Swin Transformer network model is a multi-layer structure, which contains four stages. These

Figure 1. DUWH framework
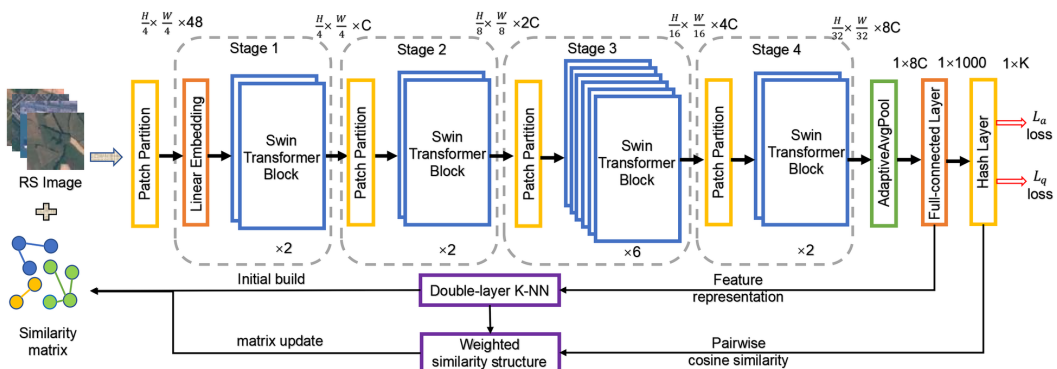
stages achieve a gradual reduction in the resolution of the feature map. After the model starts, Patch Embedding will slice the input image and embed them into Embedding. The four stages include Patch Merging and more than one Blocks, where Patch Merging part is used for the Stage to decrease the RS image resolution. The Block structure is mainly composed of LayerNorm, MLP, Window Attention and Shifted Window Attention. Finally, the output is passed through a layer of AdaptiveAvgPool and a fully connected layer with 1000 neurons. Here, a hash layer is added after the fully connected layer with k neurons.

In the previous deep hashing methods, the sign function was often used directly as an activation function of the hash layer. But the sign function is not a smooth function, and its output at the zero position is incorrect, which can easily cause the problem of "gradient disappearance" (Gu et al., 2018), leading to the failure to effectively use backward propagation to update the model parameters in the training process. So we use a tanh function that can replace the sign function:

$$h\left(x_i\right) = \tanh\left(F\left(x_i, W\right)\right) \tag{1}$$

where $h\left(x_i\right)$ represents the approximate hash encoding of image x_i, $F\left(x_i, W\right)$ represents the feature represented by image x_i as input, and W represents the network parameter.

The k-dimensional approximate hash code is output from the hash layer, which is later quantized into the final binary code with a symbolic function:

$$b_a\left(x_i\right) = \text{sgn}\left(h_a\left(x_i\right)\right) = \begin{cases} 1, & h_a\left(x_i\right) \geq 0 \\ -1, & h_a\left(x_i\right) \leq 0 \end{cases} \tag{2}$$

where $b_a\left(x_i\right)$ represents the a th element of the binary hash encoding of image x_i.

Construction of Initial Similarity Matrix

The aim of unsupervised hashing learning is to analyze the important semantic relationship between data points, learn the image depth features obtained by the deep neural network, and then capture the intrinsic visual connection between different images. The unlabeled data are encoded into binary code by training, and the pseudo-label can be combined with the intrinsic layout of the data to guide learning during model training. Typically, images from the same category should be closer to the pseudo-label space. Therefore, we first need to construct an approximate nearest neighbor matrix M as the initial pseudo-label matrix. Combined with the previous works, the pre-trained model can effectively extract the depth features containing semantic information in the image. Inspired by (Song et al., 2018), this paper adopts a double-layer K-NN method to construct the initial similarity matrix.

We adopt the Swin Transformer network model pre-trained on ImageNet-1K to realize the extraction of high-dimensional features, which is sorted by computing the cosine distance between image x_i and other images. After that, the top K_1 is selected as the initial similar images, and the remaining images are the non-similar images of image x_i. The first layer nearest neighbor matrix is given by:

$$\left(M_{1st}\right)_{ij} = \begin{cases} 1, & x_j \in nn_{K1}^c\left(x_i\right) \\ -1, & otherwise \end{cases} \tag{3}$$

where $nn_{K1}^{c}\left(\boldsymbol{x}_{i}\right)$ denotes the top K_1 in order of cosine distance.

In order to ensure the effectiveness of semantic similarity between image $\boldsymbol{x}_i$ and image $\boldsymbol{x}_j$, in addition to determining the feature distance, the similarity of image pairs can also be determined from the number of identical nearest neighbor objects of the two images. We sort the number of identical nearest neighbor objects and select the top K_2 as similar images, while the remaining images are non-similar images of image $\boldsymbol{x}_i$. The nearest neighbor matrix of the second layer is given by:

$$\left(M_{2nd}\right)_{ij} = \begin{cases} 1, & x_j \in nn_{K2}^{s}\left(x_i\right) \\ -1, & otherwise \end{cases} \tag{4}$$

where $nn_{K2}^{s}\left(\boldsymbol{x}_{i}\right)$ denotes the top K_2 according to the number of common nearest neighbors.

Ultimately, the two nearest neighbor matrices, formula (3) and formula (4), are combined to get the semantic similarity structure of RS images, that is, the initial pseudo-label matrix required for subsequent training. The final semantic similarity matrix for RS images M_{ij} is given by:

$$M_{ij} = \begin{cases} 1, & \left(M_{1st}\right)_{ij} = 1 \ and \left(M_{2nd}\right)_{ij} = 1 \\ -1, & otherwise \end{cases} \tag{5}$$

Semantic Similarity Matrix Update

Due to the improvement of spatial resolution, the detailed information such as shape structure, and texture of the features have become more and more abundant (Song & Yang, 2022), which leads to some differences in the spectral and spatial domains between RS images and natural images, and the pseudo-label generated by the pre-trained model of natural images cannot retain the semantic information of RS images well.

When the network model starts training, the model parameters are fine-tuned in each round of training, which can better mine the semantic information of RS images. By implementing the encoding of RS images through the fine-tuned model, we can extract a more effective depth feature representation and generate a new similarity matrix $\boldsymbol{M}_{ij}^{new}$ by double-layer K-NN, which is specifically as:

$$\bar{F}\left(x_i,\bar{W}\right) \overset{K-NN}{\Rightarrow} M_{ij}^{new} \tag{6}$$

where $\bar{F}\left(\boldsymbol{x}_i,\bar{\boldsymbol{W}}\right)$ is the feature representation part of the fine-tuned network model, formed by removing the final hash layer from $\bar{F}\left(\boldsymbol{x}_i,\bar{\boldsymbol{W}}\right)$.

The generation of the above similarity matrix is calculated only based on the similarity of the image feature representation without considering the neighboring image information of the two images, however, the neighboring image information helps to correct the wrong similarity relationship. When image $\boldsymbol{x}_j$ and image $\boldsymbol{x}_i$ are similar, assuming that image $\boldsymbol{x}_i$ and image $\boldsymbol{x}_{i-g}$ are similar, then image $\boldsymbol{x}_{i-g}$ and image $\boldsymbol{x}_j$ are indirectly similar. Conversely, then image $\boldsymbol{x}_{i-g}$ and image $\boldsymbol{x}_j$ are indirectly dissimilar. Therefore, we further update the new similarity matrix $\boldsymbol{M}_{ij}^{new}$ by combining the weighted similarity structure and the initial similarity matrix. The weighted similar structure diagram is shown in Figure 2.

Figure 2. Weighted similarity structure diagram
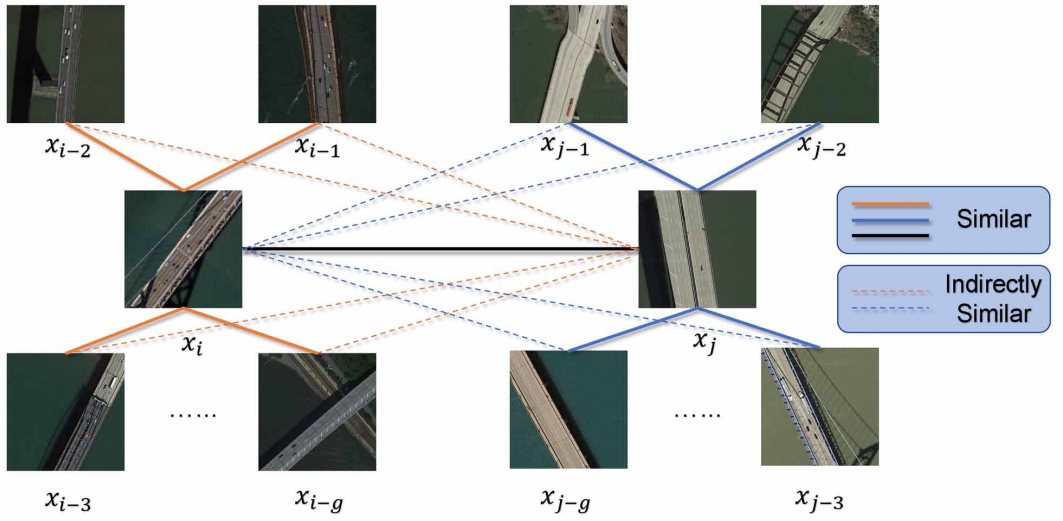
The solid line means the similarity of image pairs, and the dotted line means the contribution of adjacent images to the weighted similarity. Images x_i and x_j represent image pairs for which weighted similarity needs to be calculated. x_{i-g} is the g th adjacent image of image x_i. First, the weights $w\left(x_i, x_{i-g}\right)$ and $w\left(x_j, x_{j-g}\right)$ are calculated by the normalized similarity s_{ii-g} and s_{jj-g}, which is specifically as:

$$w\left(x_i, x_{i-g}\right) = \frac{s_{ii-g}}{\sum_{s_{ii-n} \in S} s_{ii-n}} \tag{7}$$

In which the similarity of image pairs is expressed by measuring the paired cosine similarity. The chord similarity measures the similarity between two vectors by measuring the angle of the inner product space of two vectors:

$$s_{ij} = \cos\left(h_i, h_j\right) = \frac{h_i^T h_j}{\|h_i\| \|h_j\|} \tag{8}$$

The neighboring similarity between the neighboring image of the corresponding central image and the other central image is calculated by the weight w. The neighboring similarity sw is given by:

$$sw\left(x_i\right) = \sum s_{ij-n} w\left(x_j, x_{j-n}\right) \tag{9}$$

The final weighted similarity is calculated by combining the original similarity with the adjacent similarity of the two images:

$$s_{ij}^{new} = \frac{1}{2} s_{ij} + \frac{1}{4} \left(sw\left(x_i\right) + sw\left(x_j\right) \right) \tag{10}$$

Then, similarity threshold $\boldsymbol{u}$ is used to update the similarity pairs in the new matrix, and the threshold is defined as the mean of weighted similarity of all similar image pairs in the initial similarity matrix, which is defined as:

$$\boldsymbol{u} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{s}_{ij}^{new} \tag{11}$$

where n is the number of all similar image pairs in the initial similarity matrix $\boldsymbol{M}_{ij}$, and $\boldsymbol{s}_{ij}^{new}$ is the weighted similarity calculated by the new paired cosine similarity after training.

The final semantic similarity matrix $\boldsymbol{M}_{ij}^{f}$ is obtained after updating the similarity matrix $\boldsymbol{M}_{ij}^{new}$ according to the threshold value, which is specifically as:

$$M_{ij}^{f} = \begin{cases} 1, & m_{ij}^{new} = 1 \ and \ m_{ij} \geq u \\ m_{ij}^{new}, & otherwise \end{cases} \tag{12}$$

In the similarity matrix update part, the fine-tuned model is used to capture the deep semantic information of RS images better, and a new similarity matrix is generated. Meanwhile, the part above the average similarity of similar image pairs is considered a positive sample pair, and the final update of the new similarity matrix is performed. The updated Semantic Similarity Matrix is brought into the next round of training, which enables hash learning better and more effective fine-tuning of the best network model, greatly helping to optimize the subsequent binary hash codes.

Loss Function Module Based on Hash Code

The ultimate goal of the model is to map RS images into hash codes that maintain relative similarity. The similarity preservation loss function is to model the loss of similarity structure in the data, and then optimize the hash learning. In previous hashing studies (Chen & Lu, 2020; Wu et al., 2019; Xia et al., 2014), the following formulas was often used as a similar loss function to bring similar images closer together and pull dissimilar images further apart:

$$\begin{aligned} L_s &= \sum_{i=1}^{n} \sum_{j=1}^{n} \left(s_{ij} - m_{ij} \right)^2 \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\cos\left(h_i, h_j\right) - m_{ij} \right)^2 \end{aligned} \tag{13}$$

However, during the training process, the number of non-similar image sample pairs in each batch of the training input is much more than the number of similar image sample pairs, which causes the positive and negative sample imbalance problem. In some classification tasks, the existence of unbalanced data causes the accuracy of positive samples to be close to 100%, while the accuracy of negative samples is only 0-10%. Therefore, we need to deal with this imbalance problem. The common methods used to cope with the data imbalance problem include sampling, data addition, and weighting (Krawczyk, 2016). Since the RS data samples are large enough and the proportion is not particularly different, we use a weighting method to solve the data imbalance problem within each batch, assigning different weights to different image pairs, so that the network can achieve better results.

There are many positive and negative sample pairs with different difficulties around the query image. Some are easy to distinguish, while others are difficult. Simply assigning the same weights to all image pairs would instead amplify the hash learning error when the difficult samples are misjudged. Therefore, we use adaptive weights as a conditioning factor and use the conditioning factor to readjust the similarity retention loss between simple and difficult examples. The greater the dissimilarity of the image sample pairs, the greater the assigned weights. The adaptive weights are defined as follows:

$$w_{ij} = \sum_{j=0}^{n} p\left(x_i, x_j\right) \log\left(\frac{p\left(x_i, x_j\right)}{q\left(x_i, x_j\right)}\right) \tag{14}$$

where $q\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right)$ denotes the probability that image $\boldsymbol{x_i}$ selects image $\boldsymbol{x_j}$ as the most similar sample.

After entering query image x_i, if the optimal retrieval result returned is image x_j, set positive example sample $p\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) = 1$, otherwise $p\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) = 0$. In this way the adaptive weights can be simplified, which is specifically as:

$$w_{ij} = -\log\left(q\left(x_i, x_j\right)\right) \tag{15}$$

In this paper, $q\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right)$ is defined as follows:

$$q\left(\boldsymbol{x_i}, \boldsymbol{x_j}\right) = \frac{f^n\left(\boldsymbol{s_{ij}}\right)}{\sum_{g=1}^{n}\sum_{j=1}^{n} f^n\left(\boldsymbol{s_{gj}}\right)} \tag{16}$$

It should be noted that the hyper-parameter n should be even, so that $f^n\left(\boldsymbol{s_{ij}}\right) > 0$ can be guaranteed. Combining Formula (15) with Formula (13), the adaptive weight similarity loss function can be obtained:

$$L_a = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(s_{ij} - m_{ij}\right)^2 \tag{17}$$

The ultimate goal of the loss function is to make the binary codes of similar RS images as close as possible. We use k-dimension approximate hash codes for calculating similar loss components, so we also need to introduce quantified loss to reduce quantization errors, get more efficient binary hash codes, and improve retrieval accuracy. The final loss function is given by:

$$\begin{aligned} L &= L_a + \lambda L_q \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(s_{ij} - m_{ij}\right)^2 + \lambda\left\|B - H\right\|_F^2 \end{aligned} \tag{18}$$

where B is the binary hash code of the model output. λ is a super-parameter used to adjust for overall loss.

EXPERIMENT AND ANALYSIS

To verify the effectiveness of the proposed method, experiments were conducted on two public RS image datasets, including the evaluation comparison of experimental metrics, ablation experiments, and parameter sensitivity analysis.

Datasets and Settings

We evaluated two public benchmark RS image datasets, EuroSAT (Helber et al., 2019) and PatternNet (Zhou et al., 2018), with dataset partitioning using the setting of Sun et al. (2021), as follows:

- **EuroSAT dataset:** Collected 27,000 RS images, including 10 scene categories in RS images. Each scene contains 2,000 to 3,000 images, and the size of each image was 64 × 64. Parts of this dataset are shown in Figure 3. Our experiments randomly selected samples of 100 images in each class as a test query set. The remaining 26,000 images are used as the training set.
- **PatternNet dataset:** Contains 38 different categories, each with 800 images, and each image has a resolution of 256 × 256, for a total of 30,400 images. Each category of images for this dataset is presented in Figure 4. The division is also done in such a way that 100 images in each category are used as the test query set. The remaining 26,600 images are used as the training set.

In the comparison experiments, the authors selected seven unsupervised hashing methods for comparison, including three classical shallow methods, LSH (Gionis et al., 1999), SH (Weiss et al., 2008), and ITQ (Gong et al., 2012) and four advanced unsupervised depth hashing methods, SSDH (Yang et al., 2018), BGAN (Song et al., 2018), MLS³RDUH (Tu et al., 2020), SPL-UDH (Sun et al., 2021). The parameters and architecture in the comparison method are set according to the original paper. Due to the difficulty in reproducing some methods, the authors directly refer to the results of the original paper. Here the authors extract the features by pre-trained Swin Transformer and use the fully connected layer features as the input for the shallow hashing method.

In our DUWH implementation, we use the tiny version of Swin Transformer, and the experimental code is written using the Pytorch framework. The parameter n for the loss function section is set to 2, and λ is set to 10. The batch size is set to 128, the model learning rate is set to 2e-5. The parameter K_1 for constructing the initial semantic similarity matrix is set to 20 and K_2 is set to 30. The update of the similarity matrix is performed after the first round of training.

Evaluation Criterions

To ensure that the experimental results are more convincing, Mean Average Precision (MAP), Precision of the top N retrieved image (Precision@N), and Precision-Recall curve (PR curve) are

Figure 3. Parts of the EuroSAT dataset

AnnualCrop Forest Herbaceous Vegetation Highway Industrial

Pasture PermanentCrop Residential River SeaLake

Figure 4. Parts of the PatternNet dataset

used as evaluation indexes, where MAP is a commonly used standard for image retrieval. If R images are returned for a query, the AP calculation method is given by:

$$AP = \frac{1}{N} \sum_{r=1}^{R} P(r) \delta(r) \tag{19}$$

where N denotes the number of total queries, $P(r)$ represents the precision for the r th query, and $\delta(r)$ represents whether the r th query is related to the image, that is, whether it has a corresponding label of 1.

If it carries the corresponding label, let $\delta(r) = 1$; conversely, let $\delta(r) = 0$. The MAP is the average value after several AP calculations. According to the settings of Sun et al. (2021), the number of samples in each category in EuroSAT and PatternNet datasets was set respectively to 1900 and 700. Precision@N is defined as the precision of the top N instances retrieved. The larger the area under PR Curve, the better the overall retrieval performance.

Experimental Results

Table 1 shows the comparison results between the MAP values of the proposed method and the seven methods of LSH, SH, ITQ, SSDH, BGAN, MLS³RDUH, and SPL-UDH, using 16-bit, 24-bit, 32-bit, and 48-bit binary codes for comparison.

It can be found from the results that on the two data sets, even if the high-dimensional features extracted by the pre-trained model are used as input for the shallow-level hashing method, its performance is still not as good as the deep unsupervised hashing. Since the hashing method based on a shallow layer is different from the unsupervised hashing method based on depth, and the hash encoding process and

Table 1. Comparison results of MAP values of DUWH and other methods on EuroSAT and PatternNet datasets

Methods	EuroSAT				PatternNet			
	16-bit	24-bit	32-bit	48-bit	16-bit	24-bit	32-bit	48-bit
LSH	0.318	0.317	0.405	0.442	0.220	0.252	0.324	0.404
SH	0.436	0.473	0.474	0.484	0.406	0.489	0.543	0.589
ITQ	0.481	0.510	0.552	0.559	0.433	0.486	0.545	0.605
SSDH	0.647	0.652	0.668	0.675	0.568	0.631	0.656	0.689
BGAN	0.745	0.759	0.761	0.776	0.705	0.766	0.779	0.806
MLS³RDUH	0.767	0.796	0.800	0.808	0.714	0.815	0.858	0.880
SPL-UDH	0.799	0.830	0.837	0.844	0.875	0.898	0.905	0.925
DUWH	0.833	0.854	0.861	0.887	0.879	0.910	0.923	0.958

the feature extraction process are independent of each other, so these two stages cannot interact with each other. The proposed DUWH outperforms all baselines in binary codes with different lengths. On the EuroSAT dataset, compared with the SPL-UDH algorithm, the DUWH method in this paper achieves 3.4%, 2.4%, 2.4% and 4.3% increments at 16, 24, 32 and 48 bits, respectively. On the PatternNet dataset, the DUWH method also achieves 3.3% increment at 48 bits. The improvement of DUWH retrieval performance indicates that the semantic similarity matrix update module based on weighted similarity structure and the adaptive weight loss module is overall effective.

The PR Curve of the 48-bit hash code on the EuroSAT dataset is shown in Figure 5(a) and the PatternNet dataset is shown in Figure 5(b). The precision rate in the figure is the ratio of the retrieved image of that category to all images, and the recall rate is the ratio of the retrieved image to all the images in the database. From the figures, it can be seen that as the recall rate gradually increases, the precision of all models will decrease. However, while the recall rate grows, the precision of DUWH remains at a high level, and the area under the PR curve of DUWH is larger than that of the other methods. Figure 6 (a) and (b) represent the Precision@N graph under 48-bit hash codes of each method on two datasets. It can be seen from the figure that the Precision@N of DUWH is higher than all baseline methods on both data sets, which again confirms that the retrieval performance of DUWH is more superior than that of other methods.

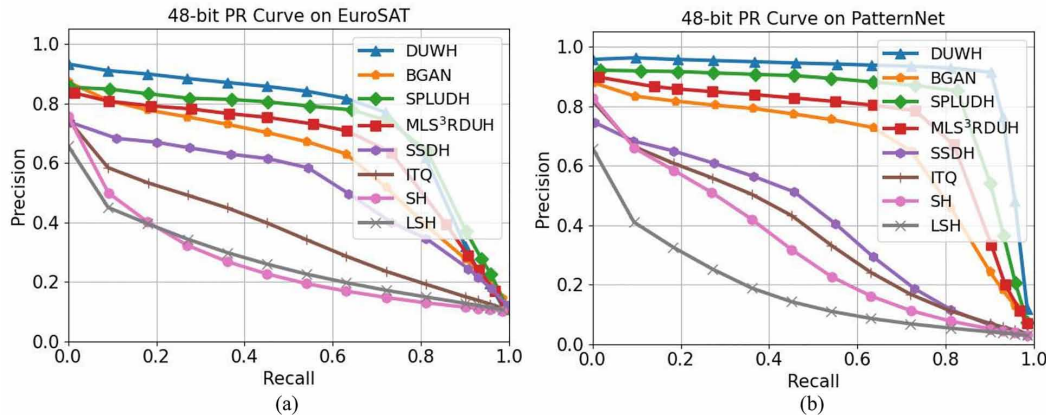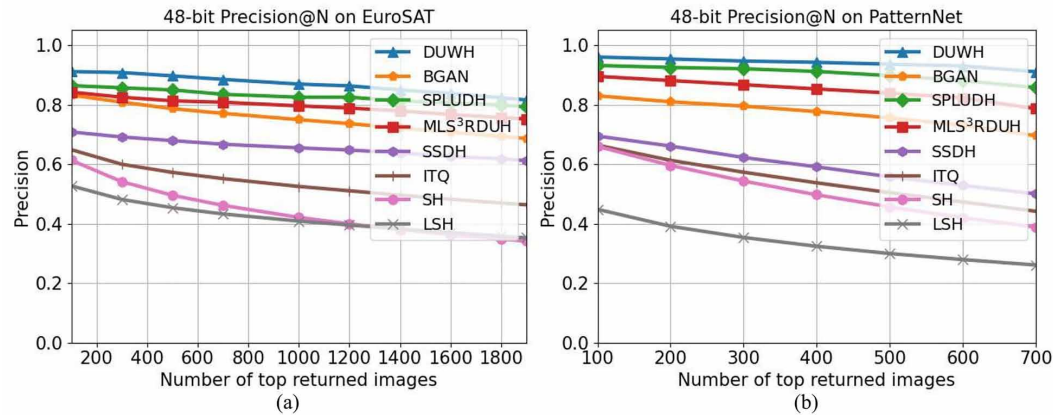Figure 5. PR curves on EuroSAT and PatternNet datasets

Figure 6. Precision@N curves on EuroSAT and PatternNet datasets

For each category of the EuroSAT dataset, the authors used the MAP values of DUWH and MLS³RDUH on 48-bit hash codes. As can be seen from Table 2, compared with MLS³RDUH, DUWH achieves a very large MAP improvement on River, Permanent Crop, and Highway. This indicates that the mechanism of adaptive weights effectively overcomes the problem of large intra-class variation and high inter-class similarity in RS image retrieval.

Visualization Analysis

To demonstrate the effectiveness of DUWH more visually, Figure 7 shows the top 10 retrieval examples returned by DUWH, MLS³RDUH on the EuroSAT dataset. An image marked with a slash means that the retrieved result is incorrect. Compared with MLS³RDUH, DUWH retrieved fewer incorrect images. During the retrieval process for the sea lake category, MLS³RDUH will retrieve some images of the river category incorrectly, while DUWH will not.

Table 2. Comparison results of MAP values of DUWH and MLS³RDUH on each category of EuroSAT dataset

Methods	DUWH	MLS³RDUH
Sea Lake	0.899	0.940
River	0.825	0.492
Residential	0.987	0.958
Annual Crop	0.901	0.917
Permanent Crop	0.820	0.690
Pasture	0.840	0.816
Herbaceous Vegetation	0.843	0.844
Forest	0.936	0.963
Industrial	0.981	0.950
Highway	0.838	0.510
All	0.887	0.808

Figure 7. Top 10 retrieved results of DUWH and MLS³RDUH on EuroSAT dataset
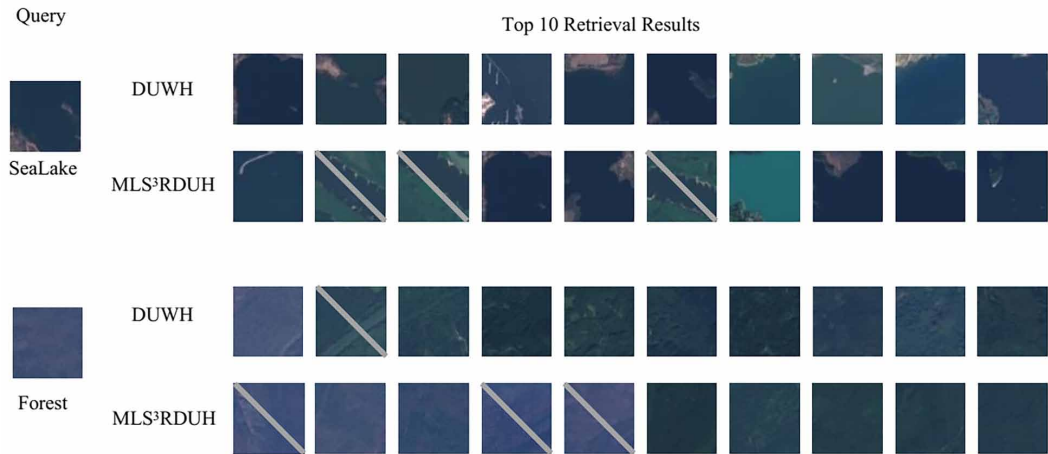
Ablation Study

In order to test the contribution of the proposed part to the obtained results, an ablation study was conducted. DUWH was divided into two experimental elements, as shown in Table 3: whether to include a similarity matrix update strategy and whether to add adaptive weights to the similarity loss.

Ablation experiments were performed on the EuroSAT dataset and the PatternNet dataset with a hash coding length of 48-bit. We remove the semantic similarity matrix update module as the first variant of the model DUWH-1. After generating the initial semantic similarity matrix as a pseudo-label, it is no longer updated during the training process. On the basis of DUWH-1, the adaptive weight similarity loss in the loss function is reduced to the ordinary similarity loss as the second variant of the model DUWH-2. Table 4 shows 48-bit MAP results on two datasets.

From the results, we can find the following two points:

1. Semantic similarity matrix update based on weighted similarity structure can improve retrieval performance. In particular, DUWH-1 decreases by 2.2% compared to DUWH when the similarity-preserving loss function loses adaptive weights on the A dataset.

Table 3. Two main components for ablation studies

Symbol	Implication
C1	Add the similarity matrix update
C2	Add adaptive weights (similarity loss)

Table 4. MAP of the method's variants for 48 bits on two datasets

Methods	C1	C2	EuroSAT	PatternNet
DUWH-2	✗	✗	0.844	0.946
DUWH-1	✗	✓	0.865	0.955
DUWH	✓	✓	0.887	0.958

2. The loss function adopts the combination of adaptive weight similarity loss and quantization loss, which can improve the hash learning ability. DUWH-1 is 2.1% and 0.9% higher than DUWH-2 on the EuroSAT dataset and the PatternNet dataset. Thus, both parts have a positive effect on improving the model performance, thus validating the effectiveness of the proposed method.

In the course of the ablation experiments, we found that the training time of the model grows a lot when the semantic similarity matrix update module C1 is added. Since the similarity matrix is updated once during the model training, where the time to regenerate the similarity matrix from the feature representation takes up a relatively large amount of time, there is still room for optimization in this part.

Parameter Sensitivity

Figure 8 shows the influence of hyper-parameters λ and n on the experimental results on two data sets when the hash code length is 48 bits. When another hyper parameter is adjusted, the other hyper parameter remains unchanged. As seen in Figure 8(a), the overall trend of MAP increases and then decreases as λ increases and performs best at λ of 10. Figure 8(b) shows that after $n > 2$, the retrieval accuracy begins to decline. According to the experimental results on the EuroSAT dataset and the PatternNet dataset, DUWH has the best stability at $\lambda = 10$ and $n = 2$.

We also analyzed the selection of the two K-NN parameters used to generate the similarity matrix. Figure 9 shows the MAP results for different values of K_1 and K_2 when the hash code length is 48 bits. The experimental results show that the different parameter selections have a slight effect on the final MAP results. DUWH retrieves the best accuracy on EuroSAT dataset and PatternNet dataset when $K_1 = 20$ and $K_2 = 30$.

CONCLUSION

In this work, the authors propose a novel unsupervised hashing method on CBRSIR tasks, called DUWH. DUWH uses a pre-trained Swin Transformer model to obtain the feature representation of the image. The similarity matrix is regenerated during model training, and the new matrix is updated with a weighted similarity structure to better capture the structure of deep semantic RS images. Meanwhile, DUWH introduces adaptive weight-based similarity loss and quantization loss to construct a new weighted loss function, which can more fully utilize the discriminative information of

Figure 8. Effect of hyperparameters on 48 bits hash code length experiments for two datasets
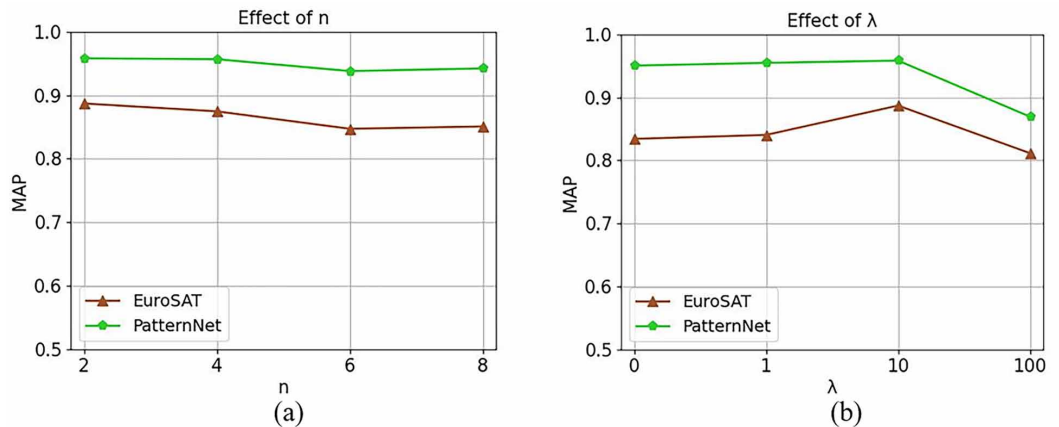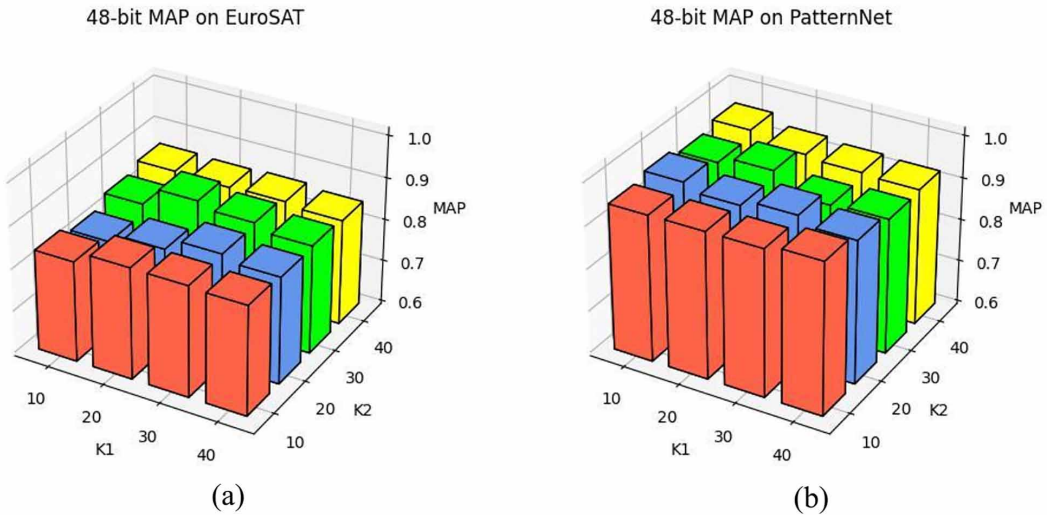
Figure 9. Effect of hyperparameters on 48 bits hash code length experiments for two datasets

sample pairs with different degrees of difficulty to learn to obtain the best hash codes. In the future, the authors plan to optimize the updating strategy of the similarity matrix to shorten the generation time of the new similarity matrix, and try new pseudo-label generation methods to make this work more efficient and accurate.

ACKNOWLEDGMENT

REFERENCES

Cao, R., Zhang, Q., Zhu, J., Li, Q., Li, Q., Liu, B., & Qiu, G. (2020). Enhancing remote sensing image retrieval using a triplet deep metric learning network. *International Journal of Remote Sensing*, *41*(2), 740–751. doi:10.1080/2150704X.2019.1647368

Chaudhuri, U., Banerjee, B., & Bhattacharya, A. (2019). Siamese graph convolutional network for content based remote sensing image retrieval. *Computer Vision and Image Understanding*, *184*, 22–30. doi:10.1016/j.cviu.2019.04.004

Chen, X., & Lu, C. (2020). An end-to-end adversarial hashing method for unsupervised multispectral remote sensing image retrieval. In *2020 IEEE International Conference on Image Processing* (pp. 1536-1540). IEEE. doi:10.1109/ICIP40778.2020.9190949

Chen, Y., Lu, X., & Wang, S. (2020). Deep cross-modal image–voice retrieval in remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(10), 7049–7061. doi:10.1109/TGRS.2020.2979273

Demir, B., & Bruzzone, L. (2014). A novel active learning method in relevance feedback for content-based remote sensing image retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(5), 2323–2334. doi:10.1109/TGRS.2014.2358804

Deng, C., Yang, E., Liu, T., Li, J., Liu, W., & Tao, D. (2019). Unsupervised semantic-preserving adversarial hashing for image search. *IEEE Transactions on Image Processing*, *28*(8), 4032–4044. doi:10.1109/TIP.2019.2903661 PMID:30872226

Dong, X., Liu, L., Zhu, L., Cheng, Z., & Zhang, H. (2020). Unsupervised deep k-means hashing for efficient image retrieval and clustering. *IEEE Transactions on Circuits and Systems for Video Technology*. Advance online publication. doi:10.1109/TCSVT.2020.3035775

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Gelly, S. (2020). *An image is worth 16x16 words: Transformers for image recognition at scale*. arXiv.

El-Nouby, A., Neverova, N., Laptev, I., & Jégou, H. (2021). *Training vision transformers for image retrieval*. arXiv.

Fan, L., Zhao, H., & Zhao, H. (2020). Global optimization: Combining local loss with result ranking loss in remote sensing image retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(8), 7011–7026. doi:10.1109/TGRS.2020.3029334

Gionis, A., Indyk, P., & Motwani, R. (1999). Similarity search in high dimensions via hashing. In *Proceedings of the 25th International Conference on Very Large Data Bases* (pp. 518-529). ACM. doi:10.5555/645925.671516

Gong, Y., Lazebnik, S., Gordo, A., & Perronnin, F. (2012). Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(12), 2916–2929. doi:10.1109/TPAMI.2012.193 PMID:24136430

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., & Cai, J. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. doi:10.1016/j.patcog.2017.10.013

Helber, P., Bischke, B., Dengel, A., & Borth, D. (2019). Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(7), 2217–2226. doi:10.1109/JSTARS.2019.2918242

Hu, Q., Wu, J., Cheng, J., Wu, L., & Lu, H. (2017). Pseudo label based unsupervised deep discriminative hashing for image retrieval. In *Proceedings of the 25th ACM International Conference on Multimedia* (pp. 1584-1590). ACM. doi:10.1145/3123266.3123403

Indyk, P., & Motwani, R. (1998). Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the Thirtieth Annual Acm Symposium On Theory Of Computing* (pp. 604-613). ACM. doi:10.1145/276698.276876

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2021). *Transformers in vision: A survey*. arXiv.

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221–232. doi:10.1007/s13748-016-0094-0

Li, P., Han, L., Tao, X., Zhang, X., Grecos, C., Plaza, A., & Ren, P. (2020). Hashing nets for hashing: A quantized deep learning to hash framework for remote sensing image retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(10), 7331–7345. doi:10.1109/TGRS.2020.2981997

Li, Y., Ma, J., & Zhang, Y. (2021). Image retrieval from remote sensing big data: A survey. *Information Fusion*, *67*, 94–115. doi:10.1016/j.inffus.2020.10.008

Li, Y., Zhang, Y., Tao, C., & Zhu, H. (2016). Content-based high-resolution remote sensing image retrieval via unsupervised feature learning and collaborative affinity metric fusion. *Remote Sensing*, *8*(9), 709. doi:10.3390/rs8090709

Lin, G., Shen, C., Shi, Q., Van den Hengel, A., & Suter, D. (2014). Fast supervised hashing with decision trees for high-dimensional data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1963-1970). IEEE. doi:10.1109/CVPR.2014.253

Lin, Q., Chen, X., Zhang, Q., Cai, S., Zhao, W., & Wang, H. (2022). *Deep unsupervised hashing with latent semantic components.* arXiv.

Lin, Q., Chen, X., Zhang, Q., Tian, S., & Chen, Y. (2021). Deep self-adaptive hashing for image retrieval. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 1028-1037). ACM. doi:10.1145/3459637.3482247

Liu, B., Cao, Y., Long, M., Wang, J., & Wang, J. (2018). Deep triplet quantization. In *Proceedings of the 26th ACM International Conference on Multimedia* (pp. 755-763). ACM. doi:10.1145/3240508.3240516

Liu, P., Gou, G., Shan, X., Tao, D., & Zhou, Q. (2020). Global optimal structured embedding learning for remote sensing image retrieval. *Sensors (Basel)*, *20*(1), 291. doi:10.3390/s20010291 PMID:31948002

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., & Guo, B. (2021). *Swin transformer: Hierarchical vision transformer using shifted windows.* arXiv.

Lu, G., & Man, Q. (2013). The image retrieval method of forest pest longicorn. *Forest Engineering*, *29*(3), 71–75. doi:10.16270/j.cnki.slgc.2013.03.021

Raginsky, M., & Lazebnik, S. (2009). Locality-sensitive binary codes from shift-invariant kernels. *Advances in Neural Information Processing Systems*, *22*, 1509–1517. doi:10.5555/2984093.2984263

Roy, S., Sangineto, E., Demir, B., & Sebe, N. (2018). Deep metric and hash-code learning for content-based retrieval of remote sensing images. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (p. 4539-4542). IEEE. doi:10.1109/IGARSS.2018.8518381

Shen, F., Shen, C., Liu, W., & Shen, T., H. (2015). Supervised discrete hashing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 37-45). IEEE. doi:10.1109/CVPR.2015.7298598

Shen, F., Xu, Y., Liu, L., Yang, Y., Huang, Z., & Shen, H. T. (2018). Unsupervised deep hashing with similarity-adaptive and discrete optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(12), 3034–3044. doi:10.1109/TPAMI.2018.2789887 PMID:29993420

Song, H., & Yang, W. (2022). GSCCTL: A general semi-supervised scene classification method for remote sensing images based on clustering and transfer learning. *International Journal of Remote Sensing*, 1–25. doi:10.1080/01431161.2021.2019851

Song, J., Gao, L., Yan, Y., Zhang, D., & Sebe, N. (2015). Supervised hashing with pseudo labels for scalable multimedia retrieval. In *Proceedings of the 23rd ACM International Conference on Multimedia* (pp. 827-830). ACM. doi:10.1145/2733373.2806341

Song, J., He, T., Gao, L., Xu, X., Hanjalic, A., & Shen, H. T. (2018). Binary generative adversarial networks for image retrieval. *Thirty-second AAAI Conference on Artificial Intelligence*. doi:10.5555/3504035.3504084

Sun, Y., Ye, Y., Li, X., Feng, S., Zhang, B., Kang, J., & Dai, K. (2021). Unsupervised deep hashing through learning soft pseudo label for remote sensing image retrieval. *Knowledge-Based Systems*, *107807*. Advance online publication. doi:10.1016/j.knosys.2021.107807

Tang, X., Zhang, X., Liu, F., & Jiao, L. (2018). Unsupervised deep feature learning for remote sensing image retrieval. *Remote Sensing*, *10*(8), 1243. doi:10.3390/rs10081243

Tu, R.-C., Mao, X., & Wei, W. (2020). MLS3RDUH: Deep unsupervised hashing via manifold based local semantic similarity structure reconstructing. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence* (pp. 3466-3472). doi:10.24963/ijcai.2020/479

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008). Academic Press.

Wang, J., Liu, W., Sun, A. X., & Jiang, Y.-G. (2013). Learning hash codes with listwise supervision. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3032-3039). IEEE. doi:10.1109/ICCV.2013.377

Weiss, Y., Torralba, A., & Fergus, R. (2008). Spectral hashing. In *Proceedings of the 21st International Conference on Neural Information Processing Systems* (pp. 1753–1760). doi:10.5555/2981780.2981999

Wu, D., Dai, Q., Liu, J., Li, B., & Wang, W. (2019). Deep incremental hashing network for efficient image retrieval. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9069-9077). IEEE. doi:10.1109/CVPR.2019.00928

Xia, R., Pan, Y., Lai, H., Liu, C., & Yan, S. (2014). Supervised hashing for image retrieval via image representation learning. *Twenty-eighth AAAI Conference on Artificial Intelligence*.

Xiong, W., Lv, Y., Cui, Y., Zhang, X., & Gu, X. (2019). A discriminative feature learning approach for remote sensing image retrieval. *Remote Sensing*, *11*(3), 281. doi:10.3390/rs11030281

Yang, E., Deng, C., Liu, T., Liu, W., & Tao, D. (2018). Semantic structure-based unsupervised deep hashing. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (pp. 1064-1070). Academic Press.

Zhang, W., Wu, D., Zhou, Y., Li, B., Wang, W., & Meng, D. (2020). Deep Unsupervised Hybrid-similarity Hadamard Hashing. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 3274-3282). ACM. doi:10.1145/3394171.3414028

Zhou, W., Newsam, S., Li, C., & Shao, Z. (2018). PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, *145*, 197–209. doi:10.1016/j.isprsjprs.2018.01.004

Weipeng Jing received the Ph.D. degree from the Harbin Institute of Technology, China. He is currently a Professor with Northeast Forestry University, China. He has published over 100 research articles in refereed journals and conference proceedings, such as IEEE Transactions on Geoscience and Remote Sensing, IEEE Geoscience and Remote Sensing Letters, and IEEE Sensors Journal etc. Besides, he served as an Publication Chair of ICPCSEE (2020, 2019, 2018, 2017, 2016), BDTA (2017, 2015), Collaborate 2017, Wicon 2016, Ficloud 2016, etc. He is now a member of the ACM, IEEE, and a member of the China Computer Federation (CCF). His research interests include modeling and scheduling for distributed computing systems, Artificial intelligence, remote sensing image intelligent mining, spatial data mining.

Zekun Xu received the M.S. degree from Northeastern Forestry University, China in 2020. He is currently pursuing his PhD at Northeastern Forestry University. His research interests include computer vision and data mining.

Linhui Li was born in Hulunbeier, Inner Mongolia Autonomous Region, China, in 1979.She received the doctor's degree in Forestry Information Engineering from Northeast Forestry University, Harbin. Her research interests include data mining, large-scale data storage, remote sensing image processing and application based on computer technology.

Jian Wang was born in 1976. He is a senior engineer in Aerospace Information Research Institute, Chinese Academy of Sciences. He received his Ph.D. degree from Institute of Software, Chinese Academy of Sciences. His research areas are high performance geographic computing, parallel computing, remote sensing big data infrastructure, etc.

Yue He received the B.S. degree from Dalian Minzu University, China in 2019. She is currently pursuing her master's degree at Northeastern Forestry University. Her research interests include cloud computing and parallel computing.

Guangsheng Chen is currently doctoral supervisor and professor in Northeast Forestry University, China. He is the member of National Innovation Methods Research Institute, and executive director of Education Information Technology Council of Education Ministry. His research interests include biomass material prediction, intelligent detection of new composite materials and big data on forestry. He has published over 30 academic papers and one monograph.