A Discriminative Feature Learning Approach With Distinguishable Distance Metrics for Remote Sensing Image Classification and Retrieval

Zhiqi Zhang, Wen Lu, Xiaoxiao Feng, Jinshan Cao, and Guangqi Xie

***Abstract*—The fast data acquisition rate due to the shorter revisit periods and wider observation coverage of satellites results in large amounts of remote sensing images every day. This brings the challenge of how to accurately search the images with similar visual content as the query image. Content-based image retrieval (CBIR) is a solution to this challenge, its performance heavily depends on the effectiveness of the image representation features and similarity evaluation metrics. Ideal image feature representations have dispersed interclass, compact intraclass distribution. However, the neural networks employed by many CBIR methods are trained with cross entropy loss, which does not directly optimize the metrics that evaluates interclass variance over intraclass variance, hence, their feature representations are suboptimal. Meanwhile, the traditional distance metrics used by many CBIR methods cannot index the similarity of feature representations well in high-dimensional space. For better CBIR performance, we propose a discriminative feature learning approach with distinguishable distance metrics for remote sensing image classification and retrieval. By balancing the diagonal elements and nondiagonal elements of the within-class scatter matrix of deep linear discriminant analysis, our proposed loss function, balanced deep linear discriminant analysis, can better optimize the Rayleigh–Ritz quotient, which measures interclass variance over intraclass variance. In addition, the proposed distance metrics, reciprocal exponential distance (RED), is more capable of maintaining distance contrast in high dimensionality, therefore, it can better index similarity for feature representations in high dimensionality. Both visual interpretations and quantitative metrics of extensive experiments demonstrated the effectiveness of our approach.**

***Index Terms*—Content-based image retrieval (CBIR), convolutional neural network, distance metrics, feature representation, linear discriminant analysis (LDA), remote sensing.**

Manuscript received 8 August 2022; revised 28 September 2022; accepted 27 December 2022. Date of publication 30 December 2022; date of current version 5 January 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61901307, in part by the Open Research Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University under Grant 20E01, and in part by the Scientific Research Foundation for Doctoral Program of Hubei University of Technology under Grant BSQD2020054 and Grant BSQD2020055. *(Zhiqi Zhang and Wen Lu contributed equally to this work.) (Corresponding author: Xiaoxiao Feng.)*

Zhiqi Zhang is with the School of Computer Science, Hubei University of Technology, Wuhan 430068, China, and also with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zzq540@hbut.edu.cn).

Wen Lu, Xiaoxiao Feng, Jinshan Cao, and Guangqi Xie are with the School of Computer Science, Hubei University of Technology, Wuhan 430068, China (e-mail: wenlu@hbut.edu.cn; 20220026@hbut.edu.cn; caojs@hbut.edu.cn; 20221074@hbut.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3233032

I. Introduction

NOWADAYS, a global earth observation system has been formed, which can quickly obtain a large amount of high spatial resolution, high temporal resolution, and high spectral resolution remote sensing imagery [1]. Image retrieval is the task that collects relevant images for a query from the imagery repository, it can prepare the auxiliary data or narrow the search space for many remote sensing images processing tasks, such as image matching, image registration, and image fusion [2], [3], [4]. Accurate retrieval of the interested images from the remote sensing imagery repository can reduce the boring workload spent on manual searching and filtering a large number of images. As the ground stations receive the remote sensing images, after initial preprocessing, they are stored in a dedicated database [5]. Traditional image retrieval methods search images based on the metadata that describes the images, and the keywords were manually input in advance. However, the high complexity of remote sensing images cannot be described easily by keywords, and a higher and higher volume of remote sensing images makes manual annotation more and more time-consuming and costly. Content-based image retrieval (CBIR) methods have been introduced to overcome the shortcomings of the traditional database query techniques, they retrieve the desired images automatically by their visual content similarities to the query image. CBIR consists of the following three steps: generation of image feature representations, similarity calculation of feature representations, and similarity ranking of feature representations.

Effective image feature representations play a decisive role in the CBIR performance, therefore, how to generate more discriminative feature representations is a research focus [6]. According to the abstract level, the feature representations are classified as the low-level, the middle-level, and the high-level. The low-level feature representations are designed by domain experts and are built by mining the spectral, texture, or shape cues of remote sensing imagery [7], [8]. By embedding the low-level hand-crafted feature descriptors into representative visual vocabulary space, the middle-level feature representations are more invariant to appearance differences caused by changes of scale, rotation, or illumination [9], [10]. Since neural networks can automatically learn strong feature extraction ability by back-propagation, its intermediate layer outputs are employed as high-level features, which include semantic information.

The high-level features usually show overwhelming superiority over the low-level and the middle-level features [11].

Neural networks can generate high-level features through unsupervised, supervised, and transfer learning methods. As an unsupervised learning method, the autoencoder attempts to reconstruct input images through encoder layers, an embedding representation layer, and decoder layers [12], [13]. Although the autoencoder saves the effort to label the images, its feature representations are less powerful than CNNs trained by supervised methods. Although the image classification and retrieval tasks differ in terms of their goals, the classification labels can guide the CNN layers to learn feature representations targeted to the labels. As transfer learning methods, some CNNs pretrained on image classification datasets of other domains, such as ImageNet [14], can also generate meaningful feature representations, but due to the large difference between remote sensing images and general images, their feature representations are less effective than those generated by the CNNs trained with the target domain remote sensing image datasets.

Some supervised learning methods rely on specially designed loss functions for more discriminative feature representations. Xiong et al. [15] utilized the center loss function to penalize the distances between the feature representations and their corresponding class centers [6]. Cao et al. [16] constructed a Triplet network with Euclidean distance metrics learning objective function to extract the representative features in a semantic space in which images from the same class are close to each other while those from different classes are far apart. For more distinguishable feature representations, Liu et al. [17] used a softmax function rather than a hinge function to deal with the limitation of triplet loss and presented a novel optimal structured loss to force the positive pairs within a limitation and push the negative ones far away from a given boundary. Liu et al. [18] proposed a center-metrics learning method and coupled it with positive–negative center loss to deal with within-class variations. The purpose of all these loss functions is to produce more discriminative feature representations with low intraclass variance and high interclass variance, however, their methods do not directly optimize metrics that evaluate interclass variance over intraclass variance, therefore, are not efficient approaches.

As a supervised dimensionality reduction and classification method, linear discriminant analysis (LDA) projects the input data into a low-dimensional subspace and finds the optimal linearly separable boundaries between classes in that subspace by maximizing the trace of between-class scatter matrix and minimizing the trace of within-class scatter matrix. Since LDA is a linear transformation, it was born not good at processing data with the nonlinear distribution. Inspired by Kernel PCA, nonlinear kernel methods are also introduced to LDA [19], but those approaches are either limited by some fixed nonlinear transformations or have very complicated computation [20]. CNN has been widely recognized as a successful nonlinear representation learner, to utilize the such talent of CNN, Dorfer et al. put LDA on top of CNN to form deep linear discriminant analysis (DLDA) [21]. In contrast with the abovementioned methods, DLDA directly optimizes the generalized Rayleigh–Ritz quotient that measures interclass variance over intraclass variance. Instead of maximizing the likelihood of target labels as the conventional cross entropy loss, the eigenvalues along the discriminant eigenvector directions are maximized. By focusing on directions in the latent space with the smallest discriminative power, DLDA learns linearly separable hidden representations with similar discriminative power in all directions of the latent space. Although the CNN trained by DLDA outperforms that trained by cross entropy, we find the proportion of nondiagonal elements to diagonal elements in the within-class scatter matrix is too large that the separative capacity of individual dimensions is suboptimal. For higher image classification accuracy and better feature representations, we propose a regularization method on within-class scatter matrix to strengthen the discriminative ability of each dimension and name it balanced deep linear discriminant analysis (BDLDA).

Similarity ranking of feature representations also plays a vital role in image retrieval accuracy. The dimensions of the feature representations output by the CNN intermediate layer usually range from several hundred to several thousand. For calculating the similarity of two feature representations in such high-dimensional feature embedding space, many CBIR methods employ the Euclidean distance or the Manhattan distance metrics as a natural extension of their traditional use in two or three-dimensional spatial applications. However, the performance of similarity indexing in high dimensions degrades rapidly, because, for a given point in high-dimensional space, the distance ratio of its nearest neighbor to its farthest neighbor is almost 1 for a wide variety of data distributions and distance functions. In such a case, the nearest neighbor problem becomes meaningless, since the contrast between the distances to different data points does not exist [22]. To relieve the detrimental effects of the high dimensionality curse, we present a new distance metrics named reciprocal exponential distance (RED), which continues to be contrasting with increasing dimensionality.

In summary, the feature representations generated by neural networks trained by conventional loss functions, for example, the cross entropy, are not so discriminative, hence restricting the performance of CBIR methods. Meanwhile, the Euclidean distance metrics and the Manhattan distance metrics used by many image retrieval methods cannot index the similarity of feature representations well in high-dimensional space. Toward these two improvement directions, we propose a discriminative feature learning approach with distinguishable distance metrics for remote sensing image classification and retrieval. The contributions of this article can be summarized in the following two aspects.

1) By balancing the diagonal elements and nondiagonal elements of the within-class scatter matrix of DLDA, our proposed loss function, BDLDA, can better optimize the Rayleigh–Ritz quotient, which measures interclass variance over intraclass variance. Therefore, the CNN trained by BDLDA has higher image classification accuracy and can generate more discriminative image feature representations.
2) Our proposed distance metrics, RED, is more capable of maintaining distance contrast in high dimensionality,

therefore, it can better index similarity for feature representations in high dimensionality.

II. Related Work

In this section, we first revisit the basic ideas of LDA and then introduce DLDA.

A. Linear Discriminant Analysis

As a supervised dimensionality reduction and classification method, LDA projects the input data into a low-dimensional subspace and finds the optimal linearly separable boundaries between classes in that subspace. Denote $\mathbb{R}^d \ni \{\boldsymbol{x}_i^{(j)}\}_{i=1}^{n_j}$ are the instances of the jth class of which there are multiple number of classes [23]. Between-class scatter matrix $\boldsymbol{S_B}$ and within-class scatter matrix $\boldsymbol{S_W}$ are introduced to measure the effectiveness of separation in the lower dimensional subspace. The between-class scatter matrix is defined as

$$\mathbb{R}^{d\times d} \ni \boldsymbol{S_B} := \sum_{j=1}^{c} n_j(\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^\top \tag{1}$$

where c is the number of classes and

$$\mathbb{R}^d \ni \boldsymbol{\mu} := \frac{1}{\sum_{k=1}^{c} n_k}\sum_{j=1}^{c} n_j\boldsymbol{\mu}_j = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i \tag{2}$$

is the weighted mean of means of classes or the total mean of data. The within-class scatter matrix is defined as

$$\mathbb{R}^{d\times d} \ni \boldsymbol{S_W} := \sum_{j=1}^{c}\sum_{i=1}^{n_j}\left(\boldsymbol{x}_i^{(j)} - \boldsymbol{\mu}_j\right)\left(\boldsymbol{x}_i^{(j)} - \boldsymbol{\mu}_j\right)^\top \tag{3}$$

where n_j is the sample size of the jth class.

The target of LDA is to find the optimal linear projection directions, $\{\boldsymbol{u}_j\}_{j=1}^{p}$ where $\boldsymbol{u}_j \in \mathbb{R}^d$ that maximize the between-class scatter matrix $\boldsymbol{S_B}$ and minimize the within-class matrix $\boldsymbol{S_W}$, so the instances belong to different classes get far from each other while the instances of the same class get close to one another. The optimization problem is

$$\max_{\boldsymbol{U}} f(\boldsymbol{U}) := \frac{\mathbf{tr}\left(\boldsymbol{U}^\top \boldsymbol{S_B}\boldsymbol{U}\right)}{\mathbf{tr}\left(\boldsymbol{U}^\top \boldsymbol{S_W}\boldsymbol{U}\right)} \tag{4}$$

where $\mathbb{R}^{d\times p} \ni \boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p]$.

The solution to the generalized Rayleigh–Ritz quotient problem is

$$\boldsymbol{U} = \mathbf{eig}\left(\boldsymbol{S_W}^{-1}\boldsymbol{S_B}\right) \tag{5}$$

where $\mathbf{eig}(\ )$ denotes the eigenvectors of the matrix that are stacked column-wise.

$\boldsymbol{S_W}$ might be singular and not invertible, especially in small sample size problems, where the number of available training samples is smaller than the dimensionality of the sample space. Meanwhile, the estimation of $\boldsymbol{S_W}$ overemphasizes high eigenvalues whereas small eigenvalues are underestimated. To make $\boldsymbol{S_W}$ invertible and correct the bias, a very small positive number λ is added to the diagonal of $\boldsymbol{S_W}$ [24], [25], [26]. In this case, the solution is

$$\boldsymbol{U} = \mathbf{eig}\left((\boldsymbol{S_W} + \lambda\boldsymbol{I})^{-1}\boldsymbol{S_B}\right). \tag{6}$$

B. Deep Linear Discriminant Analysis

Since each eigenvalue v_i is the quantitative measurement of the discrimination ability in the direction of the corresponding eigenvector $\boldsymbol{u}_i$, as a nonlinear extension of the classic LDA, DLDA replaces the cross entropy loss function with eigenvalue-based loss function to encourage CNN to learn feature representations with discriminative distribution parameters [21]. To avoid CNN focusing on maximizing the eigenvalues of which the eigenvectors already separate the classes and ignoring the eigenvalues whose eigenvectors poorly discriminate the classes, the optimization object of CNN is limited to the smallest of all $c-1$ valid eigenvalues so that all of the $c-1$ latent feature dimensions would be trained on balance. The target of DLDA is to find the optimal model parameters Θ of CNN such that the produced latent features have low intraclass variance and high interclass variance

$$\arg\max_{\Theta} \frac{1}{k}\sum_{i=1}^{k} v_i$$
$$\text{with } \{v_1, \ldots, v_k\} = \{v_j | v_j < \min\{v_1, \ldots, v_{c-1}\} + \epsilon\}. \tag{7}$$

Dorfer et al. [21] relied on the distances to the linear decision hyperplanes for classification [27]. Assume $\boldsymbol{e} = \{\boldsymbol{e}_i\}_{i=1}^{c-1}$ are the corresponding eigenvectors, based on the LDA projection matrix $\boldsymbol{A} := [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{c-1}]$ and the per-class mean hidden representations $\bar{\boldsymbol{H}}_c := [\bar{\boldsymbol{h}}_1^\top, \ldots, \bar{\boldsymbol{h}}_c^\top]$, the distances of test sample hidden representation $\boldsymbol{h}_t$ to the linear decision hyperplanes are defined as

$$\boldsymbol{d} = \boldsymbol{h}_t^\top \boldsymbol{T}^\top - \frac{1}{2}\mathbf{diag}\left(\bar{\boldsymbol{H}}_c\boldsymbol{T}^\top\right) \quad \text{with} \quad \boldsymbol{T} = \bar{\boldsymbol{H}}_c\boldsymbol{A}\boldsymbol{A}^\top \tag{8}$$

where $\boldsymbol{T}$ are the decision hyperplane normal vectors. The class probabilities vector is $\boldsymbol{p}'_c = 1/(1+e^{-\boldsymbol{d}})$ and is further normalized by $\boldsymbol{p}_c = \boldsymbol{p}'_c / \sum p'_i$. So the class with the largest probability is the prediction.

III. Proposed Method

If one observes the sample distribution on each individual output dimension of the CNN trained by cross entropy, he would find each individual output dimension only differentiates one class, leaving other classes clustered together in that dimension. In contrast, if one observes the sample distribution on each individual output dimension of the CNN trained by DLDA, he would find each individual output dimension differentiates several classes, as illustrated in Fig. 5. Although Dorfer et al. [21] showed DLDA outperforms cross entropy as a loss function for image classification task, we find the interclass variance in Fig. 5 is still not large enough and the intraclass variance is still not small enough, which causes some samples of different classes overlapping. To strengthen the discriminative ability of each individual output dimension as well as keep them complement each other, we propose a new loss function named BDLDA, which makes output dimensions have more dispersed interclass,

more compact intraclass distribution. And then we present the novel RED for better similarity indexing of feature representations in high-dimensional space.

A. Balanced Deep Linear Discriminant Analysis

Each diagonal element of the within-class scatter matrix $\boldsymbol{S_W}$ is the variance of a single output dimension, which represents the internal information within that dimension; while each nondiagonal element is the covariance of a pair of dimensions, which reflects pairwise mutual relationship. In $\boldsymbol{S_W}$, the number of diagonal elements is much fewer than that of nondiagonal elements, which causes the internal information overwhelmed by the pairwise mutual relationship. In other words, a small portion of diagonal elements and a large portion of nondiagonal elements make the CNN trained by DLDA focus on training the cooperative discriminative ability of all the dimensions, while putting less emphasis on training the separative capacity of individual dimensions. To strengthen the discriminative ability of each dimension, we propose a new loss function that balances the diagonal elements and nondiagonal elements of the within-class scatter matrix $\boldsymbol{S_W}$.

In BDLDA, $\boldsymbol{S_W}$ is regularized by the following formula to weaken the influence of nondiagonal elements that reflect the dimensional pairwise mutual relationship

$$\boldsymbol{S'_W} = \alpha \boldsymbol{S_W} + (1-\alpha)\mathbf{diag}(\boldsymbol{S_W}) + \lambda \boldsymbol{I} \quad \text{with}\, 0 \le \alpha \le 1. \tag{9}$$

The operation $\mathbf{diag}(\boldsymbol{S_W})$ means leaving the diagonal elements of $\boldsymbol{S_W}$ unchanged and setting all the nondiagonal elements to zero.

The partial derivative of the objective function with respect to hidden layer $\boldsymbol{H}$ becomes

$$\frac{\partial}{\partial \boldsymbol{H}} \frac{1}{k} \sum_{i=1}^{k} v_i = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{e}_i^{\top} \left(\frac{\partial \boldsymbol{S_B}}{\partial \boldsymbol{H}} - v_i \frac{\partial \boldsymbol{S'_W}}{\partial \boldsymbol{H}} \right) \boldsymbol{e}_i. \tag{10}$$

Compared to the original $\boldsymbol{S_W}$, the diagonal elements of $\boldsymbol{S'_W}$ remain unchanged but all the nondiagonal elements are scaled down by a factor of α. Note that if $\alpha = 1$, $\boldsymbol{S'_W} = \boldsymbol{S_W}$; if $\alpha = 0$, all the nondiagonal elements of $\boldsymbol{S'_W}$ are zero, in this case, the regularized DLDA loss function will concentrate on training the separative capacity of every single dimension while completely ignoring the coordination among dimensions.

Overemphasize cooperation at the cost of suppressing individual discretion, in this case, $\alpha = 1$, or overemphasize individual judgment at the expense of losing coordination, in this case, $\alpha = 0$, might not be the best choice in regularizing $\boldsymbol{S_W}$.

To strike a balance between the dimension internal information and the dimensions pairwise mutual relationship, the proposed method BDLDA sets

$$\alpha = \frac{1}{c-1} \tag{11}$$

where c is the number of classes.

B. Reciprocal Exponential Distance

After the image feature representations are output by the middle layer of CNN, a distance metrics is needed for similarity calculation. Not contented with the existing distance metrics, we propose a novel distance metrics for measuring proximity in high-dimensional space, which is sensitive to the number of dimensions on which two data points are similar, and is more distinguishable for distances in high-dimensional space. RED between two data points $\boldsymbol{X} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{Y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ is defined as

$$\mathbf{RED}(\boldsymbol{X}, \boldsymbol{Y}) = \frac{n}{\sum_{i=1}^{n} e^{-|x_i - y_i|}} - 1. \tag{12}$$

The range of RED is the same as the Minkowski distance, namely $[0, +\infty)$.

As $\boldsymbol{X}$ approaches $\boldsymbol{Y}$ in all dimensions,

$$\lim_{\boldsymbol{X} \to \boldsymbol{Y}} \mathbf{RED}(\boldsymbol{X}, \boldsymbol{Y}) = 0.$$

As $\boldsymbol{X}$ departs $\boldsymbol{Y}$ in all dimensions

$$\lim_{\forall (x_i - y_i) \to \infty} \mathbf{RED}(\boldsymbol{X}, \boldsymbol{Y}) = +\infty.$$

Since the denominator of RED formula is the sum of the exponentials of the negative coordinates difference in each dimension, the dimensions with close coordinate values dominate the value of the function.

Define $d_i = |x_i - y_i|$

$$\frac{\partial \mathbf{RED}(\boldsymbol{X}, \boldsymbol{Y})}{\partial d_j} = \frac{n e^{-d_j}}{(\sum_{i=1}^{n} e^{-d_i})^2}. \tag{13}$$

Assume ϵ is a finite positive number, if $\exists i \neq j$: $d_i < \epsilon$

$$\lim_{d_j \to +\infty} \frac{\partial \mathbf{RED}(\boldsymbol{X}, \boldsymbol{Y})}{\partial d_j} = 0. \tag{14}$$

According to (14), for RED, the influence of a certain dimension on the fluctuation of the distance value fades with the increasing coordinates difference.

Define $d_i = |x_i - y_i|$, MD denotes Minkowski distance

$$\frac{\partial \mathbf{MD}(\boldsymbol{X}, \boldsymbol{Y})}{\partial d_j} = \left(\sum_{i=1}^{n} d_i^p \right)^{\left(\frac{1}{p} - 1\right)} d_j^{p-1}. \tag{15}$$

Assume ϵ is a finite positive number, if $\forall i \neq j$: $d_i < \epsilon$

$$\lim_{d_j \to +\infty} \frac{\partial \mathbf{MD}(\boldsymbol{X}, \boldsymbol{Y})}{\partial d_j} = 1. \tag{16}$$

On the contrary, according to (16), for the Euclidean distance and the Manhattan distance, the dimensions with large coordinates difference values consistently play significant roles in determining the distance value, while the influences of the dimensions with close coordinate values are neglect able.

Intuitively, two data points in high dimensionality are considered similar if their coordinates in most dimensions are close. Considering such two data points, their coordinates in a majority of dimensions are close, but in a minority of dimensions are far away. If measured by the traditional Minkowski distance metrics (the Euclidean distance and the Manhattan distance),

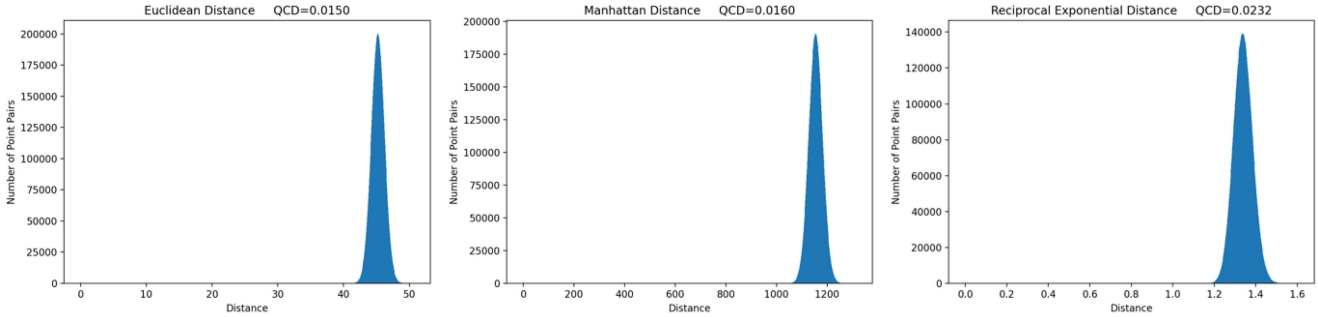

Fig. 1. Histograms of the pairwise distances calculated by the Euclidean distance, the Manhattan distance, and the RED, respectively.

their similarity value would be dominated by the minority of dimensions and, hence, be misjudged. However, different from the traditional Minkowski distance metrics that emphasize the divergent dimensions, RED judges proximity by focusing on similar dimensions, which accords with human intuition.

To verify RED can better maintain distance contrast in high dimensionality, we plotted the histograms of pairwise distances of different distance metrics and compared their dispersion. To simulate the similarity evaluation among the feature representations, we sampled 10 thousand 1024-dimensional data points from the standard normal distribution, and randomly formed 10 million pairs of data points. Then, we calculated the pairwise distances by the Euclidean distance, the Manhattan distance, and RED.

In statistics, the quartile coefficient of dispersion (QCD) is a measure of the dispersion of the distribution. It is defined as

$$\mathrm{QCD} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \tag{17}$$

where Q_1 and Q_3 denote the first and third quartiles of the dataset. We adopted this concept as the criterion for assessing the differentiating ability of distance metrics.

The histograms of the pairwise distances along with QCD of the three distance metrics are presented in Fig. 1. From the heights and shapes of the histograms as well as the QCD values, it can be seen that RED performs better than the other distance metrics in maintaining distance contrast.

IV. Experimental Results

In this section, we first introduce the image classification network used to generate feature representations of remote sensing images, then we present the image classification experimental results on two remote sensing image datasets to compare BDLDA with cross entropy and DLDA. At last, we present the image retrieval experimental results to validate the effectiveness of our proposed methods for remote sensing image retrieval tasks.

A. Experimental Network

We conducted experiments on an efficient lightweight image classification network MKANet-Class [28]. Aimed at the characteristics of top view remote sensing imagery, MKANet-Class was designed to better extract representation features of multi-scale ground objects. Its backbone consists of two initial convolutional layers and three multibranch Kernel-sharing Atrous convolution (MKA) modules, and its classification head consists of an average pooling layer and a linear layer.

TABLE I
Network Architecture of MKANet-Class

Stages	Output Size	Operation	Output Channels
Input Image	256×256		3
Stage 1	128×128	ConvS2	32
Stage 2	64×64	ConvS2	64
Stage 3	32×32	ConvS2	128
	32×32	MKA module	128
Stage 4	16×16	ConvS2	256
	16×16	MKA module	256
Stage 5	8×8	ConvS2	512
	8×8	MKA module	512
Classification Head	1	GAP	512
	1	Dropout	512
	1	Linear	number of classes

ConvS2: 3×3 convolution with stride 2, batch normalization, ReLU activation. GAP: Global Average Pooling.

As shown in Fig. 2, the MKA module includes the following three parts.

1) Multibranch kernel-sharing depthwise atrous convolutions with dilation rates (1, 2, 3).
2) Multibranch depthwise convolutions.
3) Concatenation and pointwise convolution.

The network architecture of MKANet-Class is illustrated in Fig. 3 and detailed in Table I, the 512-dimension feature vectors output by the GAP layer are used for feature representations of images in CBIR tasks.

B. Image Classification Experimental Results on RSSCN7

RSSCN7 [29] is a remote sensing image dataset (see Fig. 4), it contains wide diversity of scene images captured under changing seasons and varying weathers. It has seven typical scene classes of image size 400×400 pixels, for each class, there are 400 images sampled on four different scales with 100 images per scale. In this experiment, the total of 2800 images were split as a training set, validation set, and test set with the ratio of 3:3:4.

Training details: AdamW [30] was used as an optimizer with batch size 280, and the base learning rate was 0.001 with cosine decay. The number of epochs was 1000 with a warmup

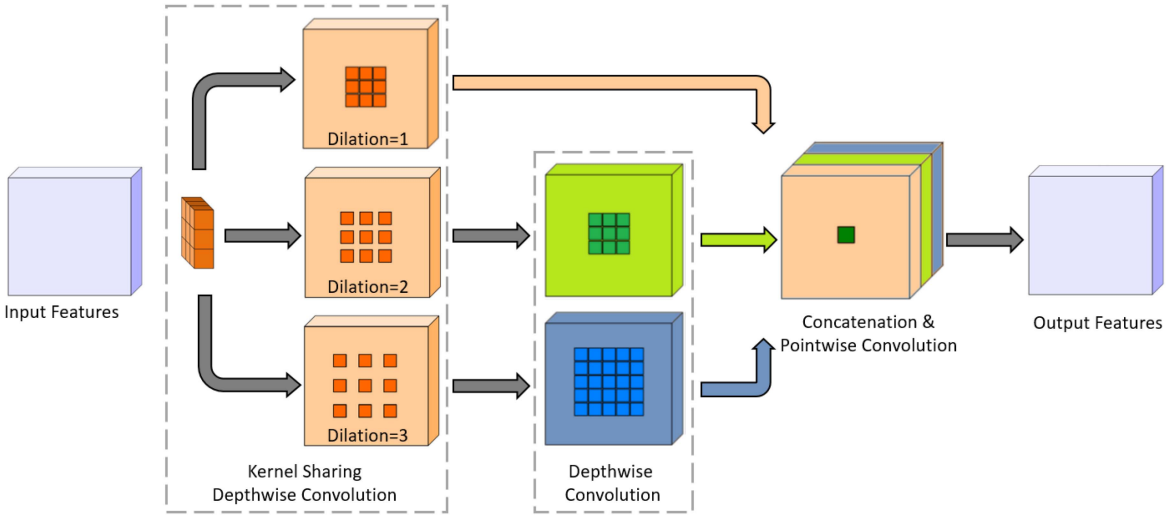

Fig. 2. Structure of the MKA module [28]. The orange square represents the kernel shared by three depthwise atrous convolutional branches with dilation rates (1, 2, 3).

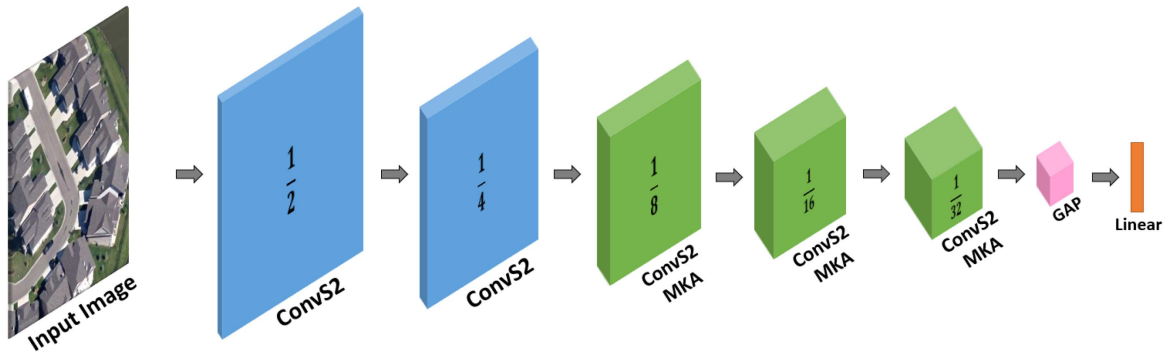

Fig. 3. Network architecture of MKANet-Class.

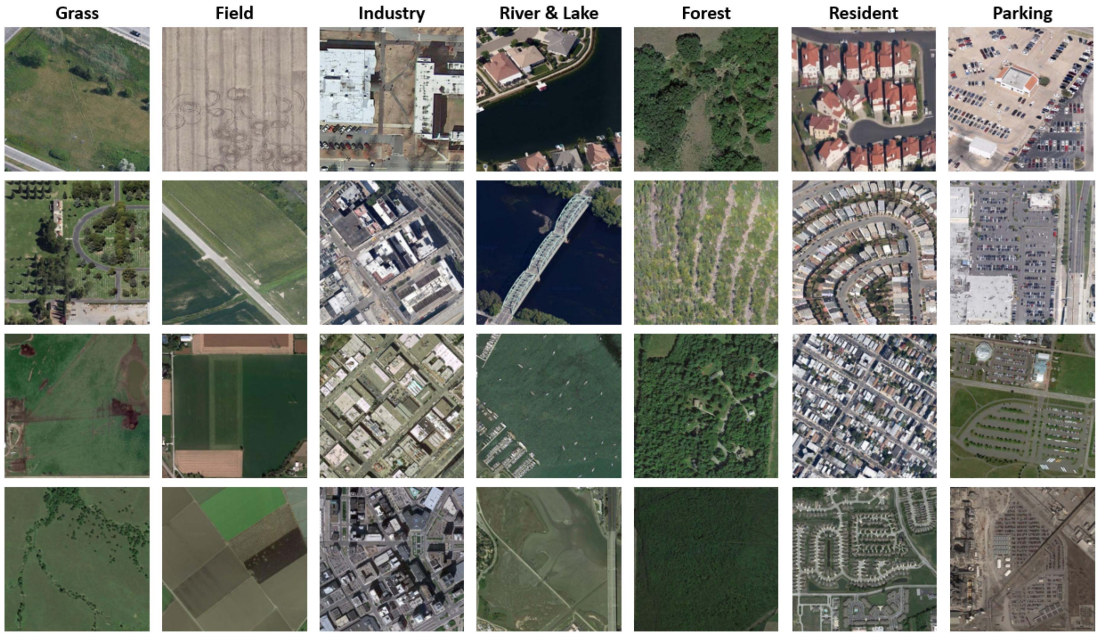

Fig. 4. Images and classes of RSSCN7 dataset.

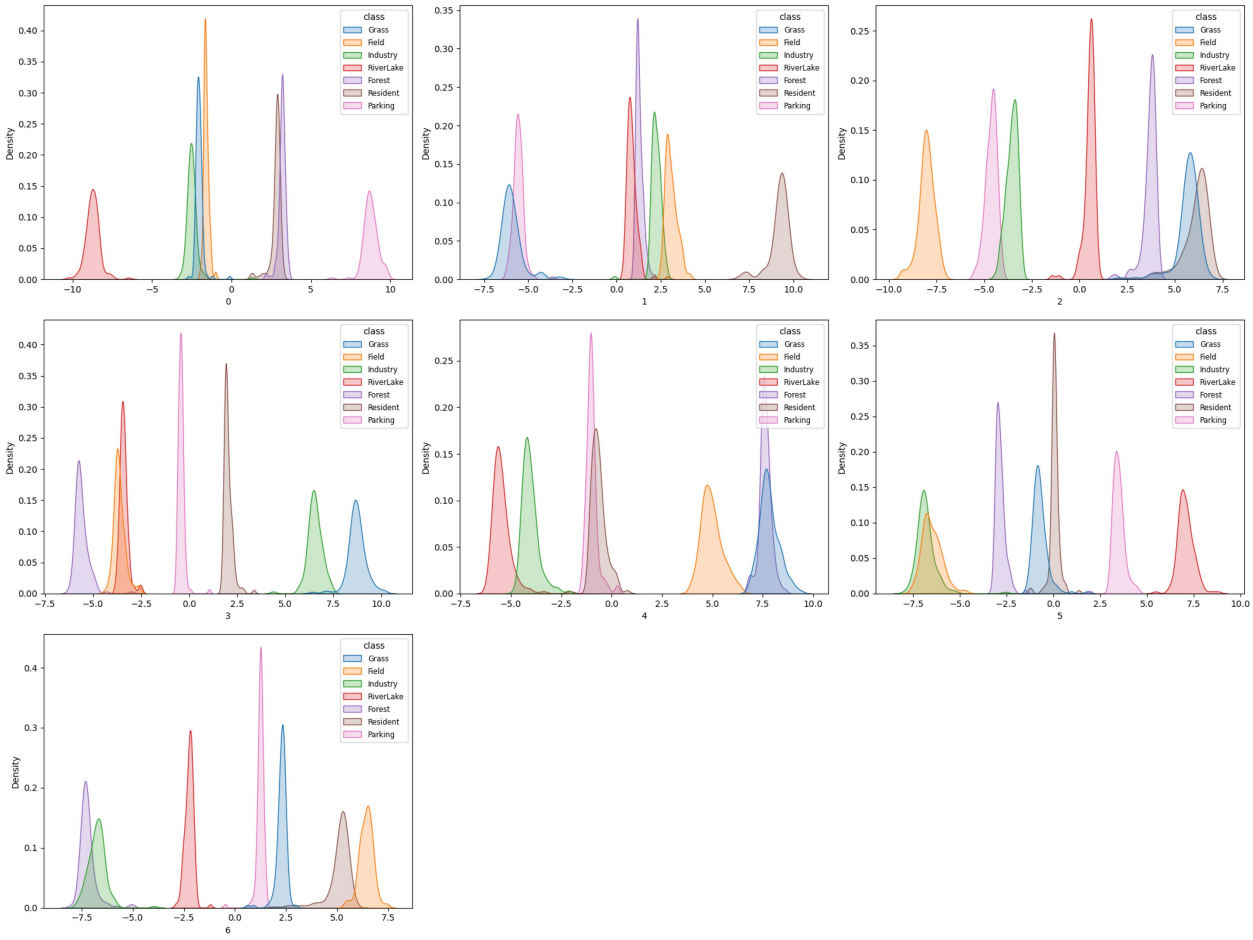

Fig. 5. Distribution of the feature representations generated by the last layer of MKANet-Class trained with DLDA, each subfigure represents an individual dimension.

strategy at the first 50 epochs. For a fair comparison, all the networks were trained from scratch without pretraining on other datasets.

Data augmentation: Random cropping into size 384 × 384 pixels, random flipping, random rotation, and color jittering operations were employed on the input images in the training process.

As presents in Table II, MKANet-Class outperformed other state-of-the-art lightweight networks when they were trained with cross entropy loss function. LDA-based loss function, DLDA and BDLDA, helped MKANet-Class widen the leading advantage. MKANet-Class trained with BDLDA shows fairly good performance, surpassing the MKANet-Class trained with cross entropy and DLDA by 2% and 1.4%, respectively. We also conducted comparative experiments on MobileNetV3, BDLDA performed better than cross entropy and DLDA by 0.81% and 0.59%, respectively.

For the training samples, the distributions of the feature representations generated by the last layer of MKANet-Class trained with DLDA are illustrated in Fig. 5. It can be seen that the interclass variance is not large enough and the intraclass variance is not small enough, which causes some samples of different classes overlap.

TABLE II
COMPARISON OF BDLDA WITH CROSS ENTROPY AND DLDA ON RSSCN7 DATASET

Method	Accuracy (%)
Cross Entropy:	
MobileNetV3 [31]	85.71 ± 0.26
REGNETX-400MF [32]	86.14 ± 0.68
ShuffleNetV2 x1.0 [33]	86.91 ± 0.59
STDC1-Class [34]	88.82 ± 0.38
ResNet34 [35]	88.86 ± 0.11
MKANet-Class	89.67 ± 0.64
DLDA:	
MobileNetV3 [31]	85.83 ± 0.13
MKANet-Class	90.22 ± 0.25
BDLDA:	
MobileNetV3 [31]	86.52 ± 0.04
MKANet-Class	91.62 ± 0.30

Each method was trained and tested 5 times to report the mean accuracy and variance.

For the training samples, the distributions of the feature representations generated by the last layer of MKANet-Class trained with BDLDA are illustrated in Fig. 6. It shows BDLDA can make MKANet-Class produce more discriminative feature representations with lower intraclass variance and higher interclass variance.

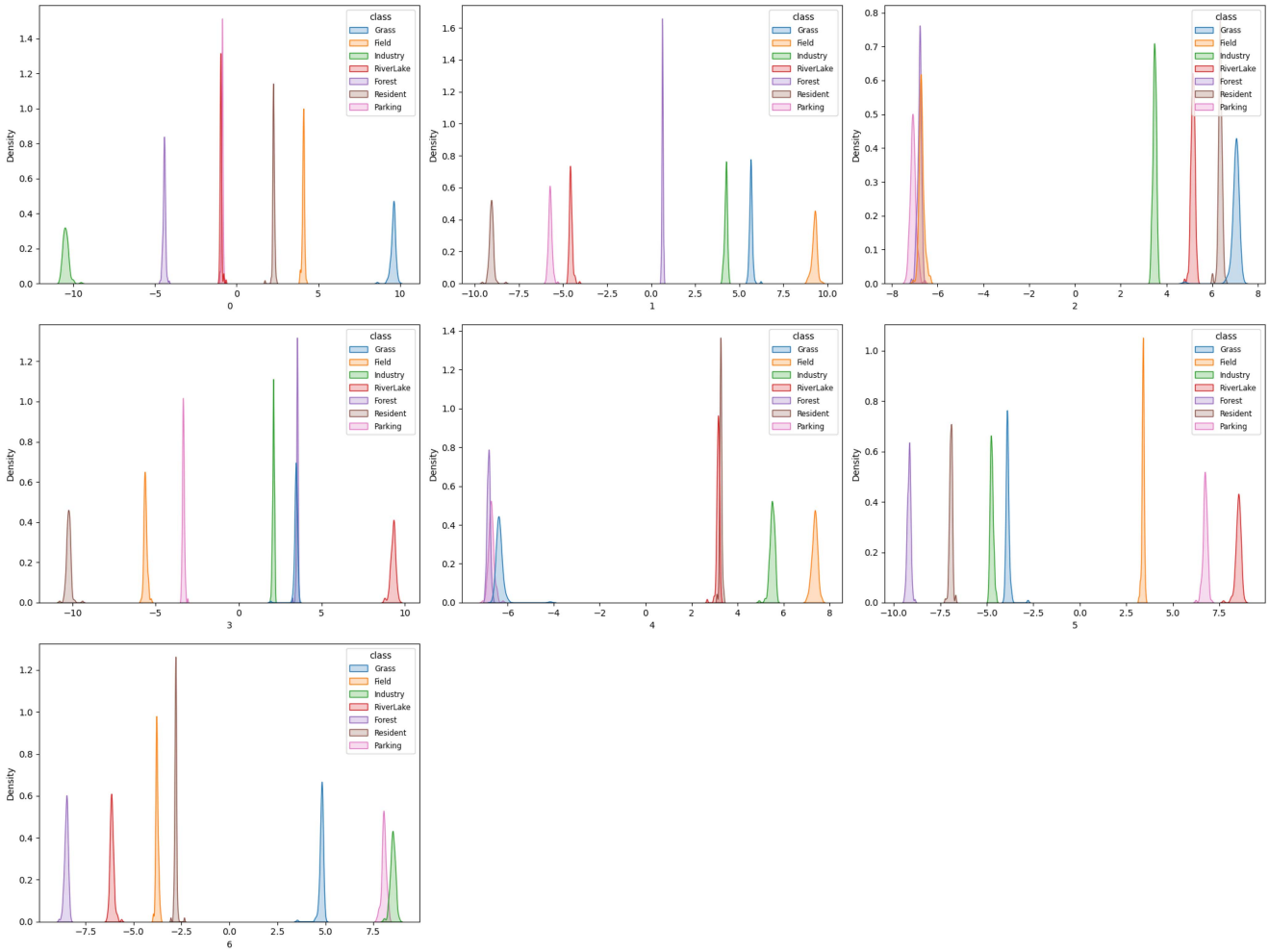

Fig. 6. Distribution of the feature representations generated by the last layer of MKANet-Class trained with BDLDA, each subfigure represents an individual dimension.

TABLE III
COMPARISON OF BDLDA WITH CROSS ENTROPY AND DLDA ON OPTIMAL-31 DATASET

Method	Accuracy (%)
Cross Entropy:	
REGNETX-400MF [32]	56.97 ± 1.21
MobileNetV3 [31]	60.17 ± 1.47
STDC1-Class [34]	61.00 ± 0.08
ShuffleNetV2 x1.0 [33]	61.83 ± 0.06
ResNet34 [35]	67.97 ± 0.70
MKANet-Class	69.95 ± 1.41
DLDA:	
MKANet-Class	76.34 ± 0.13
BDLDA:	
MKANet-Class	77.30 ± 0.07

Each method was trained and tested 5 times to report the mean accuracy and variance.

C. Image Classification Experimental Results on OPTIMAL-31

OPTIMAL-31 [36] is a remote sensing image dataset (see Fig. 7), it contains 31 scene classes of image size 256 × 256 pixels. For each class, there are 60 images. This dataset is more challenging than other datasets since it contains more classes and fewer samples per class. In this experiment, the total 1860 images were split as a training set, validation set, and test set with the ratio of 3:3:4.

Training details: AdamW was used as an optimizer with batch size 558, and the base learning rate was 0.001 with cosine decay. The number of epochs was 2000 with a warmup strategy at the first 20 epochs. For a fair comparison, all the networks were trained from scratch without pretraining on other datasets.

Data augmentation: Random cropping into size 224 × 224 pixels, random flipping, random rotation, and color jittering operations were employed on the input images in the training process.

As presented in Table III, the MKANet-Class trained by BDLDA also gained a satisfactory result, exceeding that trained by DLDA by 1%. The MKANet-Class trained with both methods surpassed that trained with cross entropy by more than 6%, which demonstrates the superiority of LDA-based loss functions over cross entropy. The reason can be explained by Figs. 8–10. As illustrated in Fig. 8, the MKANet-Class trained by cross entropy was soon overfitted. After the 1000th epoch, the training

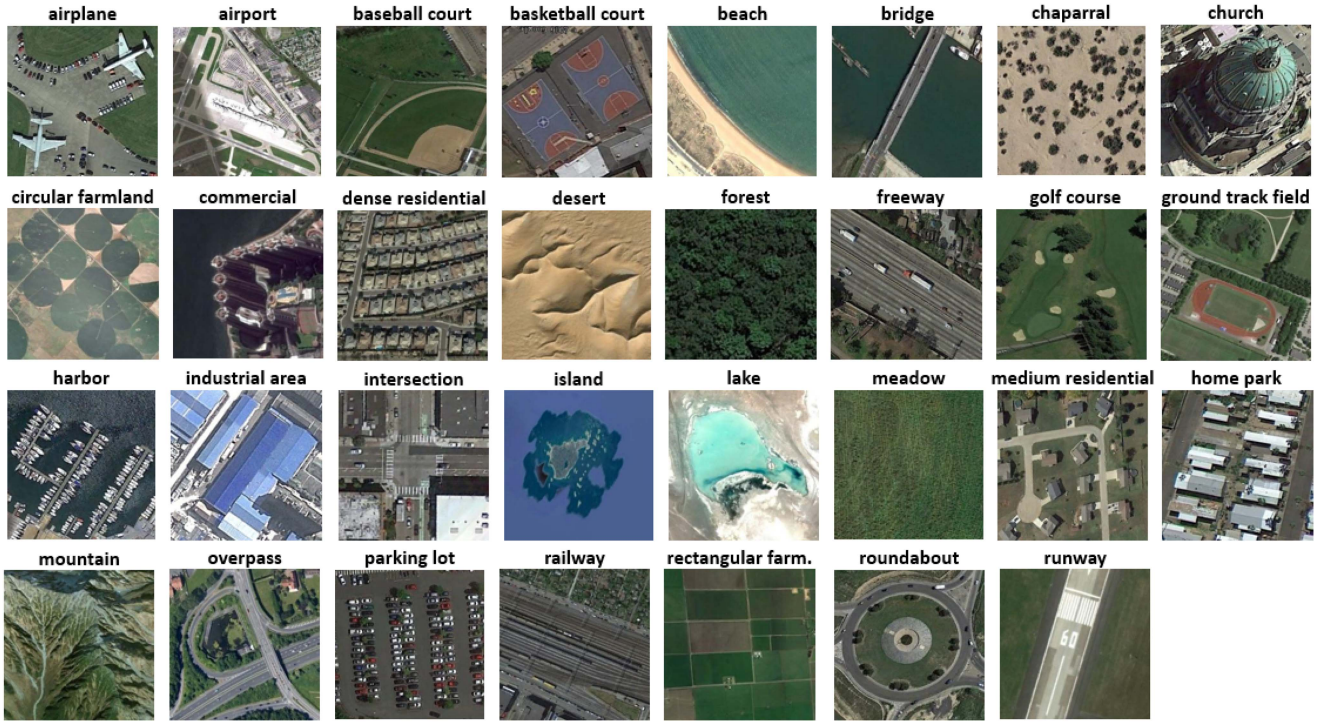

Fig. 7. Images and classes of OPTIMAL-31 dataset.

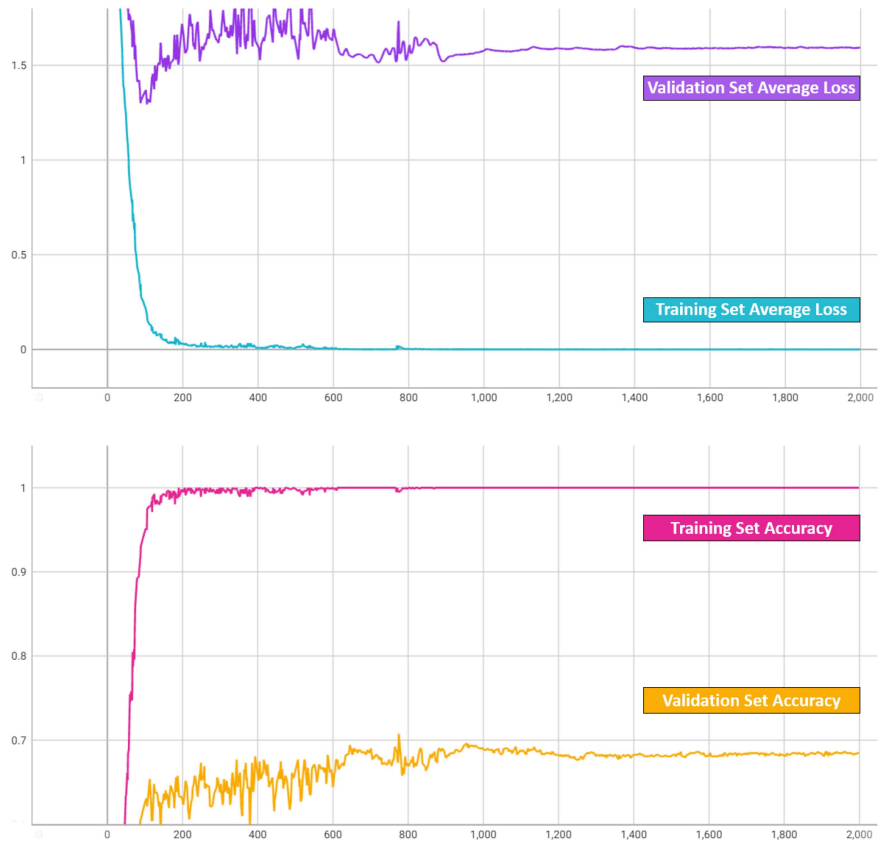

Fig. 8. Average loss and accuracy scores of training and validation set of the MKANet-Class trained with a cross entropy loss function.

Fig. 9. Training set mean eigenvalue and validation set accuracy score of the MKANet-Class trained with DLDA loss function.

Fig. 10. Training set mean eigenvalue and validation set accuracy score of the MKANet-Class trained with BDLDA loss function.

TABLE IV
IMAGE RETRIEVAL EXPERIMENTAL RESULTS OF MKANET-CLASS ON RSSCN7 DATASET

P@10			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.8691	0.8718	0.8794
DLDA	0.8862	0.8865	0.8912
BDLDA	0.8932	0.8931	0.8963
P@20			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.8695	0.8695	0.8797
DLDA	0.8860	0.8874	0.8916
BDLDA	0.8932	0.8957	0.8987
mAP@10			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.8869	0.8879	0.8927
DLDA	0.8938	0.8954	0.8990
BDLDA	0.8982	0.8991	0.9016
mAP@20			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.8808	0.8825	0.8900
DLDA	0.8922	0.8939	0.8982
BDLDA	0.8979	0.8986	0.9023
mAP			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.8461	0.8548	0.8796
DLDA	0.8951	0.8987	0.9049
BDLDA	0.9106	0.9128	0.9167

TABLE V
IMAGE RETRIEVAL EXPERIMENTAL RESULTS OF MOBILENETV3 ON RSSCN7 DATASET

P@10			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.8280	0.8279	0.8377
DLDA	0.8506	0.8509	0.8530
BDLDA	0.8554	0.8568	0.8573
P@20			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.8259	0.8280	0.8375
DLDA	0.8496	0.8507	0.8533
BDLDA	0.8584	0.8597	0.8612
mAP@10			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.8475	0.8472	0.8524
DLDA	0.8594	0.8587	0.8596
BDLDA	0.8621	0.8631	0.8661
mAP@20			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.8403	0.8409	0.8491
DLDA	0.8594	0.8586	0.8606
BDLDA	0.8616	0.8642	0.8670
mAP			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.7746	0.7933	0.8362
DLDA	0.8460	0.8478	0.8478
BDLDA	0.8471	0.8511	0.8586

set average loss was nearly zero, and the validation set accuracy stopped increasing. In contrast, as shown in Figs. 9 and 10, for the MKANet-Class trained by DLDA and BDLDA, the training set mean eigenvalue kept increasing, and the validation set accuracy grew with it.

It can also be seen from Figs. 9 and 10 that the training set mean eigenvalue of BDLDA was larger than that of DLDA. Since each eigenvalue v_i is the quantitative measurement of the discrimination ability in the direction of the corresponding eigenvector $\boldsymbol{u}_i$, it verifies the feature representations generated by the network trained with BDLDA are more discriminative than those generated by the network trained with DLDA. This is in accordance with the experimental results that the accuracy scores of the validation set and test set of BDLDA were higher than those of DLDA.

D. Image Retrieval Experimental Results on RSSCN7

We used the training set and validation set as dictionary set, the test set as query set for image retrieval experiments. Three commonly used performance metrics, mean average precision (mAP), mean average precision at k (mAP@k), and precision at k (P@k), where k is the number of retrieved images, are used to evaluate the retrieval performance. Given one query with the known label information in advance, precision at k (P@k) denotes the consistency rate that k returned results share the same label with the query [2]. In addition, mAP is defined as

$$\text{mAP} = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{n_i}\sum_{k=1}^{n_i}\text{precision}(R_{ik}) \tag{18}$$

where $q_i \in Q$ represents one query and n_i denotes the number of returned results relevant to q_i in the dataset. Suppose the relevant results are ordered as $\{r_1, r_2, \ldots, r_{n_i}\}$, and then R_{ik} represents the set of ranked retrieval results from the top result to r_k.

We calculated and ranked similarity based on the 512-dimension feature vectors output by the GAP layer of MKANet-Class. As presented in Table IV, BDLDA-based feature representations perform better than cross entropy-based and DLDA-based feature representations on all the image retrieval performance metrics. Their performance gaps are approximate to those in the image classification experiments, which demonstrates that the image feature representations play a decisive role in the CBIR performance and those generated by BDLDA are more discriminative. Meanwhile, compared with the Euclidean distance and the Manhattan distance, RED can further increase image retrieval accuracy at no additional computational cost. Our proposed two methods, BDLDA and RED, combined together achieve the best result.

We also conducted comparative experiments based on the 1024-dimension feature vectors output by the second to last linear layer of MobileNetV3. As presented in Table V, as a loss function, BDLDA outperforms cross entropy and DLDA on generating more effective image feature representations. And RED maintains superiority over the Euclidean distance and the Manhattan distance in similarity calculation of image feature representations of higher dimensions.

E. Image Retrieval Experimental Results on OPTIMAL-31

Under the same experimental setting as the RSSCN7 dataset, we conducted image retrieval experiments based on the image feature representations generated by MKANet-Class on the OPTIMAL-31 dataset. As presented in Table VI, the advantages of BDLDA and RED expand on this more challenging

TABLE VI
IMAGE RETRIEVAL EXPERIMENTAL RESULTS OF MKANET-CLASS ON OPTIMAL-31 DATASET

P@10			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.6329	0.6443	0.6648
DLDA	0.6805	0.7043	0.7255
BDLDA	0.7180	0.7335	0.7447
P@20			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.6114	0.6232	0.6500
DLDA	0.6694	0.6964	0.7296
BDLDA	0.7149	0.7344	0.7533
mAP@10			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.6924	0.6999	0.7151
DLDA	0.7299	0.7440	0.7543
BDLDA	0.7439	0.7557	0.7576
mAP@20			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.6698	0.6774	0.6971
DLDA	0.7181	0.7399	0.7563
BDLDA	0.7365	0.7500	0.7638
mAP			
Loss Function	Euclidean	Manhattan	RED
Cross Entropy	0.5694	0.5838	0.6132
DLDA	0.6148	0.6486	0.7106
BDLDA	0.6864	0.6985	0.7399

dataset which contains more classes and fewer samples per class. BDLDA and RED combined together achieve the best result.

V. CONCLUSION

As an effective solution to remote sensing image retrieval, CBIR retrieves the desired images automatically by their visual content similarities to the query image. Although the image classification and retrieval tasks differ in terms of their goals, the classification labels can guide the CNN layers to learn feature representations targeted to the labels. The feature representations are then indexed by the similarity metrics for selecting the images with similar visual content as the query image. The CBIR performance heavily depends on the effectiveness of the image representation features and similarity evaluation metrics. Ideal image feature representations have dispersed interclass, compact intraclass distribution. However, the neural networks employed by many CBIR methods are trained with cross entropy loss, which does not directly optimize the metrics that evaluates interclass variance over intraclass variance, hence, their feature representations are suboptimal. Meanwhile, the traditional distance metrics used by many CBIR methods cannot index the similarity of feature representations well in high-dimensional space. For higher image classification accuracy and better CBIR performance, we propose a discriminative feature learning approach with distinguishable distance metrics for remote sensing image classification and retrieval. By balancing the diagonal elements and nondiagonal elements of the within-class scatter matrix of DLDA, our proposed loss function, BDLDA, can better optimize the Rayleigh–Ritz quotient, which measures interclass variance over intraclass variance. In addition, the proposed distance metrics, RED, is more capable of maintaining distance contrast in high dimensionality, therefore, it can better index similarity for feature representations in high dimensionality. In the experiments, compared with cross entropy and DLDA, our proposed BDLDA can raise image classification and retrieval accuracy by several percents. In addition, RED can further increase image retrieval accuracy by around one percent at no additional computational cost.

In future research, we will investigate and improve the performance of BDLDA on unbalanced datasets. Extending the application of BDLDA to other tasks, such as semantic segmentation and object detection, is also a potential research orientation.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and members of the editorial team for their comments and suggestions.

REFERENCES

[1] Z. Zhang, Z. Qu, S. Liu, D. Li, J. Cao, and G. Xie, "Expandable on-board real-time edge computing architecture for Luojia3 intelligent remote sensing satellite," *Remote Sens.*, vol. 14, no. 15, 2022, Art. no. 3596.
[2] Y. Li, J. Ma, and Y. Zhang, "Image retrieval from remote sensing big data: A survey," *Inf. Fusion*, vol. 67, pp. 94–115, 2021.
[3] M. Wang, Z. Zhang, Y. Zhu, Z. Dong, and Y. Li, "Embedded GPU implementation of sensor correction for on-board real-time stream computing of high-resolution optical satellite imagery," *J. Real-Time Image Process.*, vol. 15, no. 3, pp. 565–581, 2018.
[4] W. Mi, Z. Zhiqi, D. Zhipeng, J. Shuying, and H. SU, "Stream-computing based high accuracy on-board real-time cloud detection for high resolution optical satellite imagery," *Acta Geodaetica et Cartographica Sinica*, vol. 47, no. 6, 2018, Art. no. 760.
[5] L. Deren, "Towards geo-spatial information science in big data era," *Acta Geodaetica et Cartographica Sinica*, vol. 45, no. 4, 2016, Art. no. 379.
[6] W. Xiong, Y. Lv, Y. Cui, X. Zhang, and X. Gu, "A discriminative feature learning approach for remote sensing image retrieval," *Remote Sens.*, vol. 11, no. 3, 2019, Art. no. 281.
[7] Z. Shao, W. Zhou, Q. Cheng, C. Diao, and L. Zhang, "An effective hyperspectral image retrieval method using integrated spectral and textural features," *Sensor Rev.*, vol. 35, pp. 274–281, 2015.
[8] Z. Shao, W. Zhou, L. Zhang, and J. Hou, "Improved color texture descriptors for remote sensing image retrieval," *J. Appl. Remote Sens.*, vol. 8, no. 1, 2014, Art. no. 083584.
[9] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, vol. 3, pp. 1470–1470.
[10] Z. Shao, W. Zhou, and Q. Cheng, "Remote sensing image retrieval with combined features of salient region," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 40, no. 6, 2014, Art. no. 83.
[11] Y. Li, Y. Zhang, C. Tao, and H. Zhu, "Content-based high-resolution remote sensing image retrieval via unsupervised feature learning and collaborative affinity metric fusion," *Remote Sens.*, vol. 8, no. 9, 2016, Art. no. 709.
[12] J. Ren, Z. Wang, and M. Xu, "An autoencoder-based learning method for wireless communication protocol identification," in *Proc. Int. Conf. Commun. Netw. China*, 2017, pp. 535–545.
[13] W. Zhou, Z. Shao, C. Diao, and Q. Cheng, "High-resolution remote-sensing imagery retrieval using sparse features by auto-encoder," *Remote Sens. Lett.*, vol. 6, no. 10, pp. 775–783, 2015.
[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.
[15] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 499–515.
[16] R. Cao et al., "Enhancing remote sensing image retrieval using a triplet deep metric learning network," *Int. J. Remote Sens.*, vol. 41, no. 2, pp. 740–751, 2020.
[17] P. Liu, G. Gou, X. Shan, D. Tao, and Q. Zhou, "Global optimal structured embedding learning for remote sensing image retrieval," *Sensors*, vol. 20, no. 1, 2020, Art. no. 291.

[18] 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.
[19] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Mullers, "Fisher discriminant analysis with kernels," in *Proc. IEEE Signal Process. Soc. Workshop*, 1999, pp. 41–48.
[20] L. Li, M. Doroslovački, and M. H. Loew, "Discriminant analysis deep neural networks," in *Proc. 53rd Annu. Conf. Inf. Sci. Syst.*, 2019, pp. 1–6.
[21] M. Dorfer, R. Kelz, and G. Widmer, "Deep linear discriminant analysis," 2015, *arXiv:1511.04707*.
[22] K. B. J. G. R. Ramakrishnan and U. Shaft, "When is 'nearest neighbor' meaningful?," in *Proc. 7th Int. Conf. Database Theory*, 1999, pp. 217–235.
[23] B. Ghojogh, F. Karray, and M. Crowley, "Fisher and kernel fisher discriminant analysis: Tutorial," 2019, *arXiv:1906.09436*.
[24] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Stat. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.
[25] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recognit. Lett.*, vol. 26, no. 2, pp. 181–191, 2005.
[26] A. Stuhlsatz, J. Lippel, and T. Zielke, "Feature extraction with deep neural networks by a generalized discriminant analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 596–608, Apr. 2012.
[27] J. Friedman et al., *The Elements of Statistical Learning* (Springer Series in Statistics), vol. 1, no. 10. New York, NY, USA: Springer, 2001.
[28] Z. Zhang, W. Lu, J. Cao, and G. Xie, "MKANet: An efficient network with Sobel boundary loss for land-cover classification of satellite remote sensing imagery," *Remote Sens.*, vol. 14, no. 18, 2022, Art. no. 4514.
[29] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 11, pp. 2321–2325, Nov. 2015.
[30] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
[31] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1314–1324.
[32] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollár, "Designing network design spaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10428–10436.
[33] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 116–131.
[34] M. Fan et al., "Rethinking BiseNet for real-time semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9716–9725.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[36] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, Feb. 2019.

Zhiqi Zhang received the B.Sc. degree in geographic information systems from Huazhong Agricultural University, Wuhan, China, the B.Eng. degree in computer science and technology from the Huazhong University of Science and Technology, Wuhan, China, the M.Eng. degree in computer technology and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2006, 2006, 2015, and 2018, respectively.

He is currently an Associate Professor with the School of Computer Science, Hubei University of Technology, Wuhan, China. His research interests include system architecture, algorithm optimization, AI, and high-performance processing of remote sensing.

Wen Lu received the B.Eng. degree in materials physics from the Wuhan University of Technology, Wuhan, China, in 2007. He is currently working toward the M.Sc. degree in computer science and technology with the School of Computer Science, Hubei University of Technology, Wuhan, China.

His research interests include computer vision, remote sensing, machine learning, and deep learning.

Xiaoxiao Feng received the B.Sc. degree in surveying and mapping from Southeast University, Nanjing, China, in 2014 and the M.Sc. degree in earth exploration and information technology from the China University of Geology, Wuhan, China, in 2017, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2021.

She is currently a Lecturer with the School of Computer Science, Hubei University of Technology, Wuhan, China. Her research interests include high spatial resolution and hyperspectral remote sensing image processing and analysis.

Jinshan Cao received the Ph.D. degree in photogrammetry and remote sensing from the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China, in 2012.

He is currently an Associate Professor with the School of Computer Science, Hubei University of Technology, Wuhan, China. His research interests include geometric calibration, sensor orientation, and image registration of high-resolution satellite imagery.

Guangqi Xie received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, in 2021.

He is currently a Lecturer with the School of Computer Science, Hubei University of Technology, Wuhan, China. His research interest includes image matching and registration, pansharpening, and image super-resolution.