IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 5, 2020, accepted June 12, 2020, date of publication June 23, 2020, date of current version July 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3004360

Query-Adaptive Remote Sensing Image Retrieval Based on Image Rank Similarity and Image-to-Query Class Similarity

FAMAO YE[1,5], XUQING ZHAO[2], WEI LUO[3], DAJUN LI[1], AND WEIDONG MIN[4], (Member, IEEE)

[1]School of Surveying and Mapping Engineering, East China University of Technology, Nanchang 330013, China
[2]School of Information Engineering, Jiangxi Technical College of Manufacturing, Nanchang 330095, China
[3]School of Information Engineering, Nanchang University, Nanchang 330031, China
[4]School of Software, Nanchang University, Nanchang 330047, China
[5]Key Laboratory for Digital Land and Resources of Jiangxi Province, East China University of Technology, Nanchang 330013, China

Corresponding authors: Dajun Li (djli@ecut.edu.cn) and Weidong Min (minweidong@ncu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 41261091, Grant 41801288, Grant 61762061, and Grant 41861052, in part by the Natural Science Foundation of Jiangxi Province, China, under Grant 20161ACB20004, and in part by the Key Laboratory for Digital Land and Resources of Jiangxi Province, East China University of Technology, China, under Grant DLLJ201908.

ABSTRACT Many image features have been proposed for image retrieval; hence, effectively fusing these features to alleviate the large variation in performance among image queries when using single image features has become a major challenge in remote sensing (RS) image retrieval. Because high-resolution remote sensing images have abundant and complex visual contents, accurately measuring the similarity between two images is another important problem. To address these challenges, we propose a novel RS image retrieval method that uses query-adaptive feature weights to fuse features and utilizes two image similarities to improve retrieval performance. First, we use the image rank similarity, which measures the similarity between two images according to their corresponding top-m image lists from a reference image collection, to calculate the similarity of each feature between a query image and each retrieved image. Then, we assign a weight to each feature to fuse these features via our query-adaptive weighting method. Finally, we take the query image and its neighborhood set selected from the retrieval dataset as the query class and utilize the image-to-query class similarity to re-rank the retrieval results. Extensive experiments are conducted on two publicly available RS image databases. Compared with the state-of-the-art methods, the proposed method can significantly enhance the retrieval precision.

INDEX TERMS Content-based remote sensing image retrieval, query-adaptive, image rank similarity, image-to-query class similarity.

I. INTRODUCTION

As sensor technology and remote sensing (RS) technology improve, both the quality and quantity of RS images are increasing quickly [1]. Researchers can now readily acquire many high-resolution remote sensing (HRRS) images that were captured from satellites or aircraft. To efficiently exploit the rapid accumulation of RS images, it is necessary to design robust and automatic tools for their retrieval, mining, and management. Consequently, the adaption of content-based image retrieval (CBIR) to this context has become a highly active research area in the RS community in the last decade [2].

CBIR retrieves the relevant images for a query image from an image database by measuring features that are extracted from the images, rather than depending on the accompanying text information. Among the content-based remote sensing image retrieval (CBRSIR) methods, the traditional global feature descriptors of image content include color, texture [2], [3], and shape [4]. However, the global descriptors might fail due to the invariance expectation if an image changes due to illumination variation, image translation, or truncation [5]. Many image retrieval methods that are based on local features have been proposed for

The associate editor coordinating the review of this manuscript and approving it for publication was Nuno Garcia.

overcoming the shortcomings of global features. Most of those methods extract image features from salient points of their inputs via feature encoding techniques, such as bag of words (BoW) [6], [7], vector of locally aggregated descriptors (VLAD) [8], and improved Fisher kernel (IFK) [9]. The most widely used method for point detection is based on the scale-invariant feature transform (SIFT).

In recent years, convolutional neural networks (CNNs) have dramatically improved the states of the art in image object recognition [10], [11], image classification [12], and image scene analysis [13], [14]. Inspired by this success, various CBRSIR methods that are based on CNNs have been put forward and seem to be becoming more popular than SIFT-based models. Region-based cascade pooling (RBCP) features, which were aggregated from convolutional layers of pre-trained or fine-tuned CNN models, were proposed for retrieving HRRS images [15]. Zhou *et al.* [16] introduced two effective CNN schemes, one of which uses pre-trained CNN architectures and the other a self-designed CNN model, for extracting CNN features for retrieval. Ge *et al.* [17] developed two CNN features for retrieving HRRS images: one is extracted directly from the outputs of high-level layers and the other is aggregated from the outputs of mid-level layers via average pooling. Wang *et al.* [18] proposed a graph-based learning method with a three-layer framework for integrating the strengths of query expansion and the fusion of holistic and local features.

The retrieval features that are discussed above were used to retrieve the images of the same scene or class from an image database according to the feature similarity between the query image and the retrieved images. There are many similarity metrics of retrieval features. Similarity measures may yield results that differ significantly using the same retrieval feature in the same retrieval task. Therefore, selecting an effective similarity measure for a retrieval task is of substantial importance. Most of the existing CBRSIR methods are based on sorting of the similarities between a query image and the retrieved images, namely, computing the similarities of only pairs of images. However, these methods ignore that a retrieved image with a high similarity score may not belong to the same class as the query image because RS images typically have complex backgrounds. In addition, these methods discard the rich information that is encoded in the relations among images, such as the potential class information and similar degrees among them [19]. It was proved that this information can be used to improve the retrieval precision [18]. To utilize this information and overcome the problems that are described above, two image similarity measures, namely, the image rank similarity and the image-to-query class similarity, are proposed in this paper. First, the image rank similarity is used to calculate the similarity between two images by considering the context information of other images that are similar to them. When a query image and a retrieved image are retrieved from the same image collection, two top-m image lists are obtained. If the two images are highly similar, there may exist many common images in the two image lists. The image rank similarity takes the number of common images and every image rank of the common images into consideration to measure the similarity between the two images. Second, the image-to-query class similarity uses the potential class information of the images that are most similar to the query image. The retrieved images should be similar to all the images in the unknown image class to which the query image belongs, not just to the query image [20]. We use the k-nearest neighbors' method to identify images that may belong to the query image class from an image collection. The similarity between a retrieved image and the query image is calculated as the average similarity between the retrieved image and all images in the query class.

RS images often represent large natural geographical scenes that have abundant and complex visual contents [21]. Therefore, it is highly difficult to accurately retrieve the desired results for all types of RS images using a single feature. It needs to combine some features effectively for RSIR. However, different features may be suitable for different query images. Given a query image, we need to automatically evaluate the effectiveness of a to-be-fused feature so that suitable features are used, while unsuitable features are ignored. Since we have no prior knowledge of the query image, it is important that we estimate unsupervised the effectiveness of a feature. In light of the above analysis, we put forward a query-adaptive feature weighting method based on an observation that suitable features have a higher head and a lower tail in the score curve than unsuitable features. The query-adaptive feature weighting method is a score-level fusion scheme and can adaptively distribute weights among the to-be-fused features for query images.

This paper proposes a suite of technical schemes for CBRSIR, which utilizes two image similarity measures and a query-adaptive weighting method to combine multiple features. The retrieval process of the proposed method is shown in Fig.1. The main contributions of this paper are as follows:

i. A novel query-adaptive weighting method for remote sensing image retrieval (RSIR) is developed. The method utilizes the score curve shapes of features to calculate the weight of each feature for various query images. Our method computes weights on the fly and is independent of the retrieval database; hence, it is highly suitable for dynamic systems.
ii. We propose the new image rank similarity, which measures the similarity between two images according to their corresponding top *m* image lists from a reference image collection. The image features that are used to compute image similarities are of various dimensions and scales. Fusing these features requires normalization procedures that can affect the retrieval accuracy. Because the image rank similarity does not depend on image features directly, it can be easily used in image retrieval tasks.
iii. We combine the query-adaptive fusion method with two similarities, namely, the image rank similarity and the image-to-query class similarity for RSIR for the

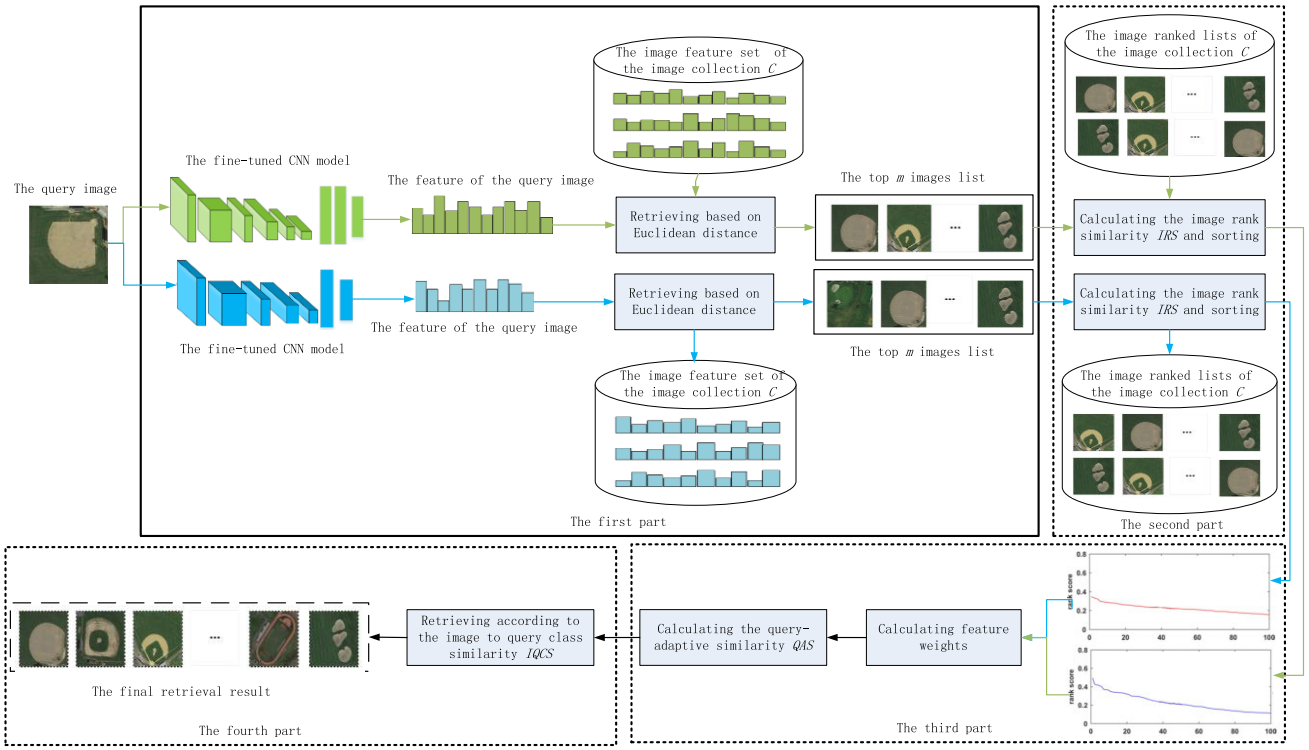

FIGURE 1. The retrieval pipeline of the proposed method. The retrieval process consists of four parts. In the first part, extract features of the query image via fine-tuned CNN models, and retrieve the query image from image collection *C* based on Euclidean distance to get the top m images list of the query image for each feature. In the second part, calculate the image rank similarity *IRS* between the top *m* images list of the query image and that of each image in Collection *C*, and sort in decreasing order. In the third part, calculate weights for each weight via the query-adaptive weight method and calculate the query-adaptive similarity *QAS* between the query image and each image in the retrieval database. In the last part, retrieve the query image from the retrieval database according to the image-to-query class similarity, and get the final retrieval result.

first time. The experimental results showed that our method can realize higher retrieval performance than some state-of-the-art RS image retrieval methods.

The remainder of this paper is organized as follows: Section II reviews related works on image similarity measurement approaches and feature fusion in RSIR. Section III presents the details of the proposed image retrieval framework. The experimental results are analyzed in Section IV. Finally, Section V presents the conclusions of this paper.

II. RELATED WORK

In this section, we will present the related work about image similarity measure metrics in RSIR, and feature fusion in RSIR in the following section.

A. IMAGE SIMILARITY MEASUREMENT APPROACHES IN RSIR

An image similarity measure, which calculates the similarity of extracted image features, is typically used to determine the image similarity between two images [22]. The Euclidean distance is one of the most common similarity measures in image retrieval, which is used in [6], [8], [15]–[17], [23]. Other similarity measures have been used by researchers: Ding *et al.* [3] used the angular similarity with weight to calculate the similarity of eigenvalues in the frequency domain. Xia *et al.* [23] used four similarity metrics, namely, the Euclidean, cosine, Manhattan, and chi-square metrics, for various feature types. The Euclidean and cosine similarity metrics yielded superior experimental results in their research. Chaudhuri *et al.* [24] proposed a graph similarity by combining the node distance and the edge distance of regions, which considers both region characteristics and their relations. Graña and Veganzones [25] presented an endmember-based distance measure, which is particularly suitable for retrieving hyperspectral images, while Veganzones *et al.* [26] developed a normalized dictionary distance. Wang and Song [21] proposed a spatial scene semantic similarity that considers the object area, attribution, orientation, and topological features. These works used features that were extracted from a single image and ignored the context information among images.

These image similarities mentioned above directly use features extracted from images. While some image similarities measure the similarity between two images according to the similarity of the ranked lists that result from using them as queries. The Jaccard similarity between two ranking lists of two images is defined as the size of the intersection divided by the size of the union of two lists.

But The Jaccard similarity does not include order information [27]. In [28], a Jaccard similarity considering dif-

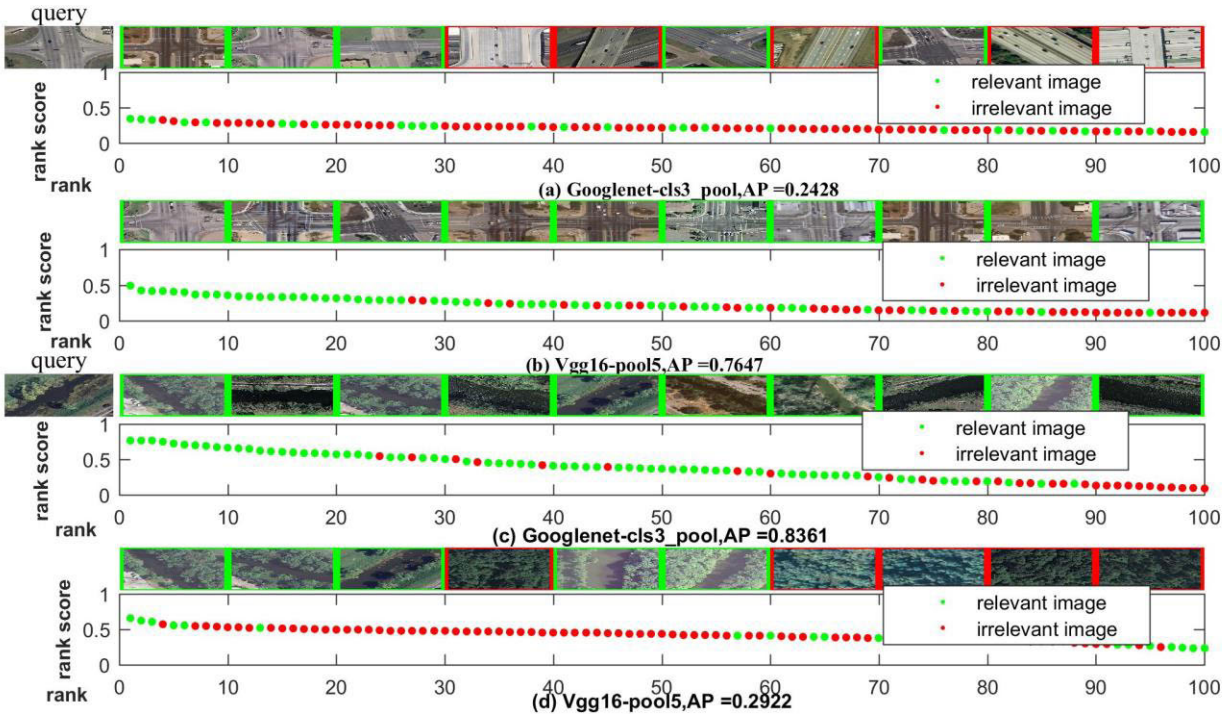

FIGURE 2. Comparison of the sorted score curve of suitable features and that of unsuitable features. For two queries in UCMD, the score lists in (a) and (c) are obtained by the Googlenet-cls3_pool feature, and the score lists in (b) and (d) are obtained by the Vgg16-pool5 feature. We plot the corresponding top 10 ranked images and the sorted scores for rank 1-100. Relevant images are marked in green, and irrelevant ones red. In (b) and (c), the features produce good performance (AP = 0.7647 and AP = 0.8361), while the features in (a) and (d) lead to a bad performance (AP = 0.2428 and AP = 0.2922). Note that well-sorted score curves have a higher score in the head and lower value in the tail than bad sorted score curves, and the retrieval performance of a feature is related to the query image.

ferent depths was presented, which gives more weight to the top ranked results than lower results. Webber *et al.* [29] also proposed a similarity, rank-biased overlap (RBO), based on a simple model in which the user compares the similarity of the two ranking lists at incrementally increasing depths. The weight of overlap measure is calculated according to probabilities defined at each depth. But the incrementally increasing depths may affect the performance of speed. Chen *et al.* [27] exploited the ranking consistency information among images obtained by Jaccard or RBO to refine an existing ranking list. In [18], a similarity measure that considers the spatial distributions of the image features between two top-ranked image lists was proposed.

B. FEATURE FUSION IN RSIR

Over the last two decades, many image features have been proposed, for example, BoW and CNN features. Because a single feature is insufficient for completely characterizing the image information content [18], feature fusion is often an efficient method that combines the complementary advantages that are offered by each feature to enhance the overall retrieval accuracy. Ge *et al*. [17] combined VGGM, VGG16, GoogLeNet, and BoW features by assigning a global weight to each feature. The global weights were obtained by manual features and texture features as the node attributes of a region to fuse them. Aptoula [2] directly combined four texture descriptors into a single vector: the circular covariance histogram (CCH), the rotation-invariant point triplets (RITs) and two texture descriptors that are based on the Fourier power spectrum (FPS) of an image's quasi-flat zone (QFZ) representation. Since a feature may differ in importance among query images, the fusion of features via a global approach in these methods may not be effective for improving the image retrieval precision. Zhang *et al*. [22] proposed a graph-based query-specific fusion approach for fusing the retrieval results based on holistic and local features for image retrieval. Wang *et al*. [18] proposed a three-layer framework for RSIR that was inspired by the above method. However, these methods require massive offline computations and the retrieval systems are inflexible to database changes. Therefore, their effectiveness cannot be preserved in an updated retrieval database [30]. Zheng *et al*. [30] proposed a simple yet effective fusion method for image search and person re-identification. It is based on the hypothesis that the sorted score curve of a suitable feature takes on an "L" shape, whereas that of an unsuitable feature descends gradually. Nevertheless, according to Fig. 2, this hypothesis is unsuitable for RSIR.

To alleviate the RSIR problem that is described above in the existing methods, in this paper, we propose a query-adaptive remote sensing image retrieval method that is based on two image similarities. We use two similarities, namely,

the image rank similarity and the image-to-query class similarity, to improve the image similarity measure between the query image and retrieved images. A new query-adaptive weighting method is utilized to combine multiple features and enhance the retrieval precision.

III. OUR PROPOSED METHOD

Our proposed method mainly consists of four parts, which are illustrated in Fig. 1. First of all, we will introduce the CNN models and features, which used in our method. Secondly, we will present what the image rank similarity is and how to calculate it based on the retrieval result using Euclidean. And then, we will introduce our query-adaptive feature weighting method. Next, image-to-query class similarity will be introduced. Finally, we will summarize the procedure of the proposed method and analyze its computational complexity.

A. CNN MODELS AND FEATURES

The hierarchical architecture of CNN models can learn parameters automatically during the training process and can automatically obtain high-level visual features for efficiently representing images [31]. In recent years, many retrieval methods that are based on CNN models have been presented and are gradually replacing methods that are based on handcrafted low-level features [32]. Several successful CNN models are utilized to extract high-level features in this study: the famous VGGNet [33], GoogleNet [34], and ResNet [35].

VGGNet won the second rank in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC-2014). VGGNet includes several CNN models, such as VGG16 and VGG19. VGG16 has presented good performance in the image retrieval [15]–[17] [31], [36], [37].

GoogleNet is another CNN model that we selected for RSIR, which was the winner of ILSVRC-2014. It widely used for image recognition [17], [32], [36].

ResNet has achieved state-of-the-art performances in many computer vision tasks, which achieved the best performance on the ILSVRC-2015. ResNet includes some CNN models with different layers, and we chose the two models, Resnet50 and Resnet152, to extract features in our method. The two models have been widely used in image retrieval [32], [36], [37] and achieved better performance than other CNN features. We select the four features from these CNN models for RSIR according to related researches and our experimental results. Many works [15]–[17], [36], [37] have shown that these features can obtain good performance in RSIR.

1) Googlenet-cls3_pool. This feature is derived from the cls3_pool layer of GoogLeNet, which has dimensions of $1 \times 1 \times 1024$.
2) Resnet50-pool5. This feature is extracted from the pool5 layer of Resnet50 and has dimensions of $1 \times 1 \times 2048$;
3) Resnet152-pool5. The feature is extracted from the pool5 layer of Resnet152 and has dimensions of $1 \times 1 \times 2048$;
4) Vgg16-pool5. This feature is extracted from the pool5 layer of VGG16 and its dimensions are reduced to $1 \times 1 \times 2048$ via a region-based cascade pooling method [15].

B. IMAGE RANK SIMILARITY

Jaccard similarity is a common statistical measure that computes the similarity between two ranked lists based on their intersection and is defined by:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \tag{1}$$

The Jaccard similarity only considers the size of the intersection and neglects order information. We develop a simple way to integrate order information into Jaccard similarity for RSIR.

For two images I_i and I_j, we extract the corresponding feature vectors, namely, F_i and F_j, using a feature extractor $\mathcal{F}$, such as SIFT or CNN. The distance between the two images can be obtained by computing the distance $\mathcal{D}\left\{F_i, F_j\right\}$ between their feature vectors F_i and F_j according to a distance function $\mathcal{D}$, for example, the Euclidean or cosine distance function.

Let $C = \{I_1, I_2, \cdots, I_M\}$ be a reference image collection and I_q be a query image. We calculate the distance $\mathcal{D}$ between the query image I_q and each image in the collection C and obtain an $M \times 1$ distance vector P, where $P_{i,1} = \mathcal{D}\left\{F_q, F_i\right\}$ is the distance between I_q and I_i for $I_i \in C$. Then, by sorting the values $P_{i,1}$ in increasing order, we obtain a ranked list: $L_q = \{I_1, I_2, \cdots, I_M\}$. Typically, the top m ($m \ll M$) image list, namely, $R_q = \{I_1, I_2, \cdots, I_m\}$, is returned as the retrieval result for the query I_q, in which the most similar images to the query image in the collection C are listed first.

Consider a retrieved image I_r from the image collection C. We also acquire a top-m ranked image list, namely, R_r. The basic strategy of the image rank similarity is to measure the degree of similarity between I_q and I_r according to the two top-m image lists, namely, R_q and R_r, which are their retrieval results from the same retrieval collection. If I_q and I_r are similar, in the typical case, we observe that the top-m image lists R_q and R_r have many images in common. If not, we do not observe the same behavior. Therefore, the more similar their ranked results are, the more similar the two images are. In addition, we consider the number of images that appear in both ranked results. Furthermore, we believe that the ranking of images in these results is important for calculating the similarity of the two images.

Denote by a_i the rank of the i^{th} image in the image list R_q. If the i^{th} image is also contained in the image list R_r and b_i is its rank in R_r, we define d_i as in (2).

$$d_i = |a_i - b_i|. \tag{2}$$

If not, d_i is calculated via (3).

$$d_i = |a_i - 2m| . \tag{3}$$

Then, the image rank distance from R_q to R_r can be defined as (4).

$$D\left(\overrightarrow{R_q, R_r}\right) = \sum_{i=1}^{m} d_i \Big/ \left(\frac{(m-1) \times m}{2} + m \times m\right). \tag{4}$$

Via the same approach, we can calculate the image rank distance from R_r to R_q: $D\left(\overrightarrow{R_r, R_q}\right)$.

Finally, the image rank distance between R_q and R_r can be defined as in (5).

$$D\left(R_q, R_r\right) = \frac{D\left(\overrightarrow{R_q, R_r}\right) + D\left(\overrightarrow{R_r, R_q}\right)}{2}. \tag{5}$$

An increasing image rank distance corresponds to an increasing disagreement between the rankings. The distance is inside the interval [0, 1] and assumes the following values:

•0 if the agreement between the image rankings is perfect, namely, if the two image rankings are the same;

•1 if the image rankings are completely independent.

We use the image rank distance $D\left(R_q, R_r\right)$ to calculate the image rank similarity (IRS) coefficient, which is denoted as $IRS\ (q, r)$, between I_q and I_r as follows:

$$IRS\ (q, r) = 1 - D\left(R_q, R_r\right). \tag{6}$$

C. QUERY-ADAPTIVE FEATURE WEIGHTING

A score-level feature fusion scheme proposed in [30] is based on the observation that the profile of a suitable feature should exhibit an "L" shape while that of an unsuitable feature a gradually descending curve. In the scheme, the initial score cures are normalized according to reference curves trained on irrelevant data. Then, feature weight is estimated as inversely correlated with the area under the normalized score curve. We cannot observe the phenomenon in RSIR according to Fig. 2. Moreover, the method needs to build a huge reference collection. We propose a new query-adaptive feature fusion method for RSIR based on feature score cures.

In RSIR, a suitable feature for image retrieval can well distinguish the relevant images from the irrelevant images; hence, the relevant images have high similarities and the irrelevant images have low similarities. With an unsuitable feature, the relevant images and the irrelevant images have close similarities and difficult to distinguish. We retrieve two query images from the UC Merced Land-Use/Land-Cover dataset (UCMD) using the Googlenet-cls3_pool feature and the Vgg16-pool5 feature. The distance function is the Euclidean distance. The retrieval results are presented in Fig.2. According to Fig. 2(c), the Googlenet-cls3_pool feature, the average precision (AP) of which is 0.8361, is a suitable feature for the second query image. Its sorted score curve has a higher score in the head and a lower value in the tail than the sorted score curve of the Vgg16-pool5 feature (Fig. 2(d)), for which the AP is only 0.2922. We observe the same phenomenon as in Fig. 2(a) and (b), namely, the sorted score curve of the suitable feature is higher in the head and lower in the tail than that of the unsuitable feature. In addition, the Vgg16-pool5 feature is a suitable feature for the first query image, while it is an unsuitable feature for the second query image. For the Googlenet-cls3_pool feature, the result is the opposite.

For an image query I_q, image I_q is retrieved from the image collection C according to the feature $\mathcal{F}^i$ and the distance function $\mathcal{D}$. We obtain a top-l ranked image list, namely, $R_q^i = \{I_1, I_2, \cdots, I_l\}$, and an initially sorted score curve, namely, $S_q^i = \{s_1^i, s_2^i, \cdots, s_l^i\}$, with respect to feature $\mathcal{F}^i$, where s_l^i is the image rank similarity of the top l image of the feature $\mathcal{F}^i$, namely, $IRS\ (q, l)$, between image I_q and image I_l. We normalize the curve S_q^i via the following formula:

$$\bar{S}_q^i = \left(S_q^i - \min\left(S_q^i\right)\right)^2 \tag{7}$$

where $\bar{S}_q^i$ is the normalized score curve that is used to estimate the feature effectiveness. The initially sorted score curves of the two retrieved images in Fig. 2 are shown in Fig. 3(a) and (c). Fig. 3(b) and (d) present their corresponding normalized score curves. As shown in Fig. 3(b) and (d), after normalization, suitable features have a large area under the score curve. Therefore, we assume that the effectiveness of a feature is positively related to the area under its normalized score curve. To evaluate the assumption, we have collected satisfactory and unsatisfactory normalized score curves of the four features from UCMD. Satisfactory score curves are those for which AP exceeds 0.8 and unsatisfactory curves are those for which AP is smaller than 0.3. The probabilities of satisfactory or unsatisfactory normalized score curves are defined as the ratio of the number of satisfactory or unsatisfactory normalized score curves to all normalized score curves which area under score curves are in the same range. We compute the probabilities of satisfactory and unsatisfactory normalized score curves against the area under the normalized score curve. According to Fig. 4, the probability of an unsuitable feature decreases as the area under its normalized score curve increases. With this approach, we can estimate the effectiveness of a feature according to the area under its normalized score curve.

For an image query I_q with K features $\left\{\mathcal{F}^i\right\}_{i=1}^K$, we have K initial sorted score curves $\left\{S_q^i\right\}_{i=1}^K$. After curves $\left\{S_q^i\right\}_{i=1}^K$ have been normalized to $\left\{\bar{S}_q^i\right\}_{i=1}^K$, we calculate the query-adaptive weight of the feature F^i as follows:

$$w_q^i = \frac{A_i}{\sum_{k=1}^K A_k} \tag{8}$$

where $A_i, i = 1, \cdots, K$ is the area under the i^{th} feature's score curve.

To obtain a global similarity measure, we employ the sum rule to combine the scores of multiple features.

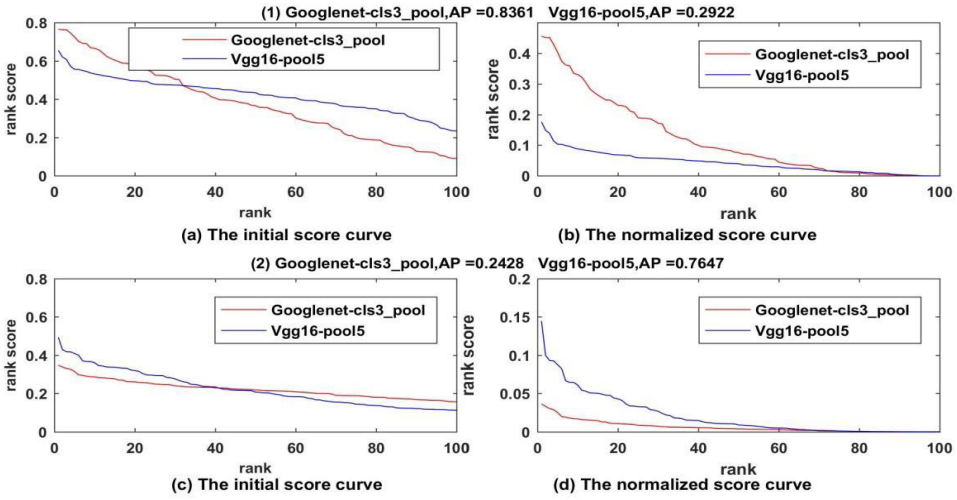

FIGURE 3. Comparison of the initial and normalized score curves. (a) and (c) are the initial score curves from Fig. 2 and (b) and (d) are the corresponding normalized score curves. In (b), AP of the Googlenet-cls3_pool feature is 0.8361, AP of the Vgg16-pool5 feature is 0.2922, and the area under the Googlenet-cls3_pool curve is greater than the area of under the Vgg16-pool5 curve. In (d), AP of the Googlenet-cls3_pool feature is 0.2428, AP of the Vgg16-pool5 feature is 0.7647, and the area under the Googlenet-cls3_pool curve is smaller than the area under the Vgg16-pool5 curve. Suitable features have a greater area under the score curve than unsuitable features.

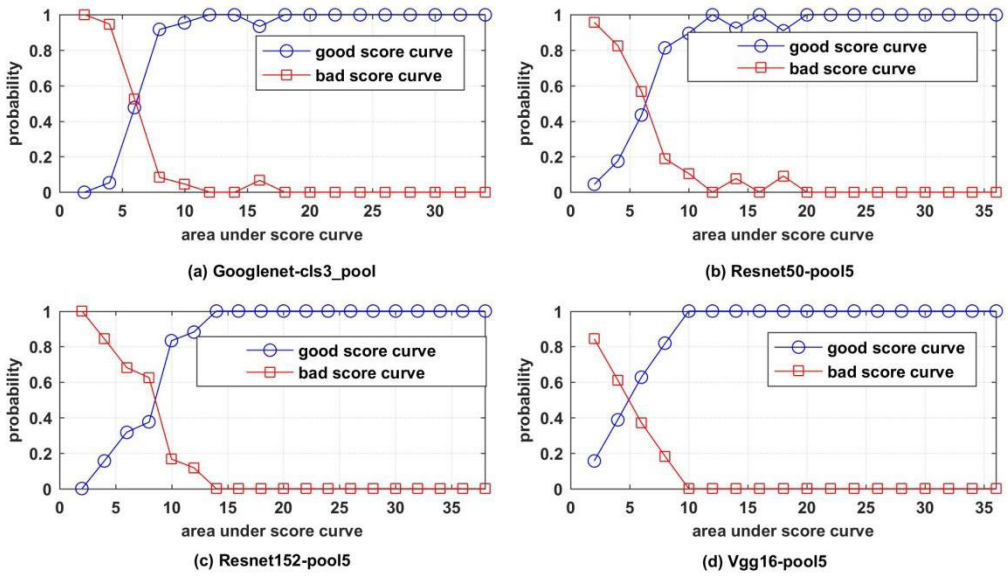

FIGURE 4. Probabilities of satisfactory and unsatisfactory normalized score curves against the area under the normalized score curve for the four features. The probability that a feature to be an unsuitable feature decreases as the area under its normalized score curve increases.

Given a retrieved image I_r, the image rank similarity score of I_r to I_q with respect to feature F^i is denoted as $IRS^i(q, r)$. Then, under the sum rule, the desired query-adaptive similarity (QAS) between q and r is calculated as follows:

$$QAS(q, r) = \sum_{i=1}^{K} w_q^i \times IRS^i(q, r), \quad where \ \sum_{i=1}^{K} w_q^i = 1. \tag{9}$$

D. IMAGE-TO-QUERY CLASS SIMILARITY

For a query image I_q and a retrieval database RD with N images, we calculate a query-adaptive similarity, namely, QAS, to I_q for each image in RD. We sort the images in descending order of their similarities and we select the top-k similar images as the retrieval result $\mathcal{R}_q$. If the similarity QAS is not perfect, irrelevant images that are not of the same class as image I_q may be found in the retrieval result $\mathcal{R}_q$. The objective of image retrieval is to retrieve the images of the

same scene or class from a database. Chen *et al.* [20] assumed that retrieved images should be similar to all the images that belong to the same class as the query image, not just to the query image. However, which images that belonged to the class of the query image are unknown. An iterative framework is used to find these images, in which the size of the query class is incrementally increased according to the previous retrieval results [20].

Here, we use the well-known k-nearest neighbor (kNN) method to identify images that may be the same class as the query image. Let $\mathcal{N}_{kNN}(q,k)$ denote the neighborhood set of the query image I_q that is obtained via this method and k is the size of the neighborhood set. It is defined as follows:

$$\mathcal{N}_{kNN}(q,k) = \{\mathcal{R} \subseteq RD, |\mathcal{R}| = k \wedge \forall x \in \mathcal{R}, y \in RD - \mathcal{R} : QAS(q,x) \geq QAS(q,y)\}. \quad (10)$$

Then, we take the query q and its neighbors $\mathcal{N}_{kNN}(q,k)$ as the query class. For a retrieved image I_r, we calculate the image-to-query class similarity (IQCS) between r and q:

$$IQCS(q,r) = \frac{1}{k+1}\left(QAS(q,r) + \sum\nolimits_{x \in \mathcal{N}_{kNN}(q,k)} QAS(r,x)\right). \quad (11)$$

E. IMAGE RETRIEVAL PROCESSING

To improve the speed of image retrieval, we divide our method into two parts: an offline part and an online part. The offline part is processed beforehand. On the offline part, we first finetune CNN models and extract CNN features of images in the retrieval database *RD*; then we generate the image collection C by randomly selecting a part of images from the retrieval database *RD*, retrieve every image from collection C using each feature and obtain four top-m image list sets; finally, we compute the query-adaptive similarity *QAS* matrix $\widehat{\boldsymbol{A}}$ between any two images in the retrieval database *RD*. On the online part, first of all, we calculate the image rank similarity between a query image and retrieved images; Then, we use the query-adaptive feature weighting method to combine these features; We sort the retrieved images according to the query-adaptive similarity *QAS*. Finally, we use the image-to-query class similarity *IQCS* to improve the retrieval result further. The online part is the retrieval process of a query image, which is shown in Fig. 1. The detail of the two parts are as follows:

1) THE OFFLINE PROCESS

Fine-tune

(1) the four pre-trained CNN models using a fine-tuning database and obtain the four fine-tuned CNN models.
(2) Randomly select a part of images on the retrieval database *RD* to build the image collection C.
(3) Extract four CNN features for every image in the image collection C using the four fine-tuned CNN models and obtain four feature sets: FC^i, for $i = 1, 2, 3, 4$. Then, retrieve every image from collection C using each feature and the Euclidean distance and obtain four top-m image list sets: $RC^i = \{R_1^i, R_2^i, \cdots, R_M^i\}$ for $i = 1, 2, 3, 4$.
(4) Apply the same approach as in collection C to every image in the retrieval database *RD* and obtain four top-m image list sets: $RS^i = \{R_1^i, R_2^i, \cdots, R_N^i\}$ for $i = 1, 2, 3, 4$.
(5) Calculate the *IRS* matrix $\boldsymbol{A}$ between any two images in RS^i, $i = 1, 2, 3, 4$, and obtain four *IRS* matrices: $\{A_1, A_2, A_3, A_4\}$.
(6) Compute the query-adaptive similarity *QAS* matrix $\hat{A}$ with size $N \times N$ between any two images in the retrieval database *RD*.

2) THE ONLINE PROCESS

(1) Given a query image I_q, use the four fine-tuned CNN models to extract the image features: $\left\{F_q^1, F_q^2, F_q^3, F_q^4\right\}$.
(2) Retrieve image I_q from the image collection C using the Euclidean distance with the four features and obtain four top-m image lists: $\left\{R_q^1, R_q^2, R_q^3, R_q^4\right\}$.
(3) Calculate and sort *IRS* between R_q^i and R_r^i in RC^i, and compute the weight set $\left\{w_q^1, w_q^2, w_q^3, w_q^4\right\}$ for the four features according to the four top-l image lists: $\left\{R_q^1, R_q^2, R_q^3, R_q^4\right\}$.
(4) Calculate *IRS* between R_q^i and R_j^i in RS^i for $i = 1, 2, 3, 4$ and $j = 1, 2, \cdots, N$ and obtain four *IRS* sets: $IS^i = \left\{IRS_{q,1}^i, IRS_{q,2}^i, \cdots, IRS_{q,N}^i\right\}$ for $i = 1, 2, 3, 4$.
(5) Compute the query-adaptive similarity *QAS* between q and each image in the retrieval database *RD*; $QS = \{QAS(q,j)\}$ for $j = 1, 2, \cdots, N$.
(6) Find the neighborhood set of the query image q, namely, $\mathrm{N}_{kNN}(q,k)$, from the retrieval database *RD*.
(7) Calculate the image-to-query class similarity *IQCS* between q and each image in *RD*, namely, $\{QAS(q,j)\}$, for $j = 1, 2, \cdots, N$, sort them in descending order and return the sorted result.

F. COMPUTATIONAL COMPLEXITY ANALYSIS

In the online part of our method, the time complexity is mainly attributed to five of the steps: (1) The time complexity of calculating the Euclidean distance in the second step is $O(lM)$, where l is the length of the feature and M is the size of the collection C. The time complexity of sorting is $O(M \log M)$. (2) The time complexity of calculating and sorting *IRS* in the third step is $O(mM)$ and $\mathrm{O}(M \log M)$, respectively. (3) Similarly, the time complexity of calculating in the fourth step is $O(mN)$. (4) The time complexity of the fifth step is $O(lN) + O(mN)$. (5) The time complexity of the sixth step is $O(N \log N)$. (6) In the seventh step, because the *QAS* between the query and each retrieved image has been calculated previously, the time complexity for calculating the image-to-query class similarity is $O(Nk)$ and that for sorting is $O(N \log N)$. Hence, the time complexity of the whole online part is $O(mN) + O(lN)$.

The time complexity of the offline part in our method mainly lies in calculating the four *IRS* matrices $\{A_1, A_2, A_3, A_4\}$. The time complexity of this step is $O(NNM)$.

The space complexity of each part mainly lies to store four *IRS* matrices $\{A_1, A_2, A_3, A_4\}$ and *QAS* matrix $\hat{A}$, it requires $O(N^2)$. We neglect the time for fine-tuning the CNN models and extracting the features.

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP

To evaluate the performance of our method, we consider two standard criteria of retrieval evaluation: the average normalized modified retrieval rank (ANMRR) [38] and the mean average precision (mAP) [39]. ANMRR considers the ranking information of relevant images among the top-retrieved images. ANMRR ranges from 0 to 1; a lower ANMRR value indicates a better retrieval performance [38]. The mAP is the average of Average Precision. Average Precision is the average of the precision value obtained for the set of top images existing after each relevant image is retrieved [39].

1) RS IMAGE DATABASES

We use three datasets below in our method.

(1) UCMD [40]: The UC Merced Land-Use/Land-Cover dataset is composed of 21 categories of land-use aerial images that were collected from the United States Geological Survey (USGS) national map. Each category is comprised of 100 images and each image has a size of 256×256 pixels.

(2) PatternNet [41]: The high-resolution RS image dataset is a recently released large-scale dataset, which contains 38 classes of RS scene images that were gathered from Google Earth imagery or via the Google Map API for the US cities. Each class has 800 samples with a size of 256×256 pixels.

(3) AID [42]: The Aerial Image Dataset is composed of 30 types of aerial scene images that were selected from Google Earth imagery. The numbers of sample images range from 220 up to 420 among the satellite scene types. The total number of samples in the AID dataset is 10000 and each image has a size of 600×600 pixels.

2) PREPROCESSING AND PARAMETER SETTINGS

Our method is implemented under Matconvnet [43] and MATLAB R2017a. It is run on a PC with an Intel CPU i7-7700 CPU @ 3.60 GHz, 16 GB of physical memory, and a graphics card GTX1080 with 8.0 GB of RAM. The machine is run on Ubuntu 14.04.

In our method, the main parameters depend on the expected number of relevant images, namely, τ, which can be estimated based on the database size and the number of classes. In the image rank similarity, the parameter m, which determines the image list length, is set to $c\_m \times \tau$, where c_m is a coefficient. The parameter l in the query-adaptive weight, which is the length of the score curves, is set to $c\_l \times \tau$. The parameter k in the image-to-query class similarity, which determines the number of neighbors of the query image is set to $c\_k \times \tau$. The following parameter values are used consistently for all evaluations: $c\_m = 0.6$, $c\_l = 1.1$, and $c\_k = 0.3$. Moreover, we consider the whole retrieval database *RD* as collection C.

We select AID for fine-tuning the four CNN models. The dataset is randomly split into training and testing data sets with about an 80%/20% split. Regarding the fine-tuning process, we adjust the number of classes of the outputs of the last fully connected or convolutional layer to match the number of AID classes, namely, 30. We randomly initialized the weights of the last layer according to a Gaussian distribution with mean 0 and variance 0.01. The weights are updated via the adaptive moment estimation (Adam) optimization algorithm [43] with a learning rate of 0.001, a momentum of 0.9, and a weight decay of 0.0005.

The UCMD and PatternNet are used to evaluate the retrieval performance. 20 percent of images on the two datasets are taken as query images, and the others are used as retrieval images.

B. ANALYSIS OF THE RETRIEVAL PERFORMANCE OF EACH STEP

Retrieving an image from an RS dataset via our proposed method involves four main steps: (1) Use the Euclidean distance to obtain the top-m image lists for each feature. (2) Calculate the image rank similarity score for each feature. (3) Compute the query-adaptive similarity. (4) Calculate the image-to-query class similarity. We conduct experiments to evaluate the retrieval performance of each step.

1) RETRIEVAL RESULT COMPARISON

The top-20 retrieval results of each step for two RS images are shown in Fig. 5 and Fig. 6. The first query image is selected from the UCMD. The top-20 retrieval results of Vgg16-pool5 using the Euclidean distance as the similarity are shown in Fig. 5(a), for which the AP is 0.3845 and which include 10 irrelevant images. Fig. 5(b) shows the result of Vgg16-pool5 that was obtained using the image rank similarity. Its AP increased by 0.3322 to reach 0.7167 and the number of irrelevant images decreased from 10 to 2. The results of fusing the four features via the query-adaptive fusion approach are shown in Fig. 5(c), the AP of which increased to 0.9007 and which include only one irrelevant image. After using the image-to-query class similarity, the AP of the final result is 0.9687 and there are no irrelevant images in Fig. 5(d). According to Fig. 5, each step in our proposed method improves the results over the previous step. This conclusion also can be drawn from Fig. 6, in which the query image is selected from PatternNet.

2) PERFORMANCE COMPARISON

We evaluate the retrieval performances on UCMD and PatternNet with mAP and ANMRR and the results are presented in Table 1. We observe positive gains with the

FIGURE 5. Retrieval results comparison in each step. The first column is the query image from UCMD, (a) the top 20 result of Vgg16-pool5 using Euclidean, (b) the top 20 results of Vgg16-pool5 using the image rank similarity, (c) the top 20 results of four features by the query-adaptive feature fusion, (d) the top 20 results of fusion feature using the image to query class similarity.

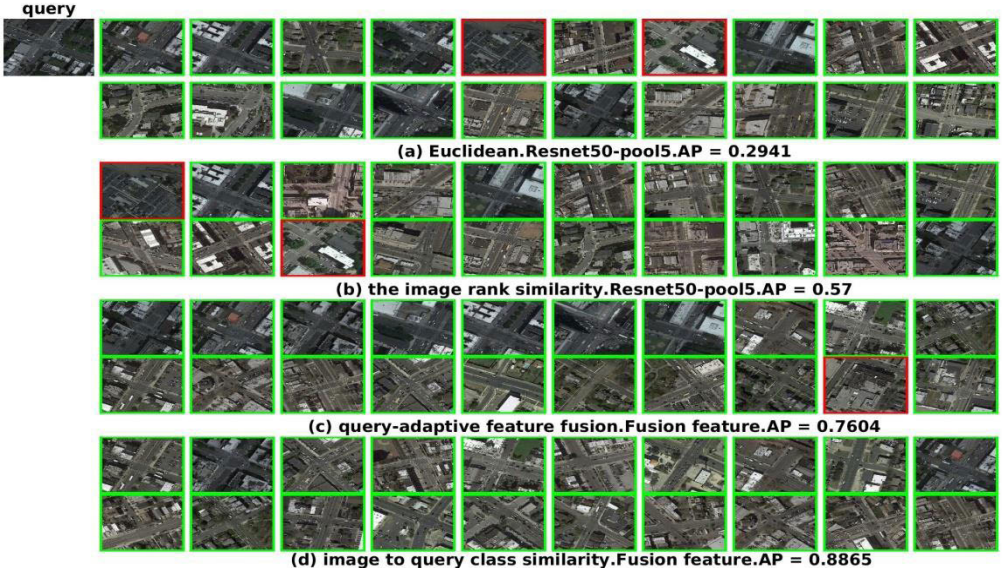

FIGURE 6. Comparison of the retrieval results in each step. The first column contains the query image from PatternNet. (a) The top-20 result of Resnet50-pool5 that were obtained using the Euclidean distance, (b) the top-20 result of Resnet50-pool5 using the image rank similarity, (c) the top-20 result of four features via query-adaptive feature fusion, and (d) the top-20 result of the fusion feature using the image-to-query class similarity.

mAP for all features in the image rank similarity step, which range from +4.86% to +9.23% in UCMD and from +8.66% to +14.82 in PatternNet. The ANMRR of all features also decreases by between −0.0495 and −0.0763 in UCMD and between −0.0751 and −0.1222 in PatternNet. Hence, the image rank similarity can greatly enhance retrieval accuracy compared to the Euclidean distance. In the query-adaptive feature fusion step, mAP increases by at least 8.01 percent in UCMD and by at least 5.55 percent in PatternNet and the ANMRR decreases by at least 0.0672 in UCMD and by at least 0.0469 in PatternNet. Therefore, our query-adaptive weight fusion method can improve retrieval

TABLE 1. Performance comparison for each step with mAP and ANMRR.

Method		UCMD		PatternNet	
		mAP (%)	ANMRR	mAP (%)	ANMRR
Euclidean distance	Googlenet-cls3_pool	63.13	0.296	67.40	0.2626
	Resnet50-pool5	66.27	0.2775	66.92	0.2671
	Resnet152-pool5	64.47	0.2900	65.73	0.2763
	Vgg16-pool5	63.36	0.2962	67.35	0.2640
Image rank similarity	Googlenet-cls3_pool	70.26	0.2300	79.58	0.1586
	Resnet50-pool5	71.13	0.2280	75.66	0.1920
	Resnet152-pool5	69.16	0.2397	74.39	0.2003
	Vgg16-pool5	72.59	0.2199	82.17	0.1418
Query-adaptive feature fusion		80.60	0.1527	87.72	0.0949
Image-to-query class similarity		83.69	0.1291	90.56	0.0759

TABLE 2. Computation time of each step.

Dataset	Euclidean distance	Image rank similarity	Query-adaptive feature fusion	Image-to-query class similarity	Total (ms)
UCMD	<1	8	<1	<1	10
PatternNet	80	1246	4	7	1337

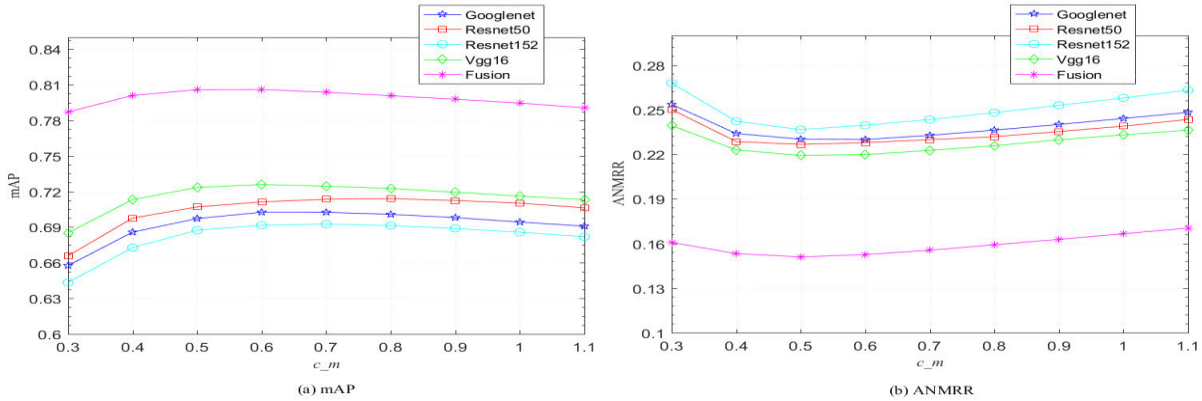

FIGURE 7. Impact of the value of parameter *c_m* on the retrieval performance with UCMD for various features. The results are (a) in terms of mAP and (b) in terms of ANMRR. If *c_m* is in the range [0.5 0.7], the retrieval performances are similar. The optimal value for most features is attained when *c_m* is approximately 0.6 in terms of mAP and 0.5 in terms of ANMRR.

performance. In the last step, the image-to-query class similarity further improves the mAP by approximately 3% and decreases the ANMRR by approximately 0.02 on both datasets.

3) COMPUTATION TIME

The computation time of each step is listed in Table 2. The total computation time of the main steps of our method is 10 milliseconds on UCMD and 1337 milliseconds on PatternNet. The image rank similarity step consumes most of the computation time. The time increases as the number of images in the retrieval dataset increases.

C. IMPACT OF THE PARAMETERS

To determine the optimal parameter values, we conducted a set of experiments on UCMD and PatternNet to evaluate the influence of the main parameters on the retrieval performance.

The parameter *c_m* in the image rank similarity varies in the interval [0.3 1.1] with a step size of 0.1. Fig. 7 and Fig. 8 present the results of the mAP and the ANMRR for various values of parameter *c_m* on both datasets. According to the two figures, their performances are similar if *c_m* is in the range [0.5 0.7] under both evaluation criteria. The best results are obtained for most features when *c_m* is approximately 0.6 in terms of mAP and 0.5 in terms of ANMRR.

The parameter *c_l* in the query-adaptive weight varies in the range [0.8 1.4]; the results are shown in Fig. 9. The retrieval performance is slightly affected by the value of parameter *c_l* in both datasets in terms of both evaluation criteria. The best result is obtained when *c_l* is approximately 1.1.

The parameter *c_k* in the image-to-query class similarity varies in the interval [0.1 1] with a step size of 0.1. Fig. 10 shows the results of the mAP and the ANMRR on UCMD and PatternNet for various values of parameter *c_k*.

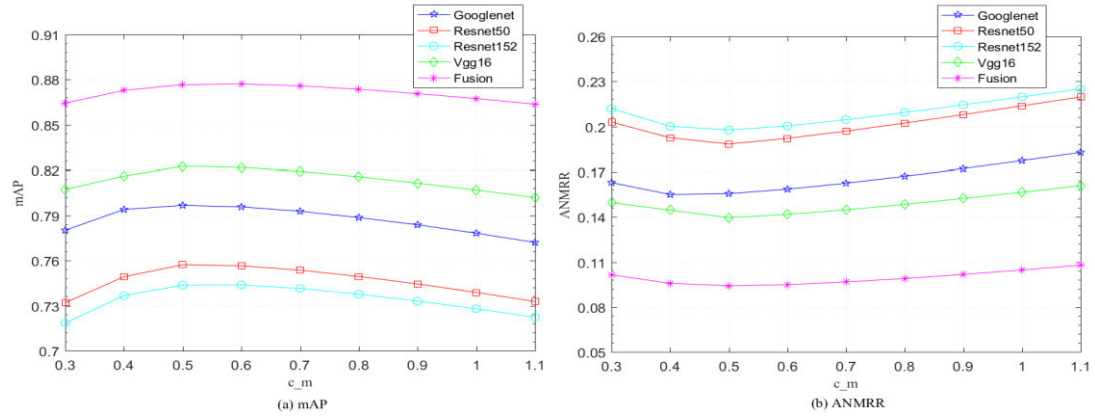

FIGURE 8. Impact of the value of parameter *c_m* on retrieval performance with Patternnet for various features. The results are (a) in terms of mAP and (b) in terms of ANMRR. The retrieval performances are similar when *c_m* is in the range [0.5 0.7]. The optimal value for most features is attained when *c_m* is approximately 0.6 in terms of mAP and 0.5 in terms of ANMRR.

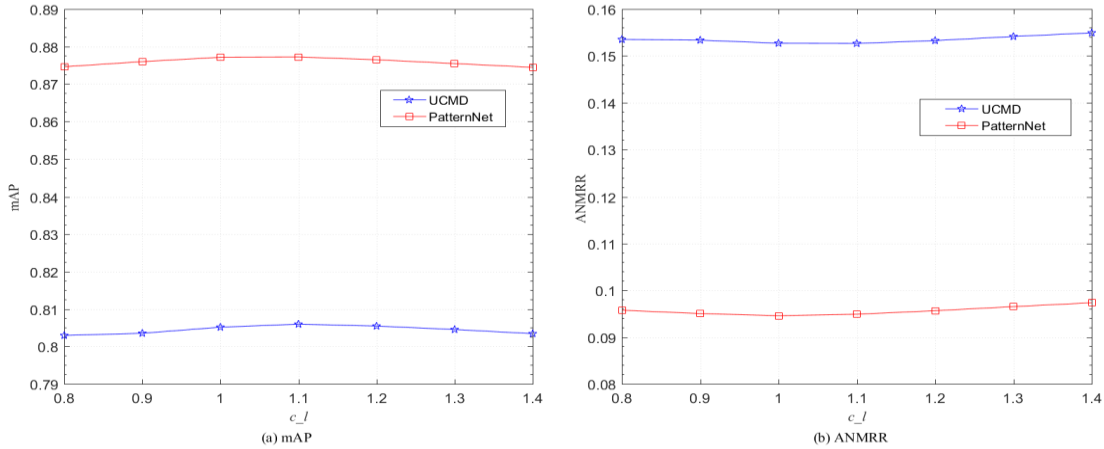

FIGURE 9. Impact of the value of parameter *c_l* on the retrieval performance for fused features on UCMD and PatternNet. The results are (a) in terms of mAP and (b) in terms of ANMRR. The retrieval performance is slightly affected by parameter *c_l* in both datasets. When *c_l* is approximately 1.1, the retrieval performance is optimal.

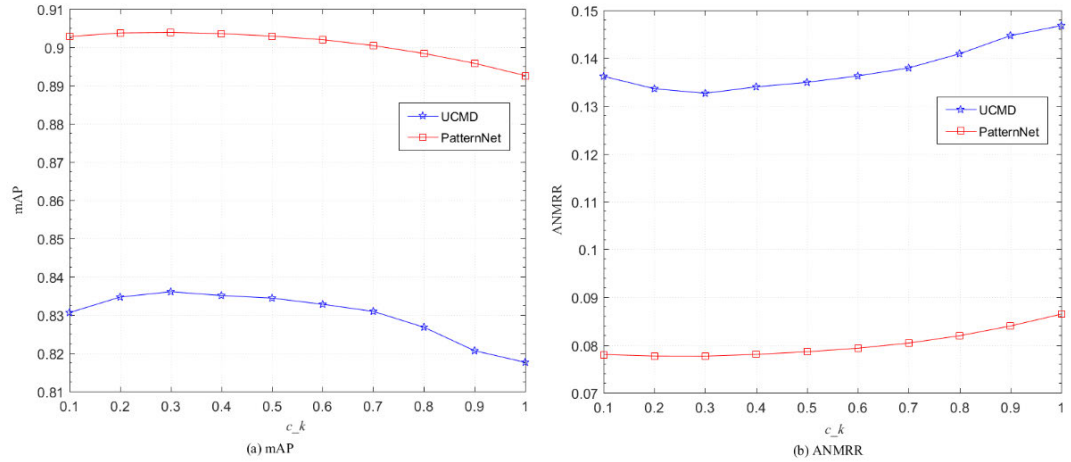

FIGURE 10. Impact of the value of parameter *c_k* on the retrieval performance for fused features on UCMD and PatternNet. The results are (a) in terms of mAP and (b) in terms of ANMRR. When *c_k* is 0.3, the retrieval performance is optimal on both datasets in terms of both evaluation criteria.

TABLE 3. Performance comparison for various sizes of the collection *C* with mAP and ANMRR on patternnet.

Method	Feature	40%		60%		80%		100%	
		mAP (%)	ANMRR	mAP (%)	ANMRR	mAP (%)	ANMRR	mAP (%)	ANMRR
Image rank similarity	Googlenet-cls3_pool	79.30	0.1586	79.51	0.1574	79.54	0.1573	79.58	0.1586
	Resnet50-pool5	75.11	0.1936	75.36	0.1921	75.40	0.1920	75.66	0.192
	Resnet152-pool5	73.84	0.2023	74.05	0.2008	74.12	0.2006	74.39	0.2003
	Vgg16-pool5	81.83	0.1458	82.06	0.1414	82.14	0.1408	82.17	0.1418
Query-adaptive feature fusion		87.46	0.0971	87.78	0.0942	87.77	0.0942	87.72	0.0949
Image-to-query class similarity		90.17	0.0785	90.39	0.0770	90.47	0.0767	90.56	0.0759

TABLE 4. Computation times for various sizes of the collection *C* on patternnet.

Percent	Euclidean distance	Image rank similarity	Query-adaptive feature fusion	Image-to-query class similarity	Total (ms)
40%	40	411	2	4	457
60%	58	690	3	5	756
80%	76	972	4	6	1058
100%	80	1246	4	7	1337

TABLE 5. Performance comparison of feature fusion methods on UCMD.

Feature Combination	Graph [22]		Global [17]		Ours	
	mAP	ANMRR	mAP	ANMRR	mAP	ANMRR
Googlenet-cls3_pool + Resnet50-pool5	74.87	0.1993	76.20	0.1873	**76.81**	**0.1812**
Resnet152-pool5+ Vgg16-pool5	76.66	0.1826	**78.12**	0.1728	78.06	**0.1716**
Googlenet-cls3_pool + Resnet152-pool5+ Vgg16-pool5	78.49	0.1703	79.18	0.1649	**79.69**	**0.1589**
Resnet50-pool5+ Resnet152-pool5+ Vgg16-pool5	77.09	0.1807	79.45	0.1635	**79.46**	**0.1621**
Googlenet-cls3_pool + Resnet50-pool5+Resnet152+ Vgg16-pool5	78.56	0.1693	80.12	0.1584	**80.60**	**0.1527**

The retrieval performance on both datasets is optimal when c_k is 0.3 in terms of both evaluation criteria.

In addition, we analyze the effect of the size of the image collection C on retrieval performance on PatternNet. We randomly select a subset (40%, 60%, and 80%) of all images in the retrieval database as the collection C. The precision results are listed in Table 3 and the computation times for various sizes of the collection C are listed in Table 4. According to the two tables, the precision increases slightly as the proportion increases; mAP increases from 90.17% to 90.56%, and ANMMR decreases from 0.0785 to 0.0759, while the computation time increases sharply from 457 milliseconds to 1337 milliseconds. The results are the average values over five runs.

D. COMPARISON WITH EXISTING METHODS

In this section, we compare the query-adaptive weight fusion method with other methods and compare our final retrieval results with the results of the other methods.

1) PERFORMANCE COMPARISONS WITH FEATURE FUSION METHODS

Our proposed query-adaptive weight fusion method is compared on UCMD with two methods: a graph-based query-specific fusion approach (Graph) [22] and a global method (Global) [17]. The main parameter, namely, k, in the graph-based query-specific fusion approach is set to 80, which is the true number of relevant images. The global method manually assigns a global weight w_i to each feature. We use a step size of 0.1 for manual tuning for each feature combination. According to Table 5, our method outperforms the other methods on all feature combinations in terms of ANMRR. Our method also outperforms the other methods in terms of mAP, except for the combination of Resnet152-pool5 and Vgg16-pool5. In addition, in our experiments, the global manual weight tuning is highly sensitive to weight changes: a small change in a feature weight may result in a substantial accuracy change. Our query-adaptive weight fusion method automatically determines feature weights and yields competitive results compared with the other two methods.

2) PERFORMANCE COMPARISONS WITH FEATURE SIMILARITIES

We compare the image rank similarity with Jaccard and RBO on UCMD and the results are shown in Table 6. The parameter p in RBO is set as 0.99. It can be seen that RBO can get the best results when only using a single feature. When query-adaptive feature fusion is used, IRS can get the best result in terms of mAP, 80.60%, while RBO can the best result in terms of ANMMR, 0.1523. The speed of RBO is very much slower than the other methods.

3) PERFORMANCE COMPARISONS WITH STATE-OF-THE-ART METHODS

We compare our proposed method with the state-of-the-art RS image retrieval methods on UCMD, which is a

TABLE 6. Performance comparison of RBO and IRS on UCMD.

Feature	Jaccard			RBO [29]			IRS		
	mAP (%)	ANMRR	Time (ms)	mAP (%)	ANMRR	Time(ms)	mAP (%)	ANMRR	Time (ms)
Googlenet-cls3_pool	69.95	0.2321	1	70.31	0.2275	12	70.26	0.2300	2
Resnet50-pool5	70.80	0.2309	1	71.38	0.2231	12	71.13	0.2280	2
Resnet152-pool5	68.87	0.2430	1	69.53	0.2330	12	69.16	0.2397	2
Vgg16-pool5	72.07	0.2237	1	72.74	0.2173	12	72.59	0.2199	2
Query-adaptive feature fusion	80.18	0.1566	4	80.35	0.1523	48	80.60	0.1527	8

TABLE 7. Performance comparison with state-of-the-art methods on UCMD.

Method	ANMRR
Bosilj *et al.* [45]	0.4720
Aptoula *et al.* [6]	0.5914
Du *et al.* [46]	0.5556
Tola *et al.* [8]	0.4604
Ge *et al.* [15]	0.2889
Wang *et al.* [18]	0.4317
Xia *et al.* [23]	0.2850
Zhou *et al.* [16]	0.3940
Ours	**0.1291**

benchmark test dataset that is used in most related works. The first four methods in Table 7 are based on hand-crafted feature representations, e.g., BOW and LSL, and the others use CNN features. According to Table 7, our proposed method yields the best results in terms of ANMRR. The ANMRR value decreases from 0.285 to 0.1291, which corresponds to a decrease of approximately 54%. Overall, our proposed method is promising and can realize higher retrieval performance.

V. CONCLUSIONS

In this paper, we propose a query-adaptive remote sensing image retrieval method that is based on two image similarities. We utilize the image rank similarity to measure the similarity for each feature between a query image and each retrieved image, which considers the number and image rank of the common images in their corresponding top-m image lists. Then, these similarities are fused via the query-adaptive weighting method, which calculates weights on the fly and is independent of the retrieval database. Finally, we obtain the neighborhood set of the query image as the query class via a k-nearest neighbor method and calculate the image-to-query class similarity between the query image and each retrieved image. We re-rank them to obtain the final retrieval result. Experiments in which the performance of each step was analyzed were conducted and the results demonstrate that the precision in the current step is higher than those in the previous steps. Therefore, the proposed method is effective for remote sensing image retrieval. In addition, we investigated the influences of various values of important parameters on retrieval performance. Comparisons of the proposed method with the state-of-the-art methods further demonstrated the strength of our method, which realizes highly competitive retrieval performance.

In future work, we will focus on (i) considering other CNN models [29] and image features, (ii) combining our query-adaptive weighting method with other supervised methods [47], and (iii) considering an iterative approach that utilizes the image rank similarity and the image-to-query class similarity.

REFERENCES

[1] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2486–2498, May 2017.
[2] E. Aptoula, "Remote sensing image retrieval with global morphological texture descriptors," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 3023–3034, May 2014.
[3] D. Yanqing, Y. Guoqing, and Z. Yanjie, "Remote sensing image content retrieval based on frequency spectral energy," *Procedia Comput. Sci.*, vol. 107, pp. 448–453, 2017.
[4] G. J. Scott, M. N. Klaric, C. H. Davis, and C.-R. Shyu, "Entropy-balanced bitmap tree for shape-based object retrieval from large-scale satellite imagery databases," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1603–1616, May 2011.
[5] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: A decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, May 2018.
[6] Y. Yang and S. Newsam, "Geographic image retrieval using local invariant features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 818–832, Feb. 2013.
[7] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.
[8] S. Ozkan, T. Ates, E. Tola, M. Soysal, and E. Esen, "Performance analysis of state-of-the-art representation methods for geographical image retrieval and categorization," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 11, pp. 1996–2000, Nov. 2014.
[9] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the Fisher kernel for large-scale image classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2010, pp. 143–156.
[10] W. Min, M. Fan, X. Guo, and Q. Han, "A new approach to track multiple vehicles with the combination of robust detection and two classifiers," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 174–186, Jan. 2018.
[11] Q. Wang, J. Gao, and Y. Yuan, "Embedding structured contour and location prior in Siamesed fully convolutional networks for road detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 230–241, Jan. 2018.
[12] Y. Yu and F. Liu, "Aerial scene classification via multilevel fusion based on deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 287–291, Feb. 2018.
[13] W. Min, H. Cui, H. Rao, Z. Li, and L. Yao, "Detection of human falls on furniture using scene analysis based on deep learning and activity characteristics," *IEEE Access*, vol. 6, pp. 9324–9335, 2018.
[14] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.

[15] Y. Ge, Y. Tang, S. Jiang, L. Leng, S. Xu, and F. Ye, "Region-based cascade pooling of convolutional features for HRRS image retrieval," *Remote Sens. Lett.*, vol. 9, no. 10, pp. 1002–1010, Aug. 2018.
[16] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval," *Remote Sens.*, vol. 9, no. 5, p. 489, May 2017.
[17] Y. Ge, S. Jiang, Q. Xu, C. Jiang, and F. Ye, "Exploiting representations from pre-trained convolutional neural networks for high-resolution remote sensing image retrieval," *Multimedia Tools Appl.*, vol. 77, no. 13, pp. 17489–17515, Jul. 2018.
[18] Y. Wang, L. Zhang, X. Tong, L. Zhang, Z. Zhang, H. Liu, X. Xing, and P. T. Mathiopoulos, "A three-layered graph-based learning approach for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6020–6034, Oct. 2016.
[19] D. C. G. Pedronette and R. da S. Torres, "Image re-ranking and rank aggregation based on similarity of ranked lists," *Pattern Recognit.*, vol. 46, no. 8, pp. 2350–2360, Aug. 2013.
[20] J. Chen, Y. Wang, L. Luo, J.-G. Yu, and J. Ma, "Image retrieval based on image-to-class similarity," *Pattern Recognit. Lett.*, vol. 83, pp. 379–387, Nov. 2016.
[21] M. Wang and T. Song, "Remote sensing image retrieval by scene semantic matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2874–2886, May 2013.
[22] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas, "Query specific rank fusion for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 4, pp. 803–815, Apr. 2015.
[23] X.-Y. Tong, G.-S. Xia, F. Hu, Y. Zhong, M. Datcu, and L. Zhang, "Exploiting deep features for remote sensing image retrieval: A systematic investigation," 2017, *arXiv:1707.07321*. [Online]. Available: http://arxiv.org/abs/1707.07321
[24] B. Chaudhuri, B. Demir, L. Bruzzone, and S. Chaudhuri, "Region-based retrieval of remote sensing images using an unsupervised graph-theoretic approach," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 987–991, Jul. 2016.
[25] M. Graña and M. A. Veganzones, "An endmember-based distance for content based hyperspectral image retrieval," *Pattern Recognit.*, vol. 45, no. 9, pp. 3472–3489, Sep. 2012.
[26] M. A. Veganzones, M. Datcu, and M. Graña, "Dictionary based hyperspectral image retrieval," in *Proc. Int. Conf. Pattern Recognit. Appl. Methods*, Vilamoura, Portugal, Feb. 2012, pp. 426–432.
[27] Y. Chen, X. Li, A. Dick, and R. Hill, "Ranking consistency for image matching and object retrieval," *Pattern Recognit.*, vol. 47, no. 3, pp. 1349–1360, Mar. 2014.
[28] C. Y. Okada, D. C. G. Pedronette, and R. da S. Torres, "Unsupervised distance learning by rank correlation measures for image retrieval," in *Proc. 5th ACM Int. Conf. Multimedia Retr. (ICMR)*, 2015, pp. 331–338.
[29] W. Webber, A. Moffat, and J. Zobel, "A similarity measure for indefinite rankings," *ACM Trans. Inf. Syst.*, vol. 28, no. 4, pp. 20.1-20.38, Nov. 2010.
[30] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, "query-adaptive late fusion for image search and person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1741–1750.
[31] F. Ye, W. Luo, M. Dong, H. He, and W. Min, "SAR image retrieval based on unsupervised domain adaptation and clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1482–1486, Sep. 2019.
[32] P. Napoletano, "Visual descriptors for content-based retrieval of remote-sensing images," *Int. J. Remote Sens.*, vol. 39, no. 5, pp. 1343–1376, Mar. 2018.
[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Sep. 2015, pp. 1–14.
[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[36] F. Ye, M. Dong, W. Luo, X. Chen, and W. Min, "A new re-ranking method based on convolutional neural network and two image-to-class distances for remote sensing image retrieval," *IEEE Access*, vol. 7, pp. 141498–141507, 2019.
[37] F. Ye, H. Xiao, X. Zhao, M. Dong, W. Luo, and W. Min, "Remote sensing image retrieval using convolutional neural network features and weighted distance," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1535–1539, Oct. 2018.
[38] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
[39] T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: An experimental comparison," *Inf. Retr.*, vol. 11, no. 2, pp. 77–107, Apr. 2008.
[40] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS)*, Nov. 2010, pp. 270–279.
[41] W. Zhou, S. Newsam, C. Li, and Z. Shao, "PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 197–209, Nov. 2018, doi: 10.1016/j.isprsjprs.2018.01.004.
[42] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.
[43] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia (MM)*, Oct. 2015, pp. 689–692.
[44] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Sep. 2015, pp. 1–13.
[45] P. Bosilj, E. Aptoula, S. Lefèvre, and E. Kijak, "Retrieval of remote sensing images with pattern spectra descriptors," *ISPRS Int. J. Geo-Inf.*, vol. 5, no. 12, p. 228, Dec. 2016.
[46] Z. Du, X. Li, and X. Lu, "Local structure learning in high resolution remote sensing image retrieval," *Neurocomputing*, vol. 207, pp. 813–822, Sep. 2016.
[47] Q. Wang, J. Gao, and Y. Yuan, "A joint convolutional neural networks and context transfer for street scenes labeling," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1457–1470, May 2018.

FAMAO YE received the B.E. degree in surveying and mapping engineering from the East China University of Technology, China, in 2000, the M.E. degree in surveying and mapping engineering from Wuhan University, China, in 2003, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Application, Chinese Academy of Sciences, China, in 2006. From 2006 to 2019, he was an Associate Professor with the School of Information Engineering, Nanchang University. He is currently an Associate Professor with the School of Surveying and Mapping Engineering, East China University of Technology. His research interests include artificial intelligence and remote sensing image processing.

XUQING ZHAO received the M.S. degree from Nanchang University, Nanchang, China, in 2019. She is currently a Teaching Assistant with the Jiangxi Technical College of Manufacturing. Her research interests include machine learning, computer vision, and data mining.

WEI LUO received the B.E. degree in information management and information system from Nanchang University, China, in 2017, where he is currently pursuing the master's degree in remote sensing image retrieval.

DAJUN LI received the B.E. degree in surveying and mapping engineering from the East China University of Technology, China, in 1990, the M.E. degree in surveying and mapping engineering from the Hefei University of Technology, China, in 1998, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, China, in 2003. Since 2003, he has been a Professor with the School of Surveying and Mapping Engineering, East China University of Technology, where he is currently the Dean. His current research interests include geographic information systems, remote sensing image processing, and artificial intelligence.

WEIDONG MIN (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in computer application from Tsinghua University, China, in 1989, 1991, and 1995, respectively. From 1994 to 1995, he was an Assistant Professor with Tsinghua University. From 1995 to 1997, he was a Postdoctoral Researcher with the University of Alberta, Canada. From 1998 to 2014, he was a Senior Researcher and a Senior Project Manager with Corel and other companies in Canada. From 2011 to 2014, he was with the School of Computer Science and Software Engineering, Tianjin Polytechnic University, China. Since 2015, he has been a Professor with Nanchang University, China, where he is currently a Professor and the Dean of the School of Software. His current research interests include image and video processing, artificial intelligence, big data, distributed systems, and smart city information technology. He is also an Executive Director of the China Society of Image and Graphics.