RESEARCH ARTICLE

Cross-Domain Remote Sensing Image Retrieval with Gabor-based CNN

Zheng Zhuo and Zhong Zhou*

State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, P.R.China

ARTICLE HISTORY
Compiled January 6, 2023

ABSTRACT
Domain adaptation is the ability to improve the learning efficiency of the target domain by using the prior knowledge of the source domain. When applied to new tasks in the target domain, the performance of domain adaptation trained in the source domain declines sharply. The purpose is to realize a new retrieval task in the target domain by using only a small number of labeled data samples, with the aid of the prior knowledge learned in the source domain. The research of this paper focuses on semi-supervised domain adaptation remote sensing image retrieval. The contributions of this paper are threefold. First, we construct Gabor-based CNNs to facilitate the networks to effectively capture the texture information of images. Second, we propose a cross-domain knowledge transfer strategy based on dual Gabor neural network learning. Third, we propose an unsupervised random feature mapping method based on probability distance. A large number of experiments have been conducted on UCM, WHU-RS, RSSCN7, AID, and PatternNet datasets. The results show that this method greatly improves the retrieval accuracy on the target domain and obtains state-of-the-art retrieval accuracy. The source code is available at http://nave.vr3i.com.

KEYWORDS
high-resolution remote sensing image retrieval; Gabor; cross-domain

1. Introduction

Remote sensing image retrieval is an important branch of image retrieval. Compared with natural images, the perspective of remote sensing images is looking down from a high altitude, with a large field of view. Several millions of remote sensing images have been delivered by various satellite sensors and stored in massive archives. Therefore, the texture of remote sensing images is relatively rich in both object classes and sensing variation. These features bring great challenges to the retrieval of remote sensing images(Zhao et al. 2021).

In recent years, researchers have used deep learning for remote sensing image retrieval and made considerable research progresses(Guo et al. 2020). Existing research works generally design corresponding retrieval mechanisms and strategies according to the needs of a specific task and the characteristics of datasets. These works greatly improve the retrieval accuracy and alleviate the bottleneck of massive remote sensing

CONTACT Zhong Zhou. Email: zz@buaa.edu.cn

image information extraction and sharing to a certain extent (Bateni et al. 2020).

A large number of annotation samples is the premise to achieve superior performance in deep learning. In many practical applications, remote sensing image labeling is difficult, time-consuming, and laborious. The number of labeled samples is limited. The trained depth neural network models are lack of generalization ability. When applied to new tasks in the target domain, the performance of CNN model trained in the source domain is greatly decline due to the change of sample distribution. However, data acquisition and tagging and network training for new tasks will require a lot of time and labor costs. Therefore, how to make full use of existing remote sensing image annotation data has become a research hot-spot in the field of remote sensing image retrieval (Guan et al. 2020).

This paper focuses on semi-supervised cross-domain remote sensing image retrieval, which uses the labeled Remote sensing image samples in the source domain to learn a priori knowledge. The target category, scene, and target domain of the remote sensing image in the source domain are not the same. Only some samples in the source domain are labeled, while only 10% of the samples in the target domain are labeled. By designing a knowledge migration strategy and an unsupervised random feature mapping method, it can achieve efficient retrieval only with the help of a small amount of sample label information in the target domain.

2. Related works

Domain adaptation is a special case of transfer learning. Its key points are 1) How to use the labeled data of the source domain to learn the prior knowledge related to the task. 2) How to design effective knowledge transfer learning strategies and use the prior knowledge of the source domain to guide the training and learning of the new task model in the target domain. The source domain and the target domain usually belong to the same task, but the data distribution and sample category are not the same.

According to the implementation stage of domain adaptation, existing domain adaptation methods can be divided into three categories (Pan and Yang 2010): 1) Sample level domain adaption. These methods find the data similar to the target domain in the source domain, adjust the weight of the data to match the new data with the data in the target domain, and then increase the weight of the sample to increase the proportion in predicting the target domain. The advantages are a simple method and easy implementation. It's simple and easy to implementation.The disadvantage lies in the selection of weight and the measurement of similarity depend on experience. Representative algorithms include ADDA (Tzeng et al. 2017), data distribution alignment (Zhao et al. 2021), DualGAN (Li et al. 2021), etc. 2) Feature level domain adaption. These method project the source domain and target domain into a common feature space. In this space, the source domain data and the target domain data have the same data distribution. Its key is how to design feature mapping rules. Representative algorithms include domain alignment (Guan et al. 2020), learning-to-learn (Tseng et al. 2020). 3) Model level domain adaption. The network model trained in the source domain is migrated to the target domain so that it can adapt to the characteristics of the data in the target domain and improve the learning performance of the model. The key to this method is how to design the migration strategy. Representative algorithms include IBN-Net (Pan et al. 2018), L2A (Luo et al. 2020), SIDHCNNs(Li et al. 2018) etc.

In recent years, with the rapid development of deep learning, the performance of remote sensing image retrieval based on deep learning has made a breakthrough(Zhuo and Zhou 2020), and the retrieval performance in some single domains has even been close to 100%(Zhuo and Zhou 2021). However, when these retrieval methods are applied to cross-domain new task retrieval, the retrieval accuracy will be greatly reduced. Therefore, cross-domain remote sensing image retrieval has aroused people's research interest.

For hyperspectral image classification of different sources, Chen et al (Chen et al. 2021) proposed a small sample learning network, which can narrow the distance within classes and push the distance between classes through the metric space, to realize the hyperspectral classification task across sources. Li et al (Li et al. 2021) proposed a discriminant feature learning model based on an adaptive matching network. Firstly, the "discriminant region" in a high-resolution remote sensing image is obtained by using a space spectral attention mechanism. Secondly, combined with ensemble learning adaptive learning matching network, unknown sample recognition under the condition of small samples is realized. Zhang et al (Zhang et al. 2021) proposed a small sample classification method for high-resolution aerial images. Based on meta-learning, this method proposed a partitioning strategy of meta-space, which can obtain a stable and transferable classification model under the condition of small samples.

Existing research works on cross-domain remote sensing image processing mainly focus on two aspects: one is how to improve the feature learning ability of the deep network so that the network can effectively learn a priori knowledge from the labeled data in the source domain; The second is to focus on how to design an effective migration strategy so that the knowledge learned from the source domain can be well migrated to the new task of the target domain.

There are still some problems in existing research. 1) There is no feature extraction method designed for the rich texture information of high-resolution remote sensing images. 2) Domain adaptive methods rely too much on empirical knowledge, and empirical parameters greatly affect the performance of cross-domain retrieval. 3) The algorithm has limitations. The target categories and scenarios of the source domain and target domain need to be the same, and the application is limited.

There are three main contributions of this paper:

(1) According to the rich texture of remote sensing images, two CNN network structures based on Gabor are proposed, which can effectively capture the texture of the images. Next, the labeled data of the source domain is used to train the two networks to express the task-related prior knowledge of the source domain. The knowledge obtained by the two networks is complementary to guide the learning of the target domain and improve the performance of cross-domain retrieval of remote sensing images.

(2) This paper proposes a cross-domain knowledge transfer strategy based on dual Gabor-based CNN network learning. In the source domain, two Gabor-based CNNs are trained by using the labeled data to obtain the trained network model. Next, the learned knowledge is transferred to the target domain to guide the feature learning of remote sensing images in the target domain. Specifically, after fine-tuning the two network models by using a small amount of sample label information of the target domain, it acts as a feature extractor to extract the features of the data of the target domain respectively, and then aggregate the extracted features to obtain the feature expression of the image.

(3) In this paper, an unsupervised random feature mapping method based on prob-

ability distance is proposed. This method uses t-distribution probability to measure the distance between two feature vectors, clusters the features by constructing a kNN tree and maps the features from high-dimensional space to low-dimensional space. It can not only effectively remove the redundant information in the feature expression of the target domain, but also narrow the distance within the class and widen the gap between classes.

3. The Proposed Cross-Domain Remote Sensing Image Retrieval Method

In this section, we introduce the overall image retrieval framework, the Gabor-based CNN construction method in our work, and the unsupervised feature mapping method based on probability distance.

3.1. *Overall Framework*

Different CNNs have different characteristic learning tendencies, and the prior knowledge about source domain tasks is also different. Therefore, we use two different CNN structures to extract the prior knowledge of the source domain and migrate it to the target domain to guide the learning of task features in the target domain. These two different features complement each other, which can give full play to the advantages of different networks, to improve the retrieval performance.

Based on the above ideas, this paper proposes a semi-supervised cross-domain remote sensing image retrieval framework, as shown in Figure 1. Firstly, we designed a controllable Gabor convolution layer, embedded it into CNN network, constructed two networks respectively, trained the two networks respectively by using the annotation data of source domain, and the two network models can express the prior knowledge related to source domain and task from different angles, and complement each other. Next, the two network models are migrated to the target domain, and the model is fine-tuned by using a small amount of labeled data in the target domain so that they can adapt to the characteristics of the target domain data. The fine-tuned network is used as the feature extractor to extract the features of the remote sensing image in the target domain, and the two features are aggregated as the feature representation of the image. Finally, we design an unsupervised feature mapping method to map features from high-dimensional space to low-dimensional space, remove the redundant information contained therein, expand the gap between classes and narrow the gap within classes, so that the features have better discrimination ability, to improve the performance of image retrieval in the target domain.

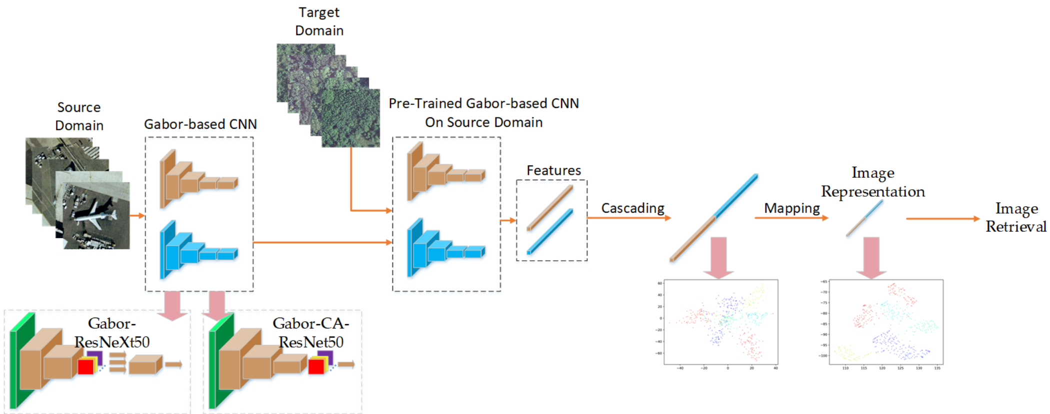

Figure 1. Semi-supervised cross-domain remote sensing image retrieval framework

Next, the implementation details of the Gabor-based CNN network, cross-domain knowledge migration, and feature dimensionality reduction will be introduced.

3.2. *Gabor-based CNN*

Anisotropic filtering technology is widely used to extract robust image features, and Gabor is the most representative filter (Luan et al. 2018). Gabor transform uses a group of Gabor filters with different frequency domain characteristics as the basis function for image transformation. Each channel can obtain some local characteristics of the input image and can describe the space-frequency structure in the image while retaining the spatial relationship information, which has multi-resolution characteristics.

Gabor transform have significant advantages in extracting local space-frequency domain features of targets because it is similar to the visual stimulus-response of human. Gabor is widely used in image processing, pattern recognition, and other fields.

3.2.1 Controllable Gabor Convolution

In this paper, a controllable Gabor convolution layer is added to CNN (Ozbulak et al. 2018). There are 5 important parameters of the Gabor kernel function, wavelength, direction, phase offset, standard deviation, and spatial aspect ratio. Through a large number of experiments, we find that the parameters such as phase offset, the standard deviation of Gaussian Factor, and spatial aspect ratio have little impact on the retrieval performance, while the direction of Gabor kernel function has a direct impact on the retrieval performance. In order not to increase the number of training parameters, this paper only considers the cases where the direction of Gabor kernel function is 0°, 45°, 90°, and 135°.

We introduce the Gabor convolution layer into ResNeXt and ResNet network based on channel attention mechanism respectively, and construct Gabor-ResNeXt and Gabor-CA-ResNet structures respectively, as shown in Figure 2 and Figure 3.

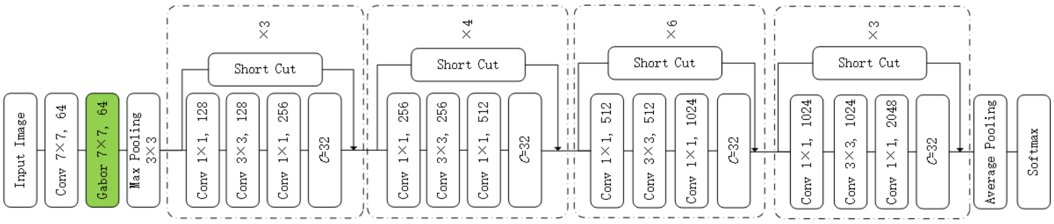

Figure 2. Gabor-ResNeXt structure

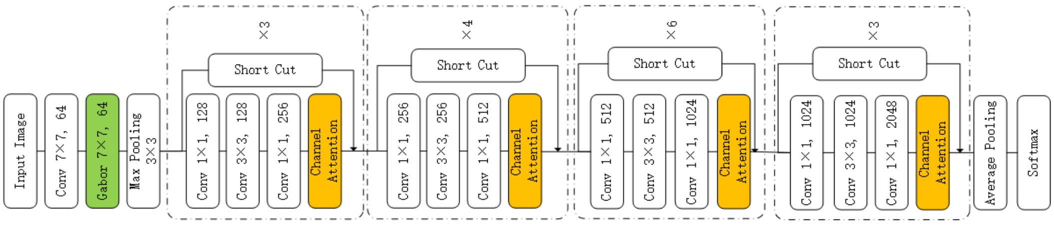

Figure 3. Gabor-CA-ResNet structure

After the input image passes through the general convolution layer once, the low-level features are enhanced through the controllable Gabor convolution layer, and the enhanced features are sent to the residual network. In the Gabor-ResNeXt network, each residual block is divided into 32 groups for convolution, which is combined and sent to the next residual block. In the Gabor-CA-ResNet network, each residual block contains a channel attention mechanism to further enhance the features. Finally, the two networks process the features of the last convolution layer through the average pooling layer, input them into Softmax, and output them as the classification results of the image. In this paper, the cross-entropy function is used to calculate the classification loss.

We extracted and visualized the feature map before and after Gabor convolution, as shown in Figure 4.

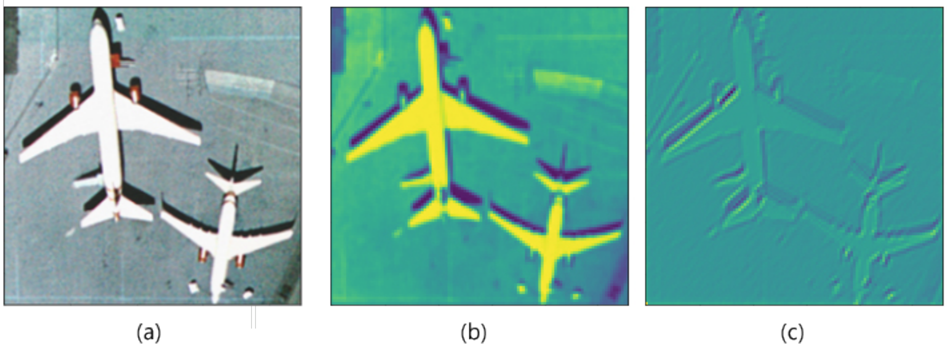

Figure 4. Gabor convolution feature visualization effect

(a) The original image (b) features extracted without Gabor convolution layer (c) feature map extracted after Gabor convolution layer is added

As can be seen from Figure 4, the edge of the feature map extracted after the Gabor convolution layer is more obvious. This figure shows that the Gabor convolution layer can effectively capture the texture which is more suitable for feature extraction of high-resolution remote sensing images.

3.2.2 Network Training

This paper adopts the strategy of "pre-training and fine-tuning" to train Gabor-ResNeXt and Gabor-CA-ResNet. Firstly, the ImageNet Deng et al. (2009) dataset is used to pre-train the ResNet network and ResNeXt network, and the model parameters of the ResNet network and ResNeXt network are obtained. Next, the Gabor-CA-ResNet network and Gabor-ResNeXt network are fine-tuned by using the labeled data of the source domain to obtain the optimized network model. The two network models contain a certain scale of parameters to express the prior knowledge of source domain tasks from different angles. The two models are represented by $W_1 = \left\{w_1^i\right\}, i = 1, 2, ..., C_1$ and $W_2 = \left\{w_2^i\right\}, i = 1, 2, ..., C_2$ respectively, where w_1^i and w_2^i represent the i-th convolution kernel weight of network 1 and network 2 respectively, and C_1 and C_2 represent the number of convolution kernels of the two networks respectively. The prior knowledge of the source domain is expressed by the weight set of the convolution kernel.

Next, the models W_1 and W_2 are migrated to the target domain as feature extractors to extract the features of remote sensing images in the target domain.

3.3. *Cross-Domain Knowledge Transfer Strategy*

Knowledge transfer refers to a learning process that uses the similarity of data, tasks, or models to apply the models learned in the wine field to new fields. It is to transfer the learned and trained model parameters to a new model for training. Considering that most of the data or tasks are relevant, the learned model parameters (also known as the knowledge learned by the model) can be shared with the new model in some way through migration learning, to speed up and optimize the learning efficiency of the model without learning from zero like most networks. The core problem of knowledge transfer is to find the similarity between the two fields. If we find this similarity, we can make rational use of it and complete the transfer learning task well.

Given a transfer learning task defined by $(D_s, T_s, D_t, T_t, f_T(\cdot))$, where D_s represents the dataset of the source domain and D_t represents the dataset of the target domain, in which a small number of data samples are labeled. T_s and T_t represent the tasks of the source domain and the target domain, respectively, transfer learning aims to improve the performance of thed predictive function $f_T(\cdot)$ for the target learning task T_t by discovering and transferring latent knowledge learned from D_s, which can be expressed as:

$$f(D_s \rightarrow T_s) \rightarrow T_t \tag{1}$$

in this paper, we propose a transfer learning strategy through reusing part of the network parameters W_1 and W_2, which had been well trained by D_s. It is well known that the features from low layers in CNN contain more low-level visual features which apply to all types of images, while the features from the high layers represent global semantic information for a specific domain. Therefore, we fine-tune the models of W_1 and W_2 by using D_T. To retain valuable knowledge of D_s, the parameters of the lower layers will be mostly frozen in the fine-tuning process, and only the fully connected layers and feature interaction layers will be fine-tuned to make the model converge, thus obtaining the networks which adapt to the characteristics of D_T. Transfer learning allows applying the powerful well pre-trained networks to obtain higher performance by using a relatively small amount of data samples, while reducing the training time by several orders of magnitude, and eliminating the requirements for optimizing hyper-parameters.

The two fine-tuned CNNs are used as the extractor to obtain the features of images in D_T, which are then aggregated together to obtain the final feature representation vector. The output of the last fully connected layer of each network is extracted, whose dimensions are both 2048. So the total dimension of the feature vector is 4096. There is redundant information among the feature vector. Therefore, we further propose an unsupervised feature mapping method based on probability distance to remove the redundant information, while increasing the distance between classes and reducing the distance within classes.

3.4. *Unsupervised Feature Mapping Based on Probability Distance*

We replace the traditional Euclidean distance measurement by calculating the probability distribution between samples, but in practical application, the sample distance is difficult to obtain. Generally, we can find the shortest path distance in the kNN graph as the approximation of the sample distance by constructing the kNN graph.

3.4.1 Construction of Efficient kNN Graph

kNN graph is a one-step composition process added to the classical kNN algorithm. Suppose there are n nodes in the space. For node v_i, find the k nearest neighbors v_1, v_2, ..., v_k, through some distance measurement (Euclidean distance, editing distance), and then connect v_i with these k neighbors to form k directed edges. All vertices in the space are processed in this way, and finally, the kNN graph is obtained.

Constructing an accurate kNN graph requires very high computational complexity, because the distance between two data points needs to be calculated, and the required computational complexity is $O(N2d)$. We adopt a neighbor search method, which is divided into two steps: the first step is to divide the space by using the random projection tree and then find the k nearest neighbor of each point to obtain an initial kNN graph. This kNN graph does not need to be completely accurate, which can accelerate the calculation speed of the probability value of sample points. In the second step, the neighbor search algorithm is used to find potential neighbors, calculate the distance between the neighbor and the current point, the neighbor's neighbor and the current point, put it into a small root heap, take the k nodes with the smallest distance as the k nearest neighbors, and finally get an accurate kNN graph.

3.4.2 Feature Mapping

In the original high-dimensional space, the current center point is regarded as the target, then the center point and its neighbor nodes constitute a positive sample, while the center point and non-neighbor points constitute a negative sample. The weights of positive samples are mapped using Gaussian distribution as:

$$p_{j|i} \frac{exp(-\left\|\vec{x_i}-\vec{x_j}\right\|^2/2\sigma_i^2)}{\sum_{(i,j\in E))} exp(-\left\|\vec{x_i}-\vec{x_k}\right\|^2\sigma_i^2)}, p_{i|i}=0 \tag{2}$$

$$\omega_{i|j}=\frac{p_{j|i}+p_{i|j}}{2N} \tag{3}$$

in low dimensional space, the coordinate position depends on the observation probability as:

$$P(e_{ij} = 1) = f(\|\vec{y_i} - \vec{y_j}\|) \tag{4}$$

$$P(e_{ij} = \omega_{ij}) = P(e_{ij} = 1)^{\omega_{ij}} \tag{5}$$

where $f(\cdot)$ is the distance function, which can be mapped by using the function with smaller distance and larger value. In other words, the closer the data points in high-dimensional space are, the closer they are in low-dimensional space. Compared with Gaussian distribution, t-distribution has significant advantages in dealing with small samples and outliers (Jones 2008). Therefore, the final optimization objective function can be defined as:

$$O = \prod_{(i,j)\in E} p(e_{ij} = 1)^{\omega_{ij}} \prod_{(i,j)\in E} (1 - p(e_{ij} = 1))^{\gamma}$$

$$\propto \sum_{(i,j)\in E} \omega_{ij}\ log\ p(e_{ij} = 1) + \sum_{(i,j)\in \overline{E}} \gamma\ log(1 - p(e_{ij} = 1)) \tag{6}$$

optimizing according to the above objective function requires a lot of computational overhead, because the number of negative edges is very large, and directly training all negative edges will increase the complexity. Therefore, we try to use a negative sampling algorithm for optimization. The negative sampling algorithm determines the probability of being sampled according to the degree of the corresponding node, that is, the greater the probability that the node with a greater degree is regarded as another node with a negative edge, and the probability meets the noise distribution:

$$p_n \propto d_j^{0.75} \tag{7}$$

thus, the objective function is transformed into:

$$O = \prod_{(i,j)\in E} \omega_{ij}\ log\ p(e_{ij} = 1) + \sum_{k=1}^{m} E_{jk} \sim P_n(j)\ \gamma\ log(1 - p(e_{ij} = 1)) \tag{8}$$

after using negative sampling and edge sampling optimization, we further use the asynchronous random gradient descent method for training, which effectively reduces the computational complexity of the algorithm.

3.5. *Similarity Measurement*

The Euclidean distance of feature vectors is adopted to measure the similarity of images. The Euclidean distance between two points $x_{1k}(k = 1, 2...n)$ and $x_{2k}(k = 1, 2...n)$ in N-dimensional space is defined as follows:

$$d_{12} = \sqrt{\sum_{k=1}^{n} (x_{1k} - x_{2k})^2} \tag{9}$$

Euclidean distance is widely used in image retrieval similarity measurement, and it is one of the most effective and widely used measurement methods.

4. Experimental Results and Analysis

To evaluate the performance of the proposed method, we have made comparison experiments on four high-resolution remote sensing image datasets including UCM, WHU-RS, RSSCN7, and AID. The experimental results are introduced and the results are analyzed.

4.1. *Datasets and Evaluation Metric*

PatternNet(Zhou et al. 2018), UCM(Yang and Newsam 2010), WHU-RS(Dai and Yang 2011), RSSCN7(Zou et al. 2015), and AID(Xia et al. 2017) are the five most commonly used high-resolution remote sensing image datasets. The information of five datasets is shown in Table 1.

Table 1. The information of five datasets

Dataset	Classes	Samples per Classes	Total Samples
PatternNet	38	800	30400
UCM	21	100	2100
WHU-RS	19	About 50	1005
RSSCN7	7	400	2800
AID	30	About 220-420	10000

The PatternNet dataset is a large-scale high-resolution remote sensing dataset collected for remote sensing image retrieval. Images in the PatternNet are collected from Google Earth images or through the Google Maps API for some American cities.

The images in the UCM dataset come from the U.S. Geological Survey's city map, which includes 21 types of scene images, such as airplanes, beaches, buildings, and dense residential areas.

The WHU-RS, RSSCN7, and AID datasets are released by Wuhan University in 2011, 2015, and 2017. The pixel size of each image is 600×600, and there are 19 types of scene images, each of them contains about 50 images, with a total of 1005 images. Because of the diversity of scenarios, the RSSCN7 dataset is very challenging.

We use the widely used mAP (mean average precision) (Deselaers et al. 2008) to evaluate retrieval performance.

4.2. *Experimental Setting*

In this paper, the dataset is randomly divided, five repeated experiments are carried out, and the average of the results is taken as the final experimental result. In the experiment, we randomly selected 20% of the samples from the PatternNet dataset to construct the dataset of the source domain, randomly selected 10% from each category of the other four datasets as the labeled data of the target domain, and the remaining 90% was used to test the retrieval performance.

This paper will build and experiment with the proposed network structure in Keras open-source framework. The computing hardware platform adopts Intel Core i7-8700, CPU 3.2GHz, 32GB memory, and an NVIDIA Geforce RTX 2080 Ti graphics card for training and testing. The number of training iterations is set to 50 rounds, the batch size is set to 20. The method of momentum and weight attenuation is used to optimize the training process to prevent overfitting. We use the stochastic gradient descent (SGD) optimizer with momentum 0.9, the learning rate is 0.0001.

When training and fine-tuning Gabor-CA-ResNet and Gabor-ResNeXt networks with source domain dataset and target domain dataset, the training samples are expanded. The expansion method is that the original image and its horizontal mirror image are rotated every 45°, and the expanded dataset is 16 times the original size.

4.3. *Comparison of Performance of Different Gabor-based CNN Structures*

To verify Gabor's ability of texture feature extraction, the Gabor convolution layer is introduced into ResNet, ResNeXt, and CA-ResNet respectively. Three network structures, Gabor-ResNet, Gabor-CA-ResNet, and Gabor-ResNeXt, are constructed and compared with ResNet, ResNeXt, and CA-ResNet. In the experiment, each dataset is randomly divided into a training set and test set according to the ratio of 1:9. The training data are expanded. The feature dimension extracted by each network is 2048, and the Euclidean distance measurement criterion is used for similarity comparison. The experimental results of six network structures on each dataset of UCM, WHU-RS, RSSCN7, and AID are shown in Table 2.

Table 2. Comparison of retrieval performance obtained by using different network structures

Networks	UCM	WHU-RS	RSSCN7	AID	PatternNet
ResNet	76.85%	64.30%	63.00%	72.21%	96.49%
Gabor-ResNet	78.52%	69.42%	64.94%	73.68%	96.79%
CA-ResNet	77.52%	67.30%	73.64%	80.98%	98.24%
Gabor-CA-ResNet	81.84%	72.13%	75.92%	81.65%	99.11%
ResNeXt	82.70%	65.33%	59.12%	81.74%	98.26%
Gabor-ResNeXt	83.38%	67.32%	72.67%	81.96%	98.42%

It can be seen from Table 2 that for different network structures, adding the Gabor convolution layer to four datasets can effectively improve the retrieval performance. Among them, Gabor-ResNet improves the retrieval performance by 1.67%, 5.12%, 1.94%, and 1.47%. Gabor-ResNeXt improves the retrieval performance by 0.68%, 1.99%, 13.55% and 0.22% respectively. Gabor-CA-ResNet improved the retrieval performance by 4.32%, 4.83%, 2.28% and 0.67% respectively. This shows that Gabor can effectively enhance the expression ability of depth network in texture, direction, and scale changes.

Compared with ResNet, the retrieval performance of the CA-ResNet network has been further improved by 0.67%, 3.00%, 10.64%, and 8.77% respectively. This shows that the channel attention mechanism further improves the expression ability of features. The depth features extracted by Gabor-CA-ResNet network architecture have strong expression and discrimination ability. Especially in the RSSCN7 dataset with few targets and rich texture, the performance is improved the most, reaching 10.64%.

The Table 2 shows that Gabor-ResNet has the lowest retrieval performance on PatternNet dataset, 96.79%, and Gabor-CA-ResNet and Gabor-ResNeXt are 2.32% and 1.63% higher than Gabor-ResNet respectively.

4.4. *Comparison of multi-network feature aggregation on retrieval performance in the target domain*

To verify the impact of multi-network feature aggregation on retrieval performance, we compared the performance of Gabor-ResNet, Gabor-CA-ResNet, and Gabor-ResNeXt networks after aggregation with the retrieval performance of the target domain obtained by using only a single network. In the experiment, the trained networks in the source domain are migrated to the target domain, the features of each network are extracted respectively and then aggregated. The aggregated features are used for retrieval. The feature dimension extracted by each network is 2048, the feature dimension after double network aggregation is 4096, and the feature dimension after three networks aggregation is 6144. Euclidean distance measurement criterion is used for similarity comparison. The comparative experimental results on UCM, WHU-RS, RSSCN7, and AID datasets are shown in Table 3.

Table 3. Impact analysis of feature aggregation on retrieval performance

Dataset	2048 dimension		
	Gabor-ResNet	Gabor-CA-ResNet	Gabor-ResNeXt
UCM	82.24%	85.64%	84.82%
WHU-RS	70.27%	72.83%	71.68%
RSSCN7	64.42%	72.27%	71.83%
AID	71.45%	78.36%	78.00%
Dataset	4096 dimension		
	Gabor-ResNet+ Gabor-CA-ResNet	Gabor-CA-ResNet+ Gabor-ResNeXt	Gabor-ResNet+ Gabor-ResNeXt
UCM	86.34%	85.95%	**87.81%**
WHU-RS	73.10%	72.86%	**75.07%**
RSSCN7	71.05%	71.41%	**75.15%**
AID	78.87%	77.91%	**82.12%**
Dataset	6144 dimension Gabor-ResNet+Gabor-CA-ResNet+Gabor-ResNeXt		
UCM	87.56%		
WHU-RS	74.73%		
RSSCN7	73.79%		
AID	80.92%		

It can be seen from the table, compared with using only a single network structure, aggregating the features of the two networks can effectively improve the retrieval performance. Among them, the feature aggregation of Gabor-CA-ResNet and Gabor-ResNeXt networks improves the most, with an increase of 2.17%, 2.24%, 2.88%, and 3.76% respectively on the four datasets. After the aggregation of the three network features, the performance decreases, which shows that the fusion of features with weak expression ability will increase the redundant information of the fused features and introduce noise interference, resulting in the decline of retrieval performance.

Therefore, this paper aggregates the characteristics of Gabor-CA-ResNet and Gabor-ResNeXt networks for the retrieval of the target domain.

4.5. *Comparison of Unsupervised Feature Mapping On Retrieval Performance in Target Domain*

Table 4 shows the comparison results of retrieval performance in four target domains when unsupervised feature mapping is used to reduce features from high dimensions to different dimensions.

Table 4. Impact of Feature Dimensionality Reduction on Retrieval Performance

Dataset	4096	16	32	64	128	256
UCM	87.81%	**93.23%**	93.21%	93.06%	92.95%	92.99%
WHU-RS	75.07%	81.04%	**81.31%**	80.63%	80.61%	80.61%
RSSCN7	75.15%	85.73%	85.50%	**86.07%**	85.62%	85.63%
AID	82.12%	88.83%	88.88%	**88.93%**	88.73%	88.51%

It can be seen that the unsupervised feature mapping method proposed in this paper has further improved the retrieval performance, with the highest improvement of 5.42%, 6.24%, 10.92%, and 6.81% respectively on the four datasets. This shows that the unsupervised feature mapping method proposed in this paper can widen the gap between classes and narrow the gap within classes, to effectively improve the distinguishing ability of features and improve the retrieval performance. Compared with the other three datasets, the performance on the WHU-RS dataset is lower. The sample data size of the WHU-RS dataset is too small, resulting in lower performance.

By comprehensively comparing the retrieval performance of each dataset under various dimensions, it can be seen that when the dimension is 64, the unsupervised feature mapping method proposed in this paper can obtain the best performance. Therefore, this paper sets the mapping dimension to 64.

To verify the impact of unsupervised feature mapping on retrieval time, we compared the retrieval time before and after dimension reduction. In the process of retrieval, each image needs to calculate the distance from other images in the dataset. We counted the overall working time and divided it by the number of images. Table 5 shows the comparison results of retrieval time in UCM dataset.

Table 5. Impact of Feature Dimensionality Reduction on Retrieval Time

<table>
<tr><th rowspan="2">Dataset</th><th rowspan="2">Samples</th><th rowspan="2">Number of Calculations</th><th colspan="2">Overall Time(ms)</th><th colspan="2">Average Time per Sample(ms)</th></tr>
<tr><th>4096</th><th>64</th><th>4096</th><th>64</th></tr>
<tr><td>UCM</td><td>1890</td><td>3,570,210</td><td>696,074</td><td>202,308</td><td>368</td><td>107</td></tr>
</table>

It can be seen that the unsupervised feature mapping method proposed in this paper improves the retrieval time, reducing the overall retrieval time to 29.06% of the original, and the average time of a single image to 29.08% of the original. The feature data dimension is reduced to 1.56%.

4.6. *Image Retrieval Results*

The results of remote sensing image retrieval are shown in Figure 5. Figure 5 shows the first five images of the partial retrieval results of the UCM dataset. From the retrieval results, the method in this paper can obtain better retrieval results, and the similarity ranking is also roughly in line with the human eye perception and high similarity to the query image. The image is ranked first in the search results.

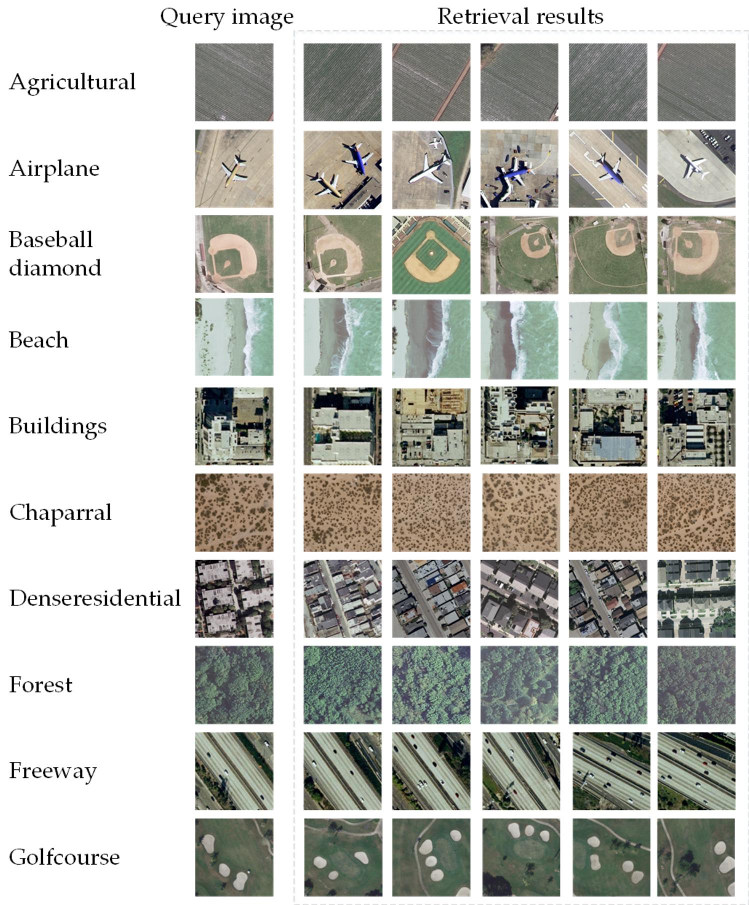

Figure 5. Top 5 image retrieval results of UCM dataset

4.7. *Comparison with Existing Methods*

To verify the effectiveness of this method, we compare it with existing cross-domain retrieval methods of high-definition remote sensing images. There is relatively little research work on cross-domain retrieval of remote sensing images. For a fair comparison, we set the experimental conditions the same as those of other methods. That is, 20% of the samples are randomly selected from the PatternNet dataset to build the dataset of the source domain, 10% are randomly selected from each category of the other four datasets as the labeled data of the target domain, and the remaining 90% are used to test the retrieval performance. The comparison results of retrieval performance obtained by this method and other methods are shown in Table 6, and the experimental results of other methods in the table are from the literature.

Table 6. Comparison results with other cross-Domain methods

Method	Source Domain	Target Domain	mAP
Schumann A. (Schumann et al. 2018)	PatternNet	UCM	87.18%
Chaudhuri U. (Chaudhuri et al. 2019)	PatternNet	UCM	87.12%
Vharkate M. (Vharkate et al. 2021)	UCM	UCM	92.30%
Ours	**PatternNet**	**UCM**	**93.04%**

As can be seen from Table 6, compared with existing methods, the retrieval performance of our method is better than other methods on the UCM dataset. This is because according to the characteristics of remote sensing images, Gabor convolution is introduced into ResNet and ResNeXt networks, which can obtain depth features with good texture description and strong expression ability. The unsupervised feature mapping method can not only effectively remove the redundant information contained in the features, but also further expand the distance between classes, reduce the distance within classes, improve the distinguishing ability of features, and obtain the optimal retrieval performance. At the same time, we also used our method to do cross-domain retrieval tests on WHU-RS, RSSCN7, and AID datasets, and the maps were 80.64%, 85.94%, and 88.71% respectively.

In order to further demonstrate the effectiveness of this method, Figure 6 shows the precision-recall curve of different methods on the UCM dataset. The curves of other methods are from the literature. It can be seen from Figure 6 that the PR curve of this method is higher than other methods, and the results prove that the retrieval effect of this method is better than that of existing methods.

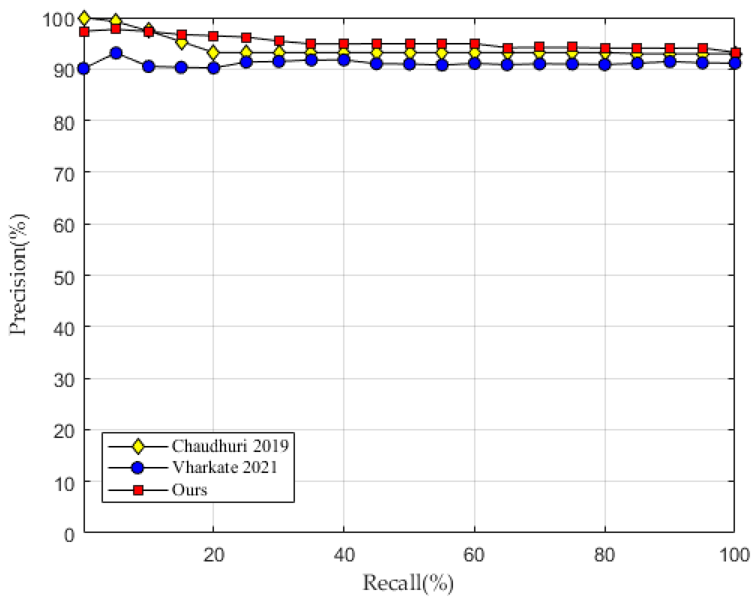

Figure 6. Comparison of PR curves between our method and other methods

5. Conclusions

This paper proposes a novel retrieval method for remote sensing images that realize new retrieval task in target domain by using only a small number of labeled data samples, with the aid of the prior knowledge learned in the source domain. Firstly, Gabor is introduced into CNN, and a depth network model based on Gabor is constructed to enhance the ability of the network to capture texture information. Secendly, a cross-domain knowledge transfer strategy based on dual Gabor-based CNN network learning is proposed. Finally, an unsupervised feature mapping method based on probability distribution is proposed. This method not only further improves the feature discrimination ability, but also reduces the feature dimension, improves the retrieval performance, and greatly reduces the storage space. We evaluated the proposed method and other retrieval methods on four high-resolution remote sensing image datasets. Experimental results show that the proposed method can achieve state-of-the-art retrieval performance. More materials will be released at http://nave.vr3i.com.

Funding

This research was funded by the National Key R&D Program of China under Grant, grant number 2018YFB2100601, and National Natural Science Foundation of China, grant number 61872023.

References

Peyman Bateni, Raghav Goyal, Vaden Masrani, Frank Wood and Leonid Sigal (2020). "Improved Few-Shot Visual Classification.". *2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Ushasi Chaudhuri, Biplab Banerjee and Avik Bhattacharya (2019). "Siamese graph convolutional network for content based remote sensing image retrieval.". *Computer Vision and Image Understanding.*

Yinbo Chen, Zhuang Liu, Huijuan Xu, Trevor Darrell and Xiaolong Wang (2021). "Meta-Baseline: Exploring Simple Meta-Learning for Few-Shot Learning.". *2021 IEEE International Conference on Computer Vision (ICCV).*

Dengxin Dai and Wen Yang (2011). "Satellite Image Classification via Two-Layer Sparse Coding With Biased Image Representation." *IEEE Geoscience and Remote Sensing Letters* 8 (1): 173-176.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Fei-Fei Li (2009). "ImageNet: A large-scale hierarchical image database.". *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami Beach, FL, USA, Jun 20-25, 2009.

Thomas Deselaers, Thomas Deselaers, Daniel Keysers, Daniel Keysers, Hermann Ney and Hermann Ney (2008). "Features for image retrieval: an experimental comparison." *Information Retrieval Journal* 11 (2): 77-107.

Jiechao Guan, Zhiwu Lu, Tao Xiang and Ji-Rong Wen (2020). "Few-Shot Learning as Domain Adaptation: Algorithm and Analysis.". *2020 International Conference on Machine Learning (ICML).*

Yunhui Guo, Noel C. Codella, Leonid Karlinsky, James V. Codella, John R. Smith, Kate Saenko, Tajana Rosing and Rogerio Feris (2020). "A Broader Study of Cross-Domain Few-Shot Learning.". *2020 European Conference on Computer Vision (ECCV).*

Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun (2016). "Deep Residual Learning for Image Recognition.". *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, Jun 27-30, 2016, IEEE.

M.C.Jones (2008). "The family and their close and distant relations.". *Journal of the Korean Statistical Society.*

Yansheng Li, Yongjun Zhang, Xin Huang, Hu Zhu and Jiayi Ma (2018). "Large-Scale Remote Sensing Image Retrieval by Deep Hashing Neural Networks." *IEEE transactions on geoscience and remote sensing* 56 (2): 950-965.

Lingjun Li, Junwei Han, Xiwen Yao, Gong Cheng and Lei Guo (2021). "DLA-MatchNet for Few-Shot Remote Sensing Image Scene Classification." *IEEE transactions on geoscience and remote sensing* 59 (9): 7844-7853.

Yansheng Li, Te Shi, Yongjun Zhang, Wei Chen and Hao Li (2021). "Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation.”. *ISPRS Journal of Photogrammetry and Remote Sensing*.

Shangzhen Luan, Chen Chen, Baochang Zhang, Jungong Han and Jianzhuang Liu (2018). ”Gabor Convolutional Networks.”. *IEEE Transactions on Image Processing*.

Canjie Luo, Yuanzhi Zhu, Lianwen Jin and Yongpan Wang (2020). ”Learn to Augment: Joint Data Augmentation and Network Optimization for Text Recognition.”. *2020 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Gökhan Ozbulak and Hazım K Ekenel (2018). ”Initialization of convolutional neural networks by Gabor filters.”. *2018 26th Signal Processing and Communications Applications Conference (SIU)*.

Sinno Jialin Pan and Qiang Yang (2010). ”A Survey on Transfer Learning.” *IEEE transactions on knowledge and data engineering* 22 (10): 1345-1359.

Xingang Pan, Ping Luo, Jianping Shi and Xiaoou Tang (2018). ”Two at Once: Enhancing Learning and Generalization Capacities via IBN-Net.”. *2018 European Conference on Computer Vision (ECCV)*.

Arne Schumann, Lars Sommer, Max Vogler and Jurgen Beyerer (2018). ”Onthology-based Masking Loss for Improved Generalization in Remote Sensing Semantic Image Retrieval.”. *ISPRS Journal of Photogrammetry and Remote Sensing*.

Hung-Yu Tseng, Hsin-Ying Lee, Jia-Bin Huang and Ming-Hsuan Yang (2020). ”Cross-Domain Few-Shot Classification via Learned Feature-Wise Transformation.”. *2020 International Conference on Learning Representation (ICLR)*.

Eric Tzeng, Judy Hoffman, Kate Saenko and Trevor Darrell (2017). ”Adversarial Discriminative Domain Adaptation.”. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Minakshi N Vharkate and Vijaya B. Musande (2021). ”Remote Sensing Image Retrieval using Hybrid Visual Geometry Group Network with Relevance Feedback.”. *International Journal of Remote Sensing*.

Gui-Song Xia, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang and Xiaoqiang Lu (2017). ”AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification.” *IEEE Transactions on Geoscience and Remote Sensing* 55 (7): 3965-3981.

Yi Yang and Shawn Newsam (2010). ”Bag-of-visual-words and spatial extensions for land-use classification.”. *2010 ACM Sigspatial International Conference on Advances in Geographic Information Systems*, San Jose, CA, USA, Nov 3-5, 2010.

Pei Zhang, Yunpeng Bai, Dong Wang, Bendu Bai and Ying Li (2021). ”Few-Shot Classification of Aerial Scene Images via Meta-Learning.” *Remote sensing (Basel, Switzerland)* 13 (108): 108.

An Zhao, Mingyu Ding, Zhiwu Lu, Tao Xiang, Yulei Niu, Jiechao Guan, Ji-Rong Wen and Ping Luo (2021). ”Domain-Adaptive Few-Shot Learning.”. *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*.

Weixun Zhou, Shawn Newsam, Congmin Li and Zhenfeng Shao (2018). ”PatternNet: A benchmark dataset for performance evaluation of remote sensing image retrieval.” *ISPRS journal of photogrammetry and remote sensing* 145: 197-209.

Zheng Zhuo and Zhong Zhou (2020). ”Low Dimensional Discriminative Representation of Fully Connected Layer Features Using Extended LargeVis Method for High-Resolution Remote Sensing Image Retrieval.” *Sensors (Basel, Switzerland)* 20 (17): 4718.

Zheng Zhuo and Zhong Zhou (2021). ”Remote Sensing Image Retrieval with Gabor-CA-ResNet and Split-Based Deep Feature Transform Network.” *Remote sensing (Basel, Switzerland)* 13 (5): 869.

Qin Zou, Lihao Ni, Tong Zhang and Qian Wang (2015). ”Deep Learning Based Feature Selection for Remote Sensing Scene Classification.” *IEEE Geoscience and Remote Sensing Letters* 12 (11): 2321-2325.