A Multi-instance Multi-label Learning Framework of Image Retrieval

Chaojun Wang, Zhixin Li, Canlong Zhang

A Multi-instance Multi-label Learning Framework of Image Retrieval

Chaojun Wang, Zhixin Li*, Canlong Zhang

College of Computer Science and Information Technology,
Guangxi Normal University, Guilin 541004, China
* Corresponding author.

wangchaojun009@126.com, lizx@gxnu.edu.cn, clzhang@gxnu.edu.cn

Abstract. Because multi-instance and multi-label learning can effectively deal with the problem of ambiguity when processing images. A multi-instance and multi-label learning method based on Content Based Image Retrieve (CBIR) is proposed in this paper, and the image processing stage we use in image retrieval process is multi-instance and multi-label. We correspond the instances with category labels by using a package which contains the color and texture features of the image area. According to the user to select an image to generate positive sample packs and anti-packages, using multi-instance learning algorithms to learn, using the image retrieval and relevance feedback, the experimental results show that the algorithm is better than the other three algorithms to retrieve results and its retrieval efficiency is higher. According to the user to select an image to generate positive sample packs and anti-packages, using multi-instance learning algorithms to learn, using the image retrieval and relevance feedback. Compared with several algorithms, the experimental results show that the performance of our algorithm is better and its retrieval efficiency is higher.

Keywords: multi-instance multi-Label learning, image retrieval, multi-points diverse density algorithm, image retrieval

1 Introduction

We live in an era in which multimedia information grows at a speed of geometric index and network technologies are popularized quickly .It is more and more difficult that the images we need are found quickly and effectively from the pictures about the ocean. We can know that text-based image retrieval and content-based image retrieval doesn't solve well the content of the image of the ambiguity through the above introduction, and a complete image includes a lot of small parts. If only a single label does not adequately describe the image, but cannot accurately expressed the users are interested in the parts, which can lead that the part of the image that we are own interested in is not retrieved .The content of this article is to divide the image into several parts, take partial features from several pieces of information as the package of the image generated by examples, and contact the examples of the package with the category of the mage. This paper presents a method of the image retrieval based on multi-instance multi-label study. The method takes use of local color of images and texture features to generate multi-sample packages, and give the appropriate subset of classes for objects, and it is no longer the only category tag. We can get the complex high-level concepts derived by the subsets of these categories marked, so that we improve the probability of recall and precision in the process of image retrieval.

2 Multi-instance multi-label learning

In multi-instance learning, every training package is made of more than one sample, and the sample has no concept, but each training package has a concept marked. When we apply this method to image retrieval, we will find that it belongs to a variety of categories for images. If we only use a single category marked words, we cannot fully describe the image, and it will be a problem of semantic gap. Considering the image content with multiple semantics, we should learn that it is actually a map from the sample set to the category tag set on. Considering the image content with multiple semantics, we should learn that it is actually a map from the sample set to the category tag set on. Multi-instance multi-label[1] defined as follows, χ represent example of space γ represent the category space, and represent by mathematical symbols follows:

Multi-instance multi-label learning: given a set of data $\{(X_1, Y_1), (X_2, Y_2),\ldots,(X_m, Y_m)\}$.Objective to learn f : $2^{\chi} \rightarrow 2^{\gamma}$. $X_i \subseteq \chi$ Is a set of sample $\{ \chi_{i1}, \chi_{i2}, \ldots, \chi_{ini} \}$, $\chi_{ij} \in \chi$ (j=1,2,…,n_i) and $Y_i \in \chi$ is a set of appropriate categories tag of X_i. $\{y_{i1},y_{i2},\ldots,y_{ili}\}$ $y_{ik} \in \gamma$(k=1,2,…,l_i). n_i is the number of samples among X_i, l_i is the number of tags among Y_i.

There is learning frameworks named traditional supervised learning (single example, single tag) and multi-label learning (single example, multiple tags) and multi-instance learning (multiple example, single tag) besides the above one of sample learning frameworks in this paper. From the above, it is not difficult to find that MIML framework can be converted into other three examples in some special cases. We know that only by using the appropriate expression can capture the important information of image. However ,dealing with the ambiguity of the image by using MIML can better deal with semantic gap problems .MIML also helps to study the high-level semantic concepts, such as elephants, lions on the picture, trees, and some grass. Through the concept of markers we can infer that this is a picture of a description of Africa. This example is the use of MIML and then study the underlying logic of high-level concepts .In order to play the ability of MIML framework, there are generally two kinds of methods, respectively, for example as Bridges and multi-label learning as a bridge.. This paper is to convert Multi-instance multi-label learning algorithm MIMLBOOST [2][3] into Multi-instance learning.

3 The structure of the package and image retrieval

3.1 The structure of the sample package process

Image of the sample bag choosing what kind of method determines whether the image semantic information is complete, this will directly affect the retrieval effect. But Maron and Yang's Multiple-Instance learning by maximizing diverse density [4] use the dividing method of fixed generated package. This method will cause originally belong to the same concept divide into different packages or belong to different concepts merged into one package. Sample package need to define in advance. These will only contain color information. Colors of different parts may be similar which can lead to low accuracy of image retrieval. The method [5] of Dai H.Blet each image fixed generates 4 sample packages. What kind of effect will be caused by this method, For a simple image that is bound to introduce noise, and compared with the complex image of some of the information will not wholly include .This paper will use a partition method which is a network clustering of Self-organizing feature Map (Self - Organizing feature Map, SOM)[6].This article uses SOM clustering image segmentation methods[7] to divide images, we use the color features as SOM network input to cluster. After go to the discrete points and region merging after some steps such as each region of the image, we respectively calculate the R, G, B, 3 kinds of color grayscale average to the corresponding color features. If we only use color features cannot describe an area, different images in similar parts of color have some differences and different parts of the color may be similar. In order to improve the accuracy of image retrieval, we have to use textures which are some characteristics of the image .Texture is an important means of visual image .Also it describes gray space distribution of image. Texture images can show some information about the arrangement of surface structure, such as roughness, contrast, direction, the line as the degrees, neat and rough degrees, etc. This paper uses co-occurrence matrix[8] method to deal with the texture feature extraction.

If we break the base of decomposition k, then we can get a $3\times(k\text{-}1)$ co-occurrence matrix. This article will choose 4 most commonly used statistics out of 10. Also it can reflect the scale of statistics as the texture feature. Finally we add color space R, G, B whose gray value is most characteristic's-occurrence matrix has a feature vector of $4\times(k\text{-}1)+3$ dimension. Regard the regions corresponding eigenvectors as an example. So it can be an image in which all areas of its corresponding sample image corresponding to the package.

3.2 Image Retrieval

Figure 1 describes the image retrieval process. There is multi-instance and multi-label learning concept: regard an image as a sample package and the image has a lot of class label. It may be more meaningful in some cases if we understand why an image has some class label, the MIML makes these possible. Using the MIML we can know there is some samples to have such class label. So we can match sample directly with class label images. And we also not only can use image package but also can use the class label between the samples for image retrieval.

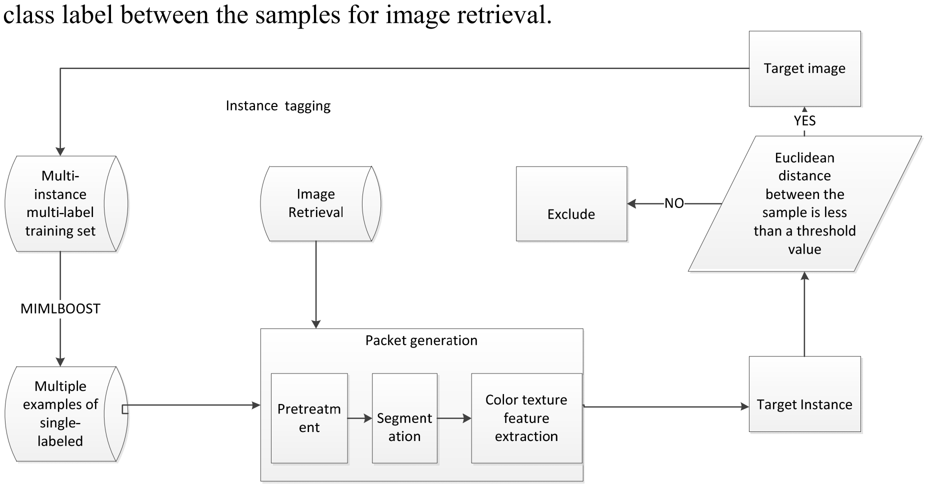

Fig. 1. Flow diagram of image retrieval

In the second chapter we introduce four models of machine learning, the special case of the MIML is the other three kinds of machine learning methods. And in this article the process of solving for multiple tag sample learning problems can be converted the problem into a single tag sample learning problems by using MIMLBOOST algorithm. What is the purpose of doing that? We can find the benefits of doing that through the experiment. This article uses the multi- diversity density algorithm (MPDD) [9] from the image training package to concentrated study some of the concepts for image retrieval. The idea of multi- diversity density algorithm is using density of multiple points of comprehensive information to describe the concept what users are interested in. This algorithm is an improvement compared with the density of the traditional diversity algorithm [10]. When using the density of diversity as a measure, we must select the point of maximum density as a target. If the dimension of the feature subspace is higher, we can’t find out the point of maximum density by iterating through the whole feature space. The other is the feature space and vector are both distributions, this will lead to exist multiple local minima points. While using the gradient descent method can’t avoid fall into local extremum but get a local optimal solution. We know that the use of one example of a sample point is difficult to contain the contents of the sample, this will cause the description of the contents of the sample packets to have certain content, in order to solve these problems it uses the MPDD algorithm in this article. MPDD algorithm calculation process is as follows.

1. The use of colour and texture extracts a features package, and we enter a positive bag and negative bag, to divided the examples and extract the description of the characteristics of the sample.
2. The use of the diversity of classic learning algorithm to find the divers density points, and output the search of the density of the point.

3. Integrated multiple search the information of the density of the point to calculate the density of the point and the distance between the unknown sample and the point. And we calculate the distance between the unknown samples and the density of multiple diversity point through the following formula.

   Set MPDD algorithm output M (M value unknown) a variety of density points, denoted by $F^* = \{F_1, F_2, \ldots, F_M\}$.An unknown sample package has N samples, denoted by $B^M = \{B_1{}^M, B_2{}^M, \ldots, B_N{}^M\}$ Then the distance between the unknown sample bag B^M and the density of multiple diversity point is

$$\mathrm{Distance}(B^M, F^*) = \frac{1}{M} \min_{i \in N} \mathrm{dis}(B^{M_i}, F_j) \tag{1}$$

In this paper we choose more density of points is selected in the sample package. In the process of calculation is to calculate the density of each variety point and test set are package sample the Euclidean distance between, then calculated the distance between the minimum value, then sum and average. Finding out and testing set minimum Euclidean distance are examples of the package contained. We know when described different images, the importance of the various features of feature space is different, if some image has obvious characteristic of image color and texture feature. Then we can give the characteristics of each assigned a weight to represent different characteristics of different degree. The distance between the instances with the target is measured weighted Euclidean distance. Expressed in mathematical formula is as follows:

$$\left\| B_{i,j} - m \right\|^2 = \sum_n \omega_i (B_{ijn} - m_n)^2 \tag{2}$$

Image retrieval process is as follows:

- With the method in the article, we can constitute the image of the sample bag by using the image texture and color. For different image, we can merge sample package with the same texture segmentation area and the color characteristics into sample package with new characteristics by using formula 2 in the article.
- We can use the K Nearest neighbour (k-Nearest Neighbour， KNN)[11] classification algorithms to classify the training focused sample packages and test focused sample packages.
- Find the feature vector of the target which is appropriate to users by using formula 1 of the article. Then we can calculate the Euclidean distance between the density point of diversity and all of the samples in the unlabeled sample package. And we can average them as the similarity measure, Order the distance from small to large. After that, present the corresponding image to user.
- According to the rank of similarity measure, we can choice the interested and uninterested images from the retrieval results and label the corresponding classification of package. After that, put it into training set to get relevant feedback.

4 Experimental Result

We do the experiment based on this method, and compare the experiment results with other methods including the Multiple-instance image retrieval[12] proposed by Dai Hongbin, Maron and Ratan and Lozano-Perez’s multiple-instance learning method known as the diverse density algorithm[13]. The image dataset this experiment uses has a medium scale. It divides into 10 classes: elephants, flowers, birds, horses, dinosaurs, buildings, and cars, mountains with snow, beaches and plants. The total number of images is up to 1500 and each class has 150 images. We pick 30 images from each class to form the training set as examples. The rest 1200 images form a test set. During each experiment, one class is selected as the target class, and then 5 images which we are care about are selected to generate the positive packet. We select 5 classes randomly from the rest classes and then select an image we are not care about from each of 5 classes randomly to form the negative packet. Now a training set is formed and it includes a positive and negative packet which has 5 images respectively. Here we use more diversity, MPDD density algorithm to learn the training set and retrieve the image in the test set. Then the results are returned by sorting the Euclidean distance of similarity. From the search results, we select the first 5 images which are sorted in target and non-target category, and add the corresponding packet and examples to the training set. We compare the experiment results using the method of Dai Hongbin, Maron and Ratan’s Single blob with neighbours method and the method of Yang and Lozano-Perez.

4.1 Experimental Comparison of four methods

HongbinDai and ZhihuaZhou use the method of extracting colour features together with the 24 texture features extracted by 24 filters of Gabor wavelet transformation. The feature vector has 27 dimensions: 3 colour features R, G, B and 24 texture features. The feature vector of every region is considered as examples.

Maron and Ratan’s Single blob with neighbours’ method constitutes the matrix by colour blobs after sampling the image. The feature vector with 15 dimensions is calculated from the R, G and B colour features in each colour blob and four blobs connected with it. There are 9 image packages and each contains a 15-dimensional feature vector.

However, Yang and Lozano - Perez's method is to convert the images to grey ones. After dividing into twenty overlapping regions, every region of the image is conducted mirror filtering, transforming and sampling. There are 40 image packages with 64- dimensional feature vector. The results of the experimental are as follows:

This article is for each category of images 10 times experiment, we used precision recall such a way to measure the efficiency of various methods of retrieval. Precision is when system is the process, the system returns results in the target image and all returned images of a ratio. In this experiment, 5 interested in image and not interested in the image of an initial search results are shown in figure2. And in the process of experiment increases 5 interested in image and not interested in image relevance feedback result is shown in figure 3. For the above all kinds of methods for initial retrieval and related feedback average retrieval time are shown in table1.

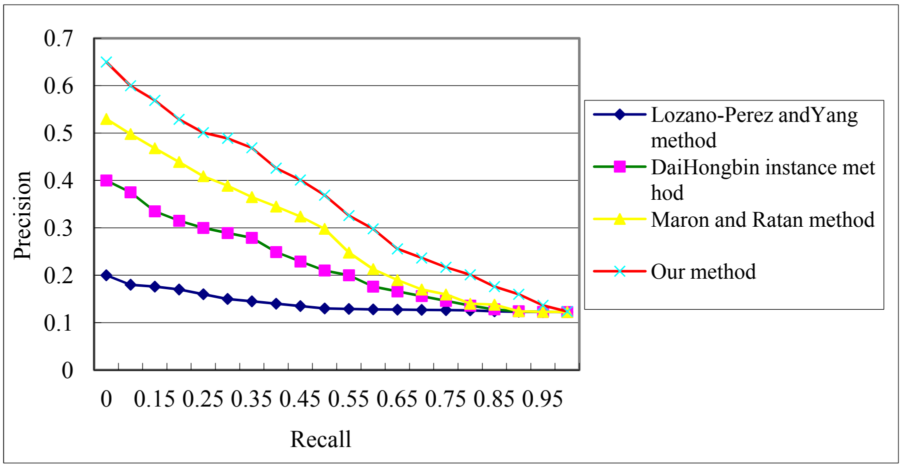

Fig. 2. Initial retrieval results of relevant and 5 irrelevant images

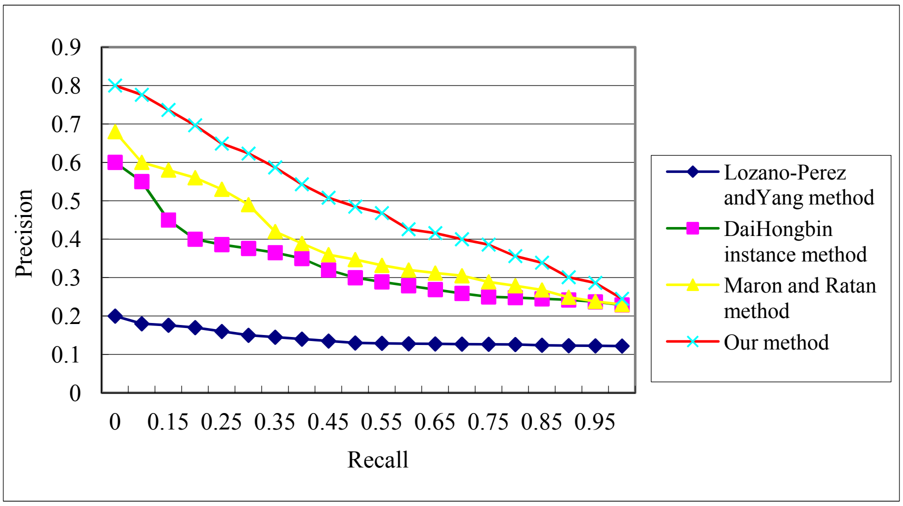

Fig. 3. Relevant feedback results of 10 relevant and 10 irrelevant images

As can be seen from the figure 2 and figure 3, the proposed method takes precedence over these three methods. In this paper, the average time for the initial retrieval and feedback time is prior to the other two methods but longer than another method. The main reason is that MIML learning methods is adopted in this paper. Linking category tag with the sample, the time we use more diversity density algorithm to find the point is less than the former two methods but is more than the density of diversity algorithm. The time this algorithm takes is less than the former two but more than Hongbin Dai multi-instance algorithm. Adopting multi-sample and multi-label algorithm, this article can well describe the image. So through next section in this paper, we learned that the algorithm has better precision and recall radio.

Table 1. Performance comparison on the task of automatic image annotation

	Maron and Ratan method	Lozano-Perez andYang method	DaiHongbin instance method	Our method
The average time of Initial retrieves (s)	3.2	7.63	1.47	2.36
Relevance feedback time (s)	9.56	8.52	2.92	4.24

4.2 Results in different image library

In order to study the efficiency and results on the different size of image library by this method, we respectively remove 500 and 1000 images from the original image library .I repeated the experiment for the method the article put forward.

From table 2 we can see that, with the increase of image database retrieval time also increases accordingly, but increased more slowly through the experiment we can find the time complexity of this article was close to linear.

Table 2. Performance comparison on the task of automatic image annotation

	500 images	1000 images	1500 images
The average time of initial retrieves (s)	1.21	1.56	1.72
Relevance feedback time (s)	2.86	3.12	3.36

Figures 4 and figure 5 show the results of image retrieval and the related feedback in different size of image library. We can see the algorithm has smooth retrieval performance with the increase number of experimental image from the experience, and the performance does not appear significant decline. So, the algorithm of this article can be an applicable method to retrieve a large scale image scale image library.

5 Conclusion

This article puts forward a kind of image retrieval based on multiple sample multi-label learning technology. The benefits of this technology are to reduce the semantic gap and solve the problem of ambiguity of the image. It enlighten us use the underlying semantic concept of high-level semantics. Through the above experiment we find this method is helpful to improve the image retrieval precision rate and recall rate. This article implementation process use the method of MIML as the first step to process image, utilize MIMIL Boost algorithm transforms the problem of multiple sample tags into multiple sample single tag questions, then the image segmentation algorithm based on SOM clustering segment image into multiple regions. We combine each region's color characteristics and co-occurrence matrix as a sample package, then use multi-points diverse algorithm for image retrieval and related feedback. It is suitable for medium-sized image library and compare with proposed method in our experiments, the results show that the algorithm of this article retrieval effect is better than that of Maron, Ratan, Yang and Lozano-Perez etc.This paper also introduced to other characteristics to improve the image of package technology. In the future research we learn how to put the MIML into the image retrieval studying, reduce the time of recall and precision of image retrieval ,and design a better algorithm to find out the target class image in the shortest time.

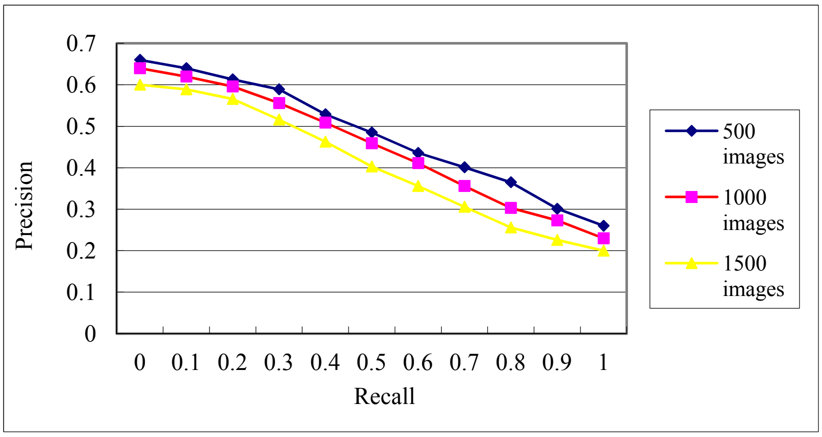

Fig. 4. Relevant feedback results of 10 relevant and 10 irrelevant images

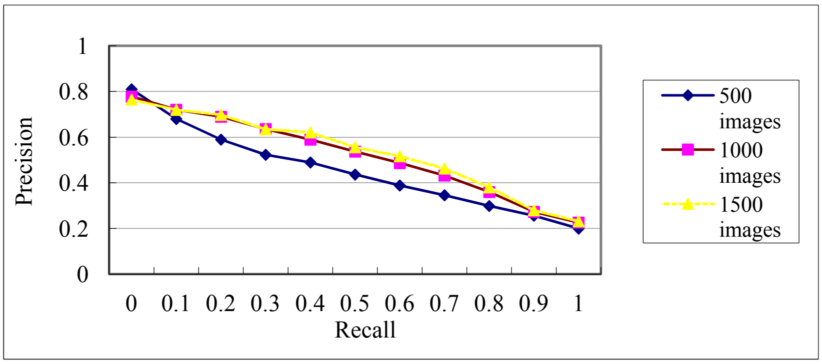

Fig. 5. Relevant feedback results of 10 relevant and 10 irrelevant images

6 Acknowledgement

This work is supported by the National Natural Science Foundation of China (Nos. 61165009, 61262005, 61363035, 61365009), the National Basic Research Program of China (No. 2012CB326403) and the Guangxi Natural Science Foundation (Nos. 2012GXNSFAA053219, 2013GXNSFAA019345) and the “Bagui Scholar” Project Special Funds.

7 References

1. Zha Z J, Hua X S, Mei T, et al. Joint multi-label multi-instance learning for image classification[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.
2. Zhou Z-H, Zhang M-L. Multi-instance multi-label learning with application to scene classi-fication [C]. // Advances in Neural Information Processing Systems, 2007, 19: 1609-1616.
3. Zhou Z-H, Zhang M-L. Multi-instance Multi-label learning [J]. Artificial Intelligence. 2012, 176(1):2291-2320.
4. Maron O, Ratan A L. Multiple-Instance Learning for Natural Scene Classification[C]. // Proceedings of ICML. 1998, 98: 341-349.
5. Yang C, Lozano-Perez T. Image database retrieval with multiple-instance Learning techniques [C]. In: Proceedings of ICDE. San Diego, CA, 2000, 233-243.
6. Kohonen T. Self-organizing maps [M]. Berlin: Springer-Verlag, 2001.
7. Jiang Y, Chen K J, Zhou Z H. SOM based image segmentation[C]. //Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing. Berlin: Springer-Verlag, 2003: 640-643.
8. Marceau D J, Howarth P J, Dubois J M M, et al. Evaluation of the grey-level co-occurrence matrix method for land-cover classification using SPOT imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 1990, 28(4): 513-519.
9. Böhm C, Krebs F. The k-nearest neighbour join: Turbo charging the KDD process[J]. Knowledge and Information Systems, 2004, 6(6): 728-749.
10. Chen Mian-shu, Yang Shu-yuan, Zhao Zhi-jie et al. Multi-points diverse density learning algorithm and its application in image retrieval [J]. Journal of jilin university: engineering science, 2011, 41 (5): 1456-1460.
11. Maron O, Lozano-Pérez T. A framework for multiple-instance learning[C]. // Advances in neural information processing systems, 1998: 570-576.
12. Dai Hongbin, Zhang Ling, Zhou Zhihua. A Multi-Instance based approach to image retrieval [J]. Journal of pattern recognition and artificial intelligence, 2006, 19(2): 179-185.
13. ChengYang, Lozano-PerezTomas. Image database retrieval with multiple-instance learning techniques[C]. In: Proceedings of the 16th International Conference on Data Engineering, 2000.