



Partial Order Structure Based Image Retrieval

Zhuoyi Li^{1,2}, Guanghua Gu^{1,2(✉)}, and Jiangtao Liu^{1,2}

¹ School of Information Science and Engineering, Yanshan University,
Qinhuangdao 066004, China
guguanghua@ysu.edu.cn

² Hebei Provincial Key Laboratory of Information Transmission and Signal
Processing, Qinhuangdao 066004, China

Abstract. Image retrieval plays an important role in the growing computer vision applications. The computation of the unrelated images in large scale image retrieval task seriously reduces the retrieval efficiency. In this paper, a new Partial Order Structure (POS) based image retrieval method is proposed. Partial order structure diagram is an effective visualization tool in Formal Concept Analysis (FCA) theory, including object partial order structure diagram and attribute partial order structure diagram. There are two contributions in this paper. First, we design an association rule according to the object partial order structure (OPOS) method to measure the correlation between the query image and the database, and then improve the database to be retrieved. Second, we perform a query expansion according to the attribute partial order structure (APOS) method to improve the generalization ability of the query information. Experimental results on two databases verify the effectiveness of the proposed algorithm.

Keywords: Image retrieval · Formal Concept Analysis · Partial order structure · Association rule · Query expansion

1 Introduction

Image retrieval is a significant part of computer vision applications. In the common retrieval methods, the retrieval process is to calculate the similarity between the feature descriptions of the query image and image database. However, many images unrelated to the query image are calculated for the similarity, which wastes a lot of time. To enhance the computation efficiency, we attempt to improve the database by removing the images with the unrelated semantics to the query image before the retrieval.

The partial order structure is a powerful visualization tool for the representational concept in Formal Concept Analysis (FCA) theory [1]. It contains the object partial order structure and the attribute partial order structure. In the object partial order structure diagram, nodes represent objects, and branches represent attributes. The cluster structure in the object partial order structure refers to the structural relationship of the objects containing the same or similar attributes. Such a structure has the role of clustering [2, 3]. In the object partial order structure, the attributes in the same cluster

The first author (Zhuoyi Li) is a student.

have the higher similarity than the attributes in different clusters. So in this paper, we set an association rule to mine the similarities of the attributes based on the object partial order structure.

In the attribute partial order structure diagram, nodes represent attributes, and branches represent objects. The similarity of two branches can be calculated by the number of identical nodes on the branches. To make features of the query image more expressive, we use the attribute partial order structure to find the several top-ranked images by calculating the similarity between the query image and the database. The query image is expanded into the merged image of the top-ranked individuals.

The deep features have strong generalization ability, which can effectively capture the semantics of the image [4, 5]. The image features extracted by CNN model can globally represent the image [6]. When the deep neural network is applied to image feature extraction, the high-dimensional features show the great advantages in image processing. Many works [7, 8] focused on replacing the traditional hand-crafted descriptors with the deep features from the fully connected layers of a pre-trained CNN model for image classification. Other methods [9, 10] used the sum or max pooled convolutional features instead of the fully connected layer and achieved better results. Additionally, some methods first divided the image into several blocks [11, 12]. The features of the image blocks are obtained through the fully connected layer or the convolution layer, and the image blocks are encoded by the BOW model or the VLAD model [13, 14]. These methods used the local information of the image block. Although these methods were commonly used in image retrieval, extracting the CNN features of each image block is complicated and inefficient [15].

Feature maps are image features produced by the original image after convolution in a neural network. In this paper, the activations of the different neuron arrays across all feature maps in a convolutional layer are treated as local features. A single transfer of an image through CNN is sufficient to obtain its local features of blocks [16]. It avoids the block segmentation and the feature extraction such that the retrieval complexity is greatly reduced. We obtain the local features of the images in the database from the last convolutional layer in the VGG network. The visual semantics are obtained by clustering local features. In this paper, visual semantics are set as the attributes, and images are set as the objects to establish the formal background. The partial order structure is built based on the formal background. According to the object partial order structure, the association rule is designed to improve the image database that contains the same or similar semantics as the query image, that is, the unrelated images to the query image are removed. Meanwhile, the query expansion is performed based on the attribute partial order structure to complete the image retrieval.

2 Related Works

In the theory of Formal Conceptual Analysis (FCA), the human cognition consists of three basic elements: objects, attributes, and relationship between objects and attributes [17]. The objects are the individuals in the database. The attributes are the characteristics of the various objects. The relationship between attributes and objects indicates the corresponding connection of both [18], which is described by the formal

background as shown in Table 1. Here, I_1 – I_6 are 6 objects, and S_1 – S_6 are 6 attributes. The formal background is a Boolean matrix including 1 or 0. Here 1 represents the object has the corresponding attribute, and 0 represents the non-correspondence.

Table 1. Formal background

	S_1	S_2	S_3	S_4	S_5	S_6
I_1	1	1	0	0	0	0
I_2	1	1	0	0	1	0
I_3	1	0	1	0	1	0
I_4	0	0	1	0	0	1
I_5	0	1	0	0	1	0
I_6	1	1	0	1	1	0

The formal background is the correspondence between objects and attributes. In this paper, we use VGG-f model to get the visual fuzzy semantics. Images are regarded as the objects and visual semantics are regarded as the attributes. The formal background is built by the objects and attributes, and then the partial order structure diagram is established. The partial order structure diagram contains the object partial order structure (OPOS) diagram and the attribute partial order structure (APOS) diagram. Both realize the hierarchical clustering of the data, which is useful to analyze the concept composition.

In the OPOS diagram, each node shows an object (image), and each branch under nodes indicates an attribute (visual semantic). OPOS brings together the objects with the similar attributes to generate a hierarchy of the database. In OPOS, the data filtering or data clustering can be completed according to the cluster structure. We put the clusters with more objects into the top of the hierarchy, and the clusters with less objects into the bottom of the hierarchy. The resulting hierarchical partial structure is helpful for data searching.

The generation process of the APOS diagram is similar to the OPOS diagram, but the node and the branch represent the opposite meaning. In the APOS diagram, each branch represents an object, and all attributes contained in this object correspond to all nodes on the branch. For an instance, the OPOS and APOS diagrams generated by the formal background in Table 1 are shown in Fig. 1.

Figure 1(a) shows an OPOS diagram. Here I_1 – I_6 indicates 6 objects (images), and S_1 – S_6 represents 6 attributes (fuzzy semantics). Each branch in Fig. 1(a) means an attribute. The number of attributes is the number of the clusters obtained by the clustering method. If two attributes contain more of the same images, the two visual semantics are more likely to appear at the same time. In other words, if the two visual semantics are more relevant, one is suitable for complementing the other. Figure 1(b) shows an APOS diagram in which each branch means an image and each node means a fuzzy semantic. If two branches have the more identical nodes, it means the branches have the higher similarity. In other words, the similarity between images can be determined by the number of identical nodes contained on the branch.

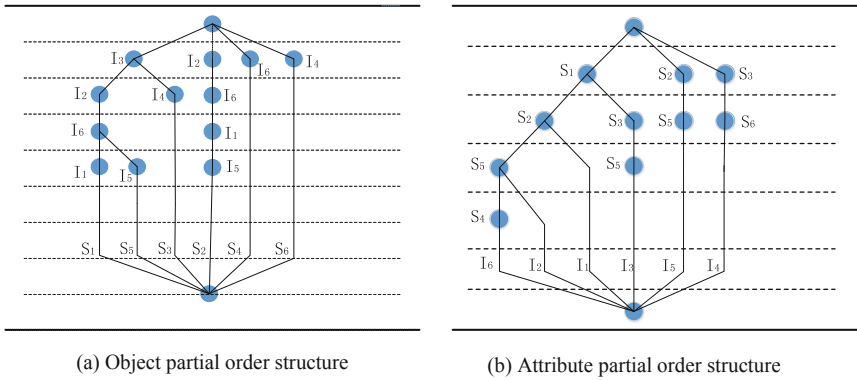


Fig. 1. An example for the OPOS diagram and APOS diagram

3 Image Retrieval Based on Partial Order Structure

The proposed method consists of three parts: Deep Feature Extraction (DFE), Partial Order Structure (POS) and Feature Similarity Metric (FSM). The main framework is shown in Fig. 2.

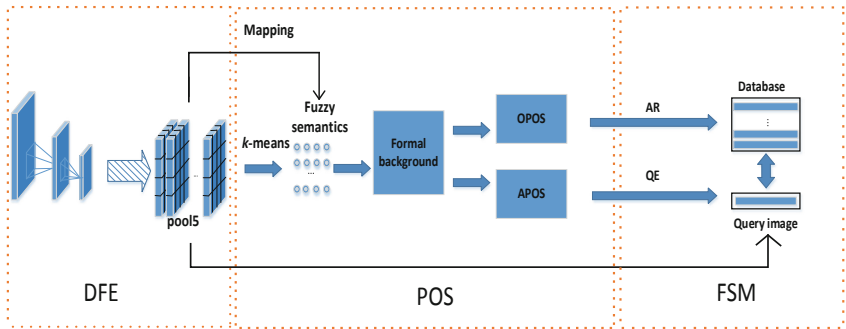


Fig. 2. Image Retrieval flowchart based on partial order structure

The DFE part contains the extraction of the local features in the network and the extraction of the global features from the full connect layer. The POS section is the primary step in the proposed method. It completes the adaptation of the database by using OPOS and association rule (AR). This part also expands the feature information of the query image, and improves the expressive ability of query images by using query expansion (QE) based on APOS. The FSM section computes the distance similarity metric between the query image and the database to generate the rank of the retrieval.

3.1 Visual Semantics

In convolutional neural network, the convolution kernel in the convolution layer acts on the receptive field [6]. The convolution calculation method is selected for each input image to form the image feature maps. Different feature maps represent different local features [19]. In the process of obtaining visual semantics, we extract the features by the last convolution layer (pool5) in the VGG-f network pre-trained on the ImageNet database. The network has not been fine-tuned. Each vector in the pool5 layer is treated as a local feature. An image extracts 6*6 local feature descriptions, each of which is 256-dimensional. It means that an image can be represented as the 36 local features with a dimension of 256.

Among the clustering methods, the k -means algorithm keeps scalability and high efficiency for dealing with the big data. In this paper, after obtaining the local features, we perform the k -means clustering to get the centers, which are regarded as the visual fuzzy semantics. Features of all images in database are mapped into the centers to form their visual fuzzy semantic representation [20]. We expect that each local feature can be properly mapped into the most appropriate cluster (i.e. the fuzzy semantics). As we all know, the number of k is essential for k -means. In this paper, a Davies Bouldin index (DBI) [21] method is used as a criterion to select the number of cluster centers.

3.2 Object Partial Order Structure

According to the OPOS diagram, we can remove the images on the branches that have no visual semantics of the query image. However, if we only leave the images with the same visual semantics as the query image, some images with the similar semantics may be ignored. It maybe leads to a bad recall rate. Therefore, it is necessary to find the similar semantics as the complementary semantics.

Some visual semantics often appear at the same time, such as airplane and sky. These simultaneous semantics are often used to complement the retrieval each other. It is necessary to calculate the association between the visual semantics by using the partial order structure, and get the related semantics of each visual semantic. This paper defines a semantic association rule to measure the degree of association between two semantics in Eq. 1.

$$D_a = \frac{\text{num}(Im(X) \cap Im(Y))}{\text{num}(G)} \quad (1)$$

Here X and Y represent two semantics respectively. $Im(X)$ and $Im(Y)$ represent the images containing the semantics X and Y , respectively. For instance, if X is S_1 and Y is S_5 in Fig. 1(a), $Im(X)$ represents the image set $\{I_3, I_2, I_6, I_1\}$ and $Im(Y)$ represents the image set $\{I_3, I_2, I_6, I_5\}$. $Im(X) \cap Im(Y)$ represents the common images $\{I_3, I_2, I_6\}$ from the vertex down. In the object partial order structure, there are some top clusters, that is, the branches under the nodes in the first level, for example the clusters under the nodes I_3, I_2, I_6 and I_4 . So as to standardize the relevance of two semantics, avoided the influence of the number of pictures between different clusters on semantic relevance, we perform correlation metrics within clusters in a partial order structure and introduce

G as the denominator. In Eq. 1, G is the image set of the top cluster in which semantics X and Y are belonging, that is, $\{I_3, I_4, I_2, I_1, I_6, I_5\}$ in Fig. 1(a). The function $num(\cdot)$ indicates the number of the image set.

As well known, in a database, a picture often has multiple semantics, the semantics are distributed across multiple clusters. For each semantic, we calculate the similarity between it and other semantics in a cluster and rank the similarity. The semantics that the similarity D_a is greater than the mean value within the cluster is considered to be the complementary semantics. The images on the branches with the higher similarity are merged to form the final image database to be retrieved. Figure 3 displays an example of the semantics complement.

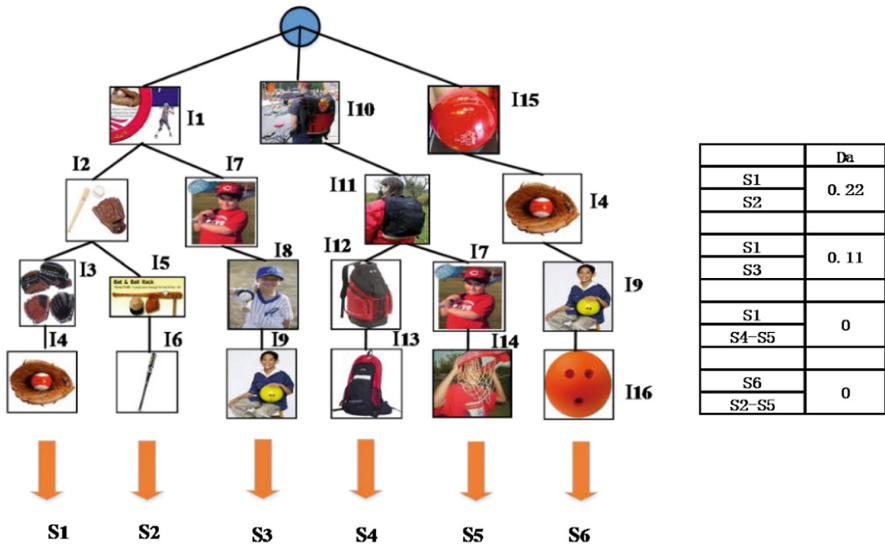


Fig. 3. Semantic supplementation based on the association rule

Suppose that the query image contains semantics S_1 and S_6 , the association rule algorithm based on the OPOS diagram in Fig. 3 is as follows:

Input: Query image, database, and OPOS.

1. Get the branch $\{S_1, S_6\}$ of the query image from OPOS, and take the images on the branches as a new database DB $\{I_1-I_4, I_9, I_{15}-I_{16}\}$.
2. Calculate the degree of association D_a between $\{S_1, S_6\}$ and other semantics $\{S_2, S_3, S_4, S_5\}$ by Eq. 1.
3. Calculating the mean value (0.165) of the semantic similarity D_a in the cluster where S_1 is located.
4. Update DB by adding images on branches $\{S_2\}$, that is $\{I_1-I_6, I_9, I_{15}-I_{16}\}$.

End

In Fig. 3, the whole image database includes 16 images (i.e. I_1-I_{16}). The OPOS diagram has three top clusters corresponding to the node I_1, I_{10} and I_{15} , respectively.

Suppose the query image contains semantic S_1 and semantic S_6 . The right of Fig. 3 is the D_a value of $\{S_1, S_6\}$ and other semantics. After the semantics complement by the association rule, the optimized image database includes 9 images (i.e. I_1 – I_6 , I_9 , I_{15} – I_{16}), and images (I_7 – I_8 , I_{10} – I_{14}) are removed. If we do not perform the association rule, only the images on the branch S_1 and S_6 will be retrieved, and the images (such as I_5 and I_6) with the similar semantics as the query image will be missed. It will cause the missing of the retrieval.

3.3 Attribute Partial Order Structure

In the APOS diagram, nodes represent attributes (visual semantics), and branches represent objects (images). The degree of similarity between two branches is evaluated by the number of identical nodes on the branches (images) in the APOS diagram. As well known, the images contain more identical attributes means that they have the higher similarity. In this paper, we calculate the number of identical attributes on the corresponding branches between the database and the query image, and sort the images of the database. The features of the top-ranked N images and the query image are averaged to obtain a new feature that replaces the feature of query image for the subsequent image retrieval. The algorithm of query expansion based on the APOS diagram is as follows:

Input: Query image, database and APOS.

1. Calculate the degree of association between query image and database according to the APOS diagram.
2. Sort the similarity of the query image and the database, and select the top N images that are most similar to the query image.
3. Average the features of the top-ranked N pictures and the feature of the query picture.
4. The obtained mean feature is used as the query feature for image retrieval.

End

The sorted result obtained by APOS is shown in Fig. 4, where the first column shows the query picture and the last is the top 5 pictures. It can be seen that most of the top 5 images are correctly retrieved. This paper performs the query expansion by selecting the top 5 images to improve the representation ability of the query image.

4 Experiments

4.1 Dataset

Two commonly used image databases, DupImage and Paris Buildings [22], are applied to evaluate the proposed retrieval method. DupImage database has 33 categories of icon images with a total number of 1188 images. Paris Buildings database contains 11 classes of building images with a total number of 6412 images. For DupImage database, all the images are used for the retrieval in the experiment. For Paris Buildings

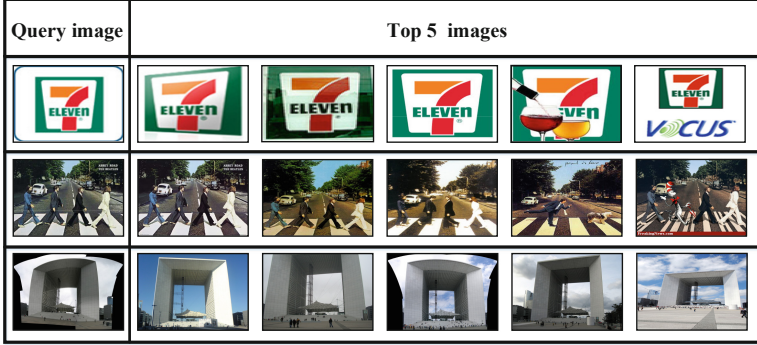


Fig. 4. The top 5 ranked results obtained by APOS

database, we randomly select 200 images for each category thus a total 2200 images to perform the image retrieval.

In Paris Buildings database, each query image has a groundtruth file which contains four types of labels: ‘good’, ‘ok’, ‘junk’ and ‘bad’. The images with the four labels are manually classified according to the similarity between the database and the query image. Visually, the images with the ‘good’ and ‘ok’ labels are similar to the query image, while the images with the ‘junk’ and ‘bad’ labels are not similar. Many previous works only used ‘good’ and ‘ok’ labels for image retrieval. In order to highlight the filtering ability of the partial order structure, this paper treats all the images under the category of the query image as similar images instead of using the ‘good’ and ‘ok’ labels given by the database.

4.2 Setup

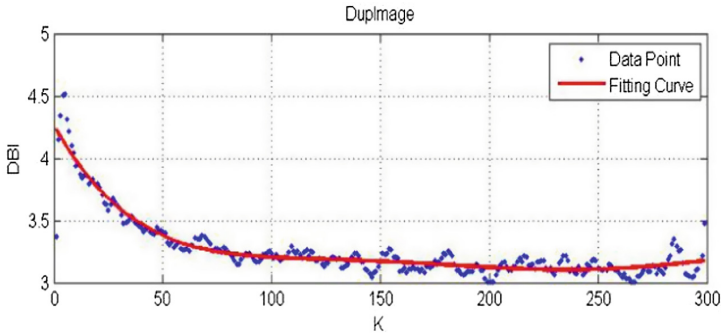
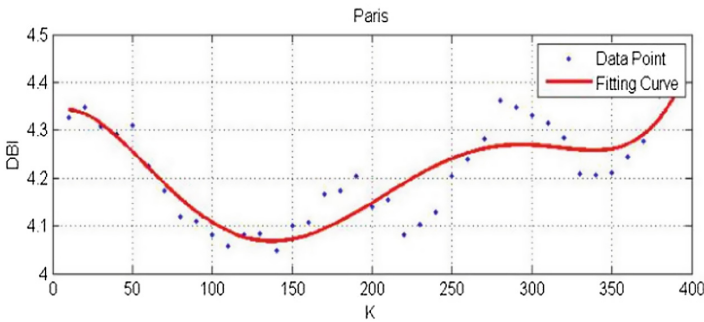
Suppose that there are M actual semantics for the database. After the clustering by the k -means method, k semantic centers are produced. If k is bigger than M , the semantics is over-clustering, that is, the semantics are not representative. If k is less than M , some actual semantics are merged, that is, some actual semantics are missing. Therefore, choosing an appropriate k value is necessary for the clustering. In this paper, we utilize Davies Bouldin index (DBI) to achieve the optimal k value automatically. The definition of DBI is as follows:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{\bar{C}_i + \bar{C}_j}{\|w_i - w_j\|_2} \right) \quad (2)$$

$$C_i = \frac{1}{T_i} \left(\sum_{p=1}^{T_i} |X_p - w_i|^2 \right)^{1/2} \quad (3)$$

Here C_i represents the average distance between all the data points in the i^{th} cluster and its center, which indicates the dispersion degree of the data points in clusters. X_p

represents the p^{th} data point. T_i represents the number of data points in the i^{th} cluster. w_i represents the clustering center of the i^{th} cluster. The data clustering is performed on two databases to obtain the *DBI* curve and get the appropriate k value. To observe the best k value clearly, the data points are fitted to a curve as shown in Fig. 5.

(a) *DBI* curve on DupImage database(b) *DBI* curve on Paris Building database**Fig. 5.** *DBI* curves on two image databases

A smaller *DBI* value means a better clustering effect. From Fig. 5(a) we can see that the value of *DBI* in the DupImage database becomes smaller as the k value increases. When the k value reaches 100, the curve tends to be stable. From Fig. 5(b), when the k value is about 150, the *DBI* of the Paris Building database reaches the lowest value. Therefore, we take 100 and 150 as the best k values for the DupImages database and Paris Building database, respectively.

In this paper, the parameters including the mean Average Precisions (*mAP*), Precision (*P*), Recall (*R*) and *F*-measure (*F*) are taken as the evaluation indexes to test the proposed OPOS and APOS methods. The result is reported as the average of the results from the 5 individual runs. The definition of *F* is as follows:

$$F = \frac{2PR}{P+R} \quad (4)$$

4.3 Experimental Results

In this section, we randomly select five images for each category in the Paris Building and DupImage databases as the query images, that is, there are 55 query images for the Paris Building database and 165 query images for the DupImage database. To demonstrate the efficiency of the experiment, we calculate the mAP , P , R and F to evaluate the proposed OPOS and APOS methods.

The P-R curves on two databases are plotted in Fig. 6. We compare the retrieval performance of three methods, OPOS, OPOS+APOS and CNN. The OPOS method indicates the database adaption by the semantics association rule based on the OPOS diagram. The OPOS+APOS method means the OPOS method combined with the query expansion based on the APOS diagram. The CNN method directly measures similarity of features extracted by the fully connected layer in VGG-f network without the database adaption and the query expansion. The number of the returned retrieved images is ranging from 40 to 200 with an increment by 20 for the DupImage database, and ranging from 50 to 600 with an increment by 50 for the Paris Building database.

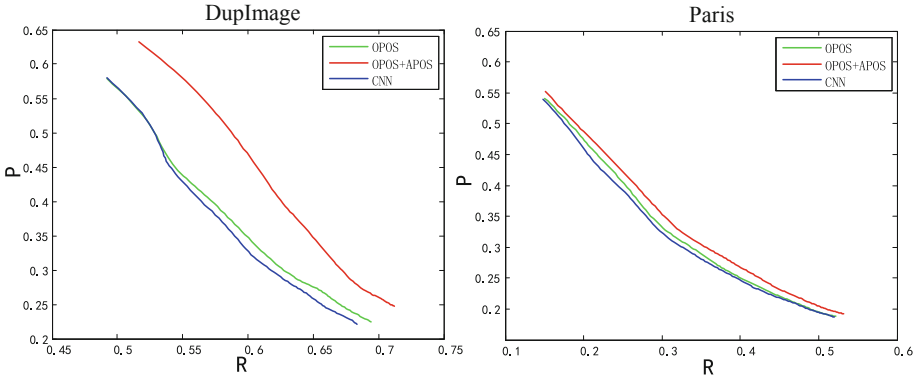


Fig. 6. P-R Statistical curves

From Fig. 6, the OPOS+APOS method outperforms other methods on two databases. The CNN method has the worst performance. Since the images in the different categories of the Paris Building database are highly similar, it is hard to increase the performance for the Paris Building database. From Fig. 6 we can see that the OPOS+APOS method improves better performance on DupImage database than Paris Building database.

In all, the OPOS method removes images that are not related to the query image and recalls some similar semantic images to improve the original image database, which indicates the clustering ability of partial order structure and the ability to filter

association rules. The APOS method enhances the expressive ability of the query image according to query expansion based on attribute partial order structure, Table 2 shows the average of P , R and F for the first W returned retrieved images, where W is the number of images of each category in the database. For two databases, the F value of the OPOS+APOS method is better than that of CNN method by about 4%.

Table 2. Comparison of the P , R , F on two databases

Method	DupImage			Paris		
	CNN	OPOS	OPOS+APOS	CNN	OPOS	OPOS+APOS
P	0.58	0.59	0.62	0.32	0.33	0.35
R	0.45	0.46	0.49	0.30	0.31	0.33
F	0.51	0.52	0.55	0.31	0.32	0.34

To further verify the superiority of the OPOS and APOS methods, Table 3 displays the mAP values of the three methods on the two image databases. For two databases, the mAP of the OPOS+APOS method is better than that of CNN method by about 4%. Need to point out, in Paris Building database, Since this paper does not use the groundtruth file that the database has given, compared with some works that only returns ‘good’ and ‘ok’ images as the similar images of the query image, the mAP obtained in this paper is lower.

Table 3. Comparison of mAP values on two databases

	DupImage	Paris
OPOS+APOS	0.56	0.38
OPOS	0.53	0.36
CNN	0.52	0.34

5 Conclusion

To better explore the relationship between images and the corresponding visual semantics, in this paper, we construct the formal background and establish the object partial order structure and attribute partial order structure according to the image and the visual semantics. We filter the irrelevant images and supplement the similar semantic images from the original image database by using the object partial order structure to improve the efficiency of the image retrieval. Also, we use the attribute partial order structure diagram to expand the query information of the database, which generalizes the characteristics of the query image. Experimental results demonstrate the validity of the proposed image retrieval method based on Object Partial Order Structure (OPOS) and Attribute Partial Order Structure (APOS). However, there is still a problem to be solved. Because the visual semantics obtained by the clustering are ambiguous, a few similar semantic images as the query image may be lost in the filtering process.

Acknowledgements. This work was partly supported by National Natural Science Foundation of China (No. 61303128), Natural Science Foundation of Hebei province (F2017203169), Key Scientific Research Projects of Colleges and Universities in Hebei province (ZD2017080), and Science and Technology Foundation for Returned Overseas People of Hebei Province (CL201621).

References

1. Li, S., Song, J., Hong, W., et al.: Partially ordered structure loop: a new method for visualization of big data's partial ordered structure. *J. Yanshan Univ.* **5**, 409–415 (2014)
2. Fan, F., Hong, W., Song, J., et al.: A visualization method for Chinese medicine knowledge discovery based on formal concept analysis. *Express Lett. Part B Appl. Int. J. Res. Surv.* **4**, 801–808 (2013)
3. Ganter, B., Eklund, P., Wille, R., et al.: Formal Concept Analysis. *Electron. Notes Discrete Math.* **2**(3), 199–200 (1999)
4. Yan, Z., Zhang, H., Piramuthu, R., et al.: HD-CNN: hierarchical deep convolutional neural network for large scale visual recognition. In: *ICCV* (2015)
5. Wan, J., Wang, D., Hoi, S.C.H., et al.: Deep learning for content-based image retrieval: a comprehensive study. In: *ACM International Conference on Multimedia* (2014)
6. Ozeki, M., Okatani, T.: Understanding convolutional neural networks in terms of category-level attributes. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) *ACCV 2014. LNCS*, vol. 9004, pp. 362–375. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16808-1_25
7. Babenko, A., Slesarev, A., Chigorin, A., Lempitsky, V.: Neural codes for image retrieval. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014. LNCS*, vol. 8689, pp. 584–599. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_38
8. Razavian, A.S., Azizpour, H., Sullivan, J., et al.: CNN features off-the-shelf: an astounding baseline for recognition. In: *CVPRW*, pp. 512–519 (2014)
9. Babenko, A., Lempitsky, V.: Aggregating local deep features for image retrieval. In: *ICCV*, pp. 1269–1277 (2015)
10. Sharif Razavian, A., Sullivan, J., Maki, A., et al.: A baseline for visual instance retrieval with deep convolutional networks. In: *ICLR* (2015)
11. Liu, Y., Guo, Y., Wu, S., et al.: Deep index for accurate and efficient image retrieval. In: *ICMR* (2015)
12. Gong, Y., Wang, L., Guo, R., Lazebnik, S.: Multi-scale orderless pooling of deep convolutional activation features. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014. LNCS*, vol. 8695, pp. 392–407. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10584-0_26
13. Jegou, H., Douze, M., Schmid, C., et al.: Aggregating local descriptors into a compact image representation. In: *CVPR*, pp. 3304–3311 (2010)
14. Arandjelović, R., Gronat, P., Torii, A., et al.: Net-VLAD: CNN architecture for weakly supervised place recognition. *Trans. Pattern Anal. Mach. Intell. TPAMI* **40**(6), 1437–1451 (2017)
15. Donahue, J., Jia, Y., Vinyals, O., et al.: DeCAF: a deep convolutional activation feature for generic visual recognition. In: *ICML* (2014)
16. Mohedano, E., Salvador, A., McGuinness, K., et al.: Bags of local convolutional features for scalable instance search. In: *ICMR* (2016)

17. Hong, W., Luan, J., Zhang, T., et al.: Knowledge discovery based on partial ordered structure theory. *J. Yanshan Univ.* **5**, 395–402 (2016)
18. Zhang, T., Ren, H., Wang, X., et al.: A calculation of formal concept by attribute topology. *ICIC Express Lett. Part B Appl. Int. J. Res. Surv.* **4**, 793–800 (2013)
19. Liu, S., Deng, W.: Very deep convolutional neural network based image classification using small training sample size. In: *ACPR*, pp. 730–734 (2015)
20. Ng, Y., Yang, F., Davis, L.: Exploiting local features from deep networks for image retrieval. In: *CVPRW*, pp. 53–61 (2015)
21. Davies, L., Bouldin, W.: A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1**(2), 224–227 (1979)
22. Philbin, J., Chum, O., Isard, M., et al.: Lost in quantization: improving particular object retrieval in large scale image databases. In: *CVPR* (2008)