Computer Modeling in Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2022.021287

ARTICLE

Refined Sparse Representation Based Similar Category Image Retrieval

Xin Wang, Zhilin Zhu and Zhen Hua*

School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai, China

*Corresponding Author: Zhen Hua. Email: hzsdgs@163.com

Received: 06 January 2022 Accepted: 14 March 2022

ABSTRACT

Given one specific image, it would be quite significant if humanity could simply retrieve all those pictures that fall into a similar category of images. However, traditional methods are inclined to achieve high-quality retrieval by utilizing adequate learning instances, ignoring the extraction of the image's essential information which leads to difficulty in the retrieval of similar category images just using one reference image. Aiming to solve this problem above, we proposed in this paper one refined sparse representation based similar category image retrieval model. On the one hand, saliency detection and multi-level decomposition could contribute to taking salient and spatial information into consideration more fully in the future. On the other hand, the cross mutual sparse coding model aims to extract the image's essential feature to the maximum extent possible. At last, we set up a database concluding a large number of multi-source images. Adequate groups of comparative experiments show that our method could contribute to retrieving similar category images effectively. Moreover, adequate groups of ablation experiments show that nearly all procedures play their roles, respectively.

KEYWORDS

Similar category; image retrieval; saliency detection; multi-level decomposition; cross mutual sparse coding

1 Introduction

With the advent of the big data era, similar category image retrieval plays a very significant application value in e-commerce, education, life science and other fields. In contrast, it is different from modern mainstream image retrieval technology. Actually, modern image retrieval is a process that involves the browsing, searching, and retrieval of target images from a database. Early on, the development of image retrieval spurred the development of many feature descriptors, such as color histogram [1], texture histogram [2], SIFT [3], rgSIFT [4], PHOG [5], and GIST [6]. Meanwhile, a large number of retrieval models, such as the Bayesian model [7,8], random forest model [9], and SVM model [10–12], directly support the retrieval process. Moreover, quite a few improved retrieval models [13–15] increase the accuracy of image retrieval significantly. However, the traditional retrieval models and other improved models are on the foundation of adequate learning instances. It is almost impossible to achieve high-quality retrieval with just a few learning instances, let alone only one reference image. Moreover, the fields of image classification [16], image annotation [17], and image recognition [18] also face similar issues.

The emergence of deep learning models [19,20] has completely changed the traditional retrieval mode. Traditional methods often rely on feature operators and retrieval models to achieve high precision retrieval. In contrast, the deep learning model directly uses a large number of learning instances to predict categories, trends and other information, fundamentally replacing some procedures of traditional methods. Without a doubt, more learning instances should be provided and more computing resources are consumed, but the accuracy of classification, retrieval and detection is significantly improved, which is highly recognized by scholars and engineering technicians. For instance, many remarkable application-based models [21,22] have been recognized fully owing to their remarkable experimental results. Especially in recent years, many innovative models have been applied to image retrieval, which improves the retrieval accuracy significantly. Specifically, innovative networks [23,24] were used to improve the efficiency of retrieval and more powerful feature representations [25,26] directly contribute to high-quality image retrieval.

As mentioned above, machine learning-based retrieval methods need adequate learning instances to support higher retrieval accuracy, especially the number of learning instances required by deep learning couldn't be underestimated. However, the classical methods cannot meet our requirements if we want to achieve similar category image retrieval based on reference image. Because the essence of machine learning is that a large number of learning instances could be used for tendency prediction, in contrast, similar category image retrieval is impossible without adequate learning instances. Although it has obvious limitations, which significantly increases the difficulty of implementation, we could still take advantage of the new strategy combined with valuable experience from traditional methods to achieve it. After all, it only needs to retrieve images of similar categories and the requirement is relatively lower than other classic methods. It is important to note that each image contains abundant semantic information [27,28] beyond our imagination, but it is not fully extracted by traditional classical methods. Moreover, the obvious distinctions between essential content and background content are not fully considered. More importantly, the obvious spatial information of the image is often ignored as well. Some previous methods [29,30] consider the above problems to a certain extent, and they can also retrieve many semantically valid images. However, the above methods do not deeply mine the essential features of the image, and do not fully consider the spatial location information of the image. These problems will directly lead to the decline of image retrieval accuracy.

Given the above problems, this paper has carried on further targeted research. On the one hand, saliency detection and multi-level decomposition are used to consider the image's essential information and spatial information into consideration to the maximum. On the other hand, the cross mutual sparse coding method is used to extract image features more accurately. In the process of similarity calculation, salient regions and multi-level regions are provided with different weights. More importantly, we have designed very comprehensive experiments to verify the effectiveness of the proposed method. The experimental results also fully prove the validity and robustness of our proposed method, especially for each critical procedure. On the whole, although our method can just retrieve similar category images limited by the semantic information of one reference image, it still has high-level practical application value. Our contributions are more reflected in the perspective of methods and ideas. In addition to machine learning, there are still many other image retrieval strategies having practical application value.

2 Algorithm

In the process of algorithm implementation, we first detect the salient region of a specific image. Next, we divide the specific image into different levels and spatial regions on different levels are

represented. In order to achieve a more accurate representation, the cross mutual sparse representation is used to realize it. Essentially, in the process of similar category image retrieval, the salient and spatial information are fully considered, which provides direct support for similar category image retrieval.

2.1 Pretreatment of Refined Sparse Representation

Since saliency detection can be realized conveniently, the Graph-Based Manifold Ranking strategy [31] was used for saliency detection. In essence, we ranked the similarity of the image elements (pixels or regions) with foreground cues or background cues via graph-based manifold ranking. Furthermore, as shown in Fig. 1, a two-stage scheme was used to achieve saliency detection by extracting background regions and foreground salient objects efficiently. Since manifold ranking is one important procedure, we illustrate it more clearly. The optimal ranking of queries would be computed through optimization problem solving, which is highly similar to the PageRank and spectral clustering algorithms.

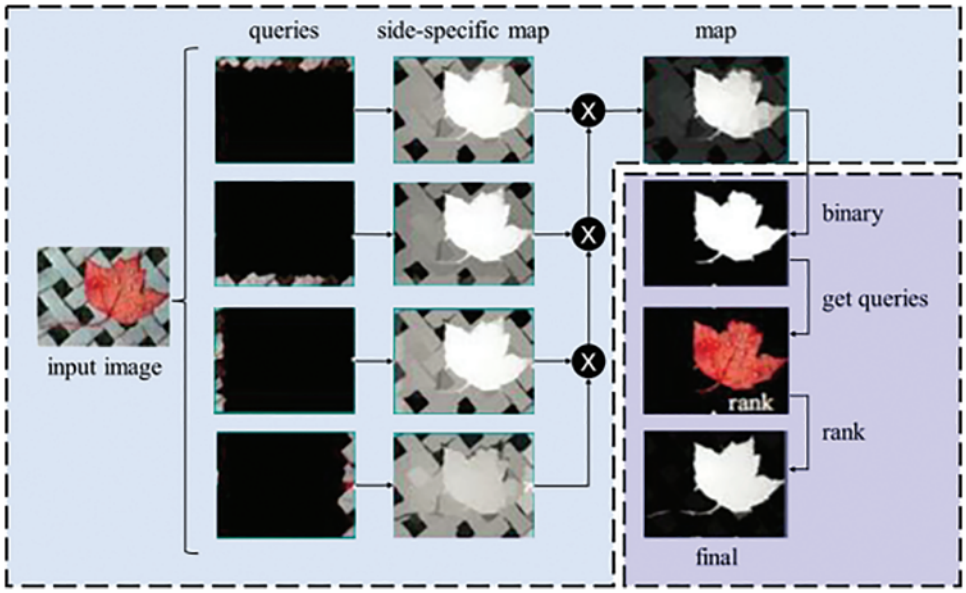

Figure 1: A two-stage scheme [31] was used to achieve saliency detection through extracting background regions and foreground salient objects efficiently

$$\mathbf{f}^* = \arg\min_f \frac{1}{2}\left(\sum_{i,j=1}^{n} w_{ij}\left\|\frac{f_i}{\sqrt{d_{ii}}} - \frac{f_j}{\sqrt{d_{jj}}}\right\|^2 + \mu\sum_{i=1}^{n}\|f_i - y_i\|^2\right) \tag{1}$$

In Eq. (1), the first term is smoothness constraint and the second term is fitting constraint, and they are balanced by the parameter μ. Specifically, a high-level ranking function should not change too much between surrounding points and should not differ too much from the initial query assignment. Actually, the minimum solution could be computed through setting the derivative of above function to be zero. Next, resulted ranking function could be expressed as Eqs. (2)–(3), where I is an identity matrix and S is the normalized Laplacian matrix.

$$\mathrm{f}^* = (\mathrm{I} - \alpha\mathrm{S})^{-1}\mathrm{y} \tag{2}$$

$$\alpha = 1/(1+\mu) \tag{3}$$

The ranking algorithm [32] draws lessons from semi-supervised learning [33] for classification. Essentially, the manifold ranking could be considered a one-class classification problem.

Next, in order to further consider the spatial information of the image, we divide an image into different levels. Spatial information will be more detailed with a higher level. As shown in Fig. 2, when an image is decomposed into more levels, spatial information will be considered more completely. Actually, the quality of image feature extraction and similarity calculation largely depends on the fine degree of image decomposition.

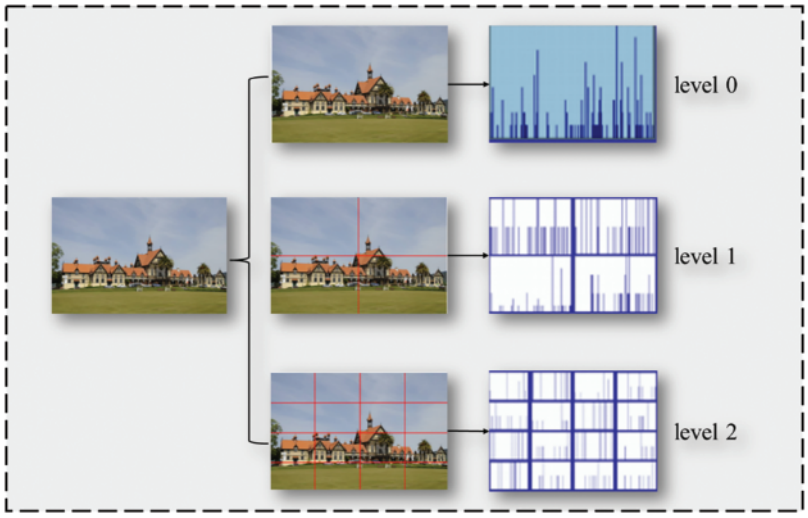

Figure 2: A specific image is divided into different levels and spatial information could be considered by decomposition

As the essential semantic region of the whole image, the saliency region could be detected, which should be given prioritized in the similarity calculation process. Moreover, the spatial information of the image also needs to be fully considered. After all, even if the same content appears in different regions, it may represent different meanings. The above algorithms fully consider these problems and lay a foundation for accurate image similarity calculation.

2.2 Refined Sparse Representation Based Retrieval

On the foundation of the pretreatment process, we use a cross mutual sparse coding method to represent images. In order to achieve accurate and effective image representation, we train millions of images to reduce the error caused by randomness. When each image is represented precisely, image similarity can be calculated expediently.

The sparse coding based model [34,35] represents one image by training a group of patches to minimize the following step-by-step equation, where x_i refers to input learning images, ϕ indicates words in the dictionary, and $a_{i,j}$ represents sparse parameters. In training, m, k, j, and i are adjusted multiple times to determine the best base groups. Eq. (4) and procedures (a) and (b) present the sparse-coding process.

$$\min_{a,\phi} \sum_{i=1}^{m} ||x_i - \sum_{j=1}^{k} a_{i,j}\phi_j||^2 + \lambda \sum_{i=1}^{m} \sum_{j=1}^{k} |a_{i,j}| \tag{4}$$

(a) Adjustment while locking ϕ to minimize the target function.

(b) Adjusting ϕ while locking a to minimize the target image.

One group of bases can be obtained through sufficient iterations until convergence. Although this set of bases can comparatively provide better representation for images, we find this set of bases is still not unreasonable after adequate experiments. Because of the aforementioned problem, we do more groups of experiments to obtain different groups of bases. More importantly, let them test and verify with each other. On the foundation, we select the most reasonable set of bases. Each image could be expressed more accurately owing to a more precise set of base vectors.

$$\min_{a} \sum_{i=1}^{m} ||x_i - \sum_{j=1}^{k} a_{i,j}\phi_j||^2 + \lambda \sum_{i=1}^{m} \sum_{j=1}^{k} |a_{i,j}| \tag{5}$$

When a set of high-quality base vectors has been obtained, a similar cost function minimization process is used to implement a sparse representation. Compared with Eqs. (4) and (5), it is no need to calculate the base vector. It only needs to obtain the sparse vector by using other parameters. Although the base vector has been locked, the other parameters still need to be adjusted step by step to minimize the objective function. When the objective function has been minimized, we obtain the encoding coefficients for achieving a sparse coding representation of the image. Fig. 3 shows one typical instance of an image region represented by one group of bases. Although the process is relatively complex, it can not only realize the accurate image representation, but also provide support for the similarity calculation of different images.

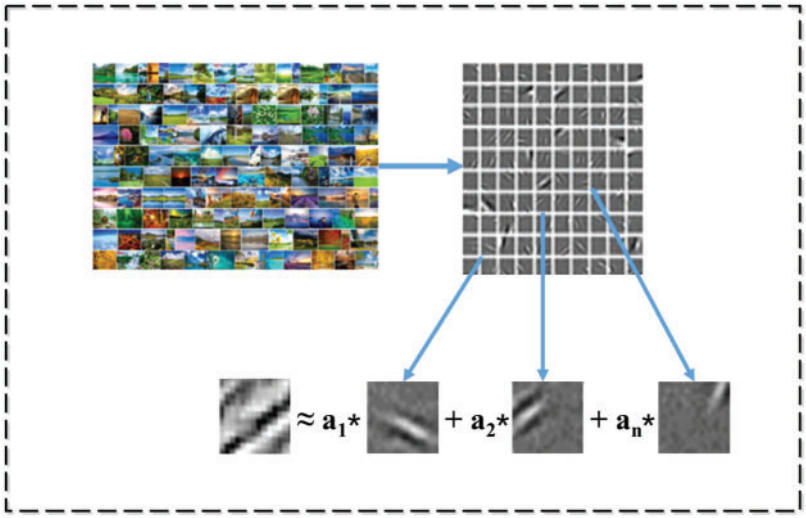

Figure 3: One typical instance of image region represented by one group of bases

With precise sparse representation, we can calculate the similarity of different regions. Specifically, suppose we want to calculate the similarity between regions (a) and (b). In that case, be expressed as

the Eq. (6), where C represents sparse coding-based function and k represents the dimension of sparse coding.

$$S(a,b) = \sqrt{\sum_{k=1}^{k=n} (C_k(a) - C_k(b))^2} \tag{6}$$

On this basis, we calculate the similarity of different images and characterize it as two parts. For the first part, the similarity measure mainly focuses on the similarity of saliency regions between different images. For the second part, the similarity measure mainly focuses on the similarity of different levels between different images. When the similarity between the candidate retrieved image and the reference image exceeds the set threshold, we consider it as one targeted retrieval image owing to the high similarity between each other.

The above section shows the key modules of our method in detail, especially for the essential modules; we lay extra emphasis on them with more targeted descriptions. In general, each module supports and complements each other mutually. What's more, there is no very high dependence between each module. In other words, even if a single module could not play a significant role, it still would not have a negative influence on other modules, which further improves the method's feasibility and lays a foundation for superior experimental results in a convincing database.

3 Experiments

In order to make the experimental results more credible, we selected 257,823 images from the Internet database, personally collected database and some classic databases. More importantly, images with different resolutions or complicated semantic content would be selected as more as possible. Meanwhile, in order to make the experimental results more objective, we use a subjective system to evaluate the image similarity. Specifically, we then determine the correctness when two of the three volunteers consider that the reference image and retrieval images are highly similar. It should be noted that some baseline methods are not suitable for retrieval directly, but we use their strategies or essential algorithms to achieve comparison.

From the AP values in Fig. 4, our retrieval accuracy has significantly improved compared to the traditional baseline methods. Meanwhile, the experimental results are evaluated by AUC values in Fig. 5. Similarly, AUC values further illustrate the effectiveness of our method.

In order to test the effectiveness of saliency detection, multi-level decomposition and cross mutual sparse representation, we added some ablation experiments. The experimental results in Fig. 6 fully reflect the necessity of the three procedures, which provide direct support for improving the retrieval accuracy. Similarly, AUC values in Fig. 7 further illustrate the necessity of the three procedures.

Considering the diversity of the images, we have carried out further experiments on complex semantic-based images, low resolution-based images and weak saliency region-based images. From the experimental results by AP and AUC values in Figs. 8 and 9, we perceive these special images could not be retrieved very realistically, but our method still has a certain application value for general retrieval.

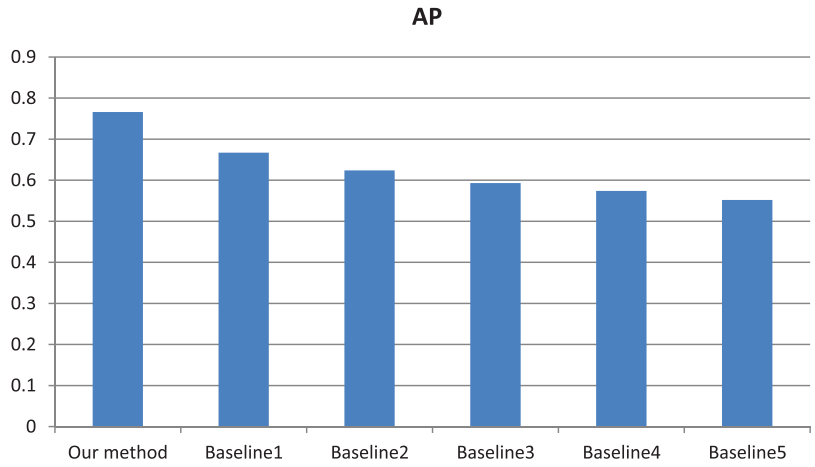

Figure 4: AP values calculated using different methods. Baseline 1: The improved sparse coding based method [36]. Baseline 2: Weighted spatial pyramid matching based method [37]. Baseline 3: The practical-based method [38]. Baseline 4: The spatial pyramid matching based method [39]. Baseline 5: The GIST-based similarity calculation method [40]

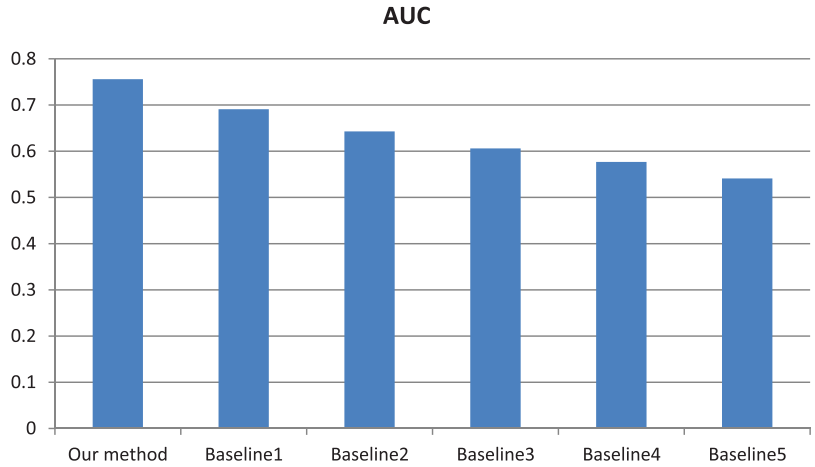

Figure 5: AUC values calculated using different methods. Baseline 1: The improved sparse coding based method [36]. Baseline 2: Weighted spatial pyramid matching based method [37]. Baseline 3: The practical-based method [38]. Baseline 4: The spatial pyramid matching method [39]. Baseline 5: The GIST-based similarity calculation method [40]

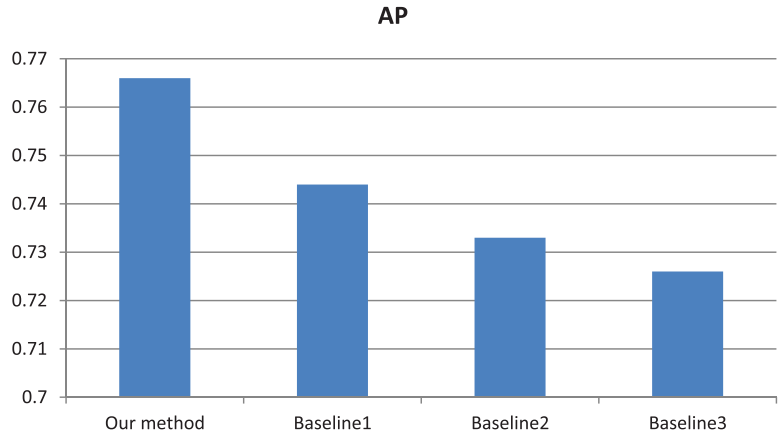

Figure 6: Comparison of AP values between our model and baseline models. Baseline 1 model is that we retrieve similar category images using our model just without the saliency detection step. Baseline 2 model is that we retrieve similar category images using our model just without using multi-level decomposition step. Baseline 3 model is that we retrieve similar category images using our model just without using cross mutual sparse representation

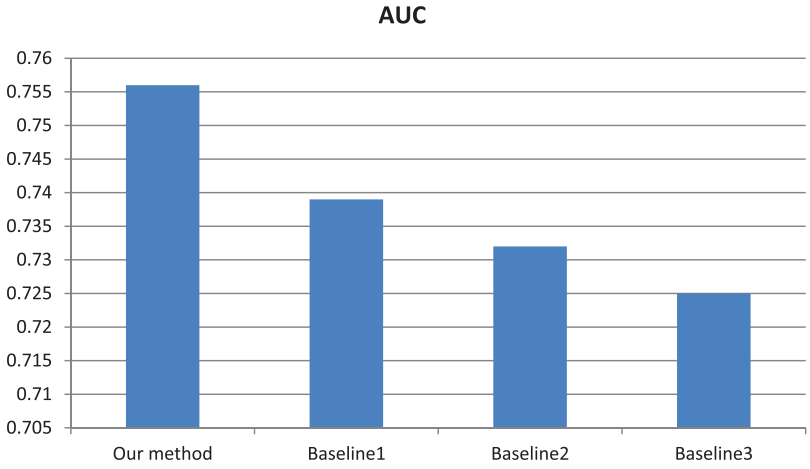

Figure 7: Comparison of AUC values between our model and baseline models. Baseline 1 model is that we retrieve similar category images using our model just without the saliency detection step. Baseline 2 model is that we retrieve similar category images using our model just without using a multi-level decomposition step. Baseline 3 model is that we retrieve similar category images using our model just without using cross mutual sparse representation

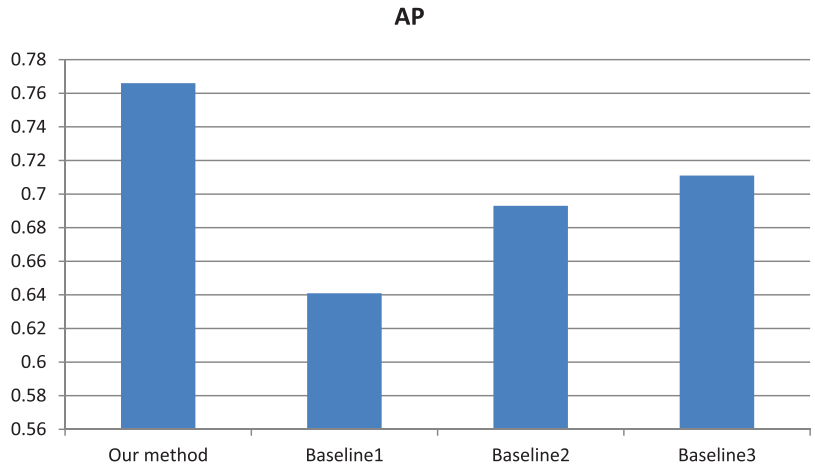

Figure 8: Comparison of AP values by our method based on different types of images. The AP values are experimental results based on universal images, complex semantic-based images, low resolution-based images and weak saliency region-based images, respectively

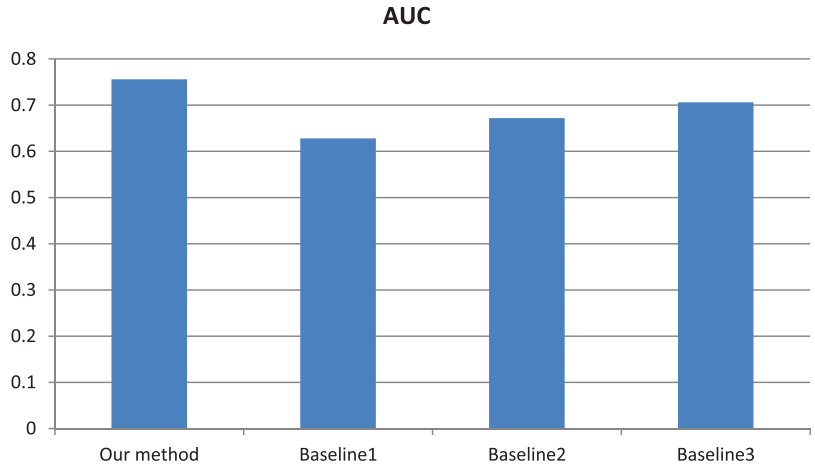

Figure 9: Comparison of AUC values by our method based on different types of images. The AUC values are experimental results based on universal images, complex semantic based images, low resolution based images and weak saliency region based images, respectively

Furthermore, to prove the universality of our method, we have added new groups of experiments. The image database was randomly divided into five groups, each of which underwent the corresponding retrieval experiment. It can be seen from the experimental results in Figs. 10 and 11 show no significant difference between each group of experimental results, which further proves that our method is robust. In addition, in order to make the experimental results more visible and intuitive, we spread four typical groups of retrieval experimental results in Figs. 12–15. Although the experimental results that can be displayed are limited, these experimental results are still highly representative. It

should be noted that the traditional feature operator-based method is also very effective while saving considerable computing time for some non-complex images. Therefore, for some non-complex images, we use the traditional feature operator-based retrieval method directly considering the efficiency factor. In order to further verify our method, we divided the images into architecture, scene, and other categories and tested them, respectively. Moreover, we artificially add some noise to some images. We find that the experimental results are basically unchanged, which further proves the effectiveness and robustness of our method.

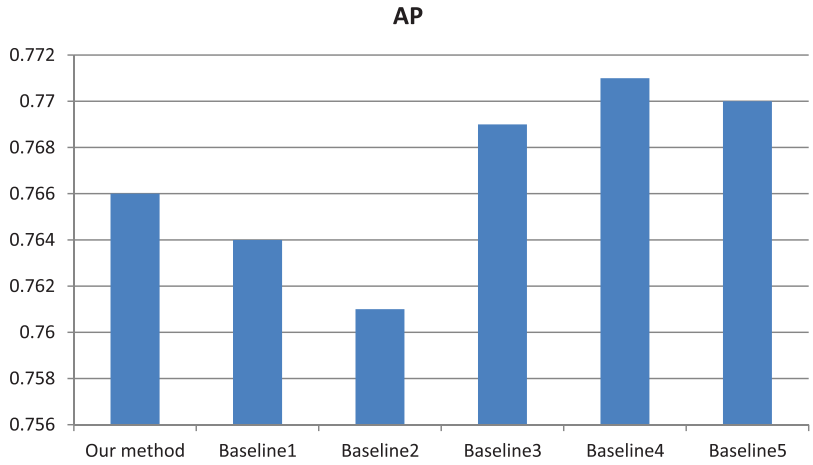

Figure 10: Comparison of AP values by our method based on different groups of images. The AP values are experimental results based on the original database and five other random divided groups of images, respectively

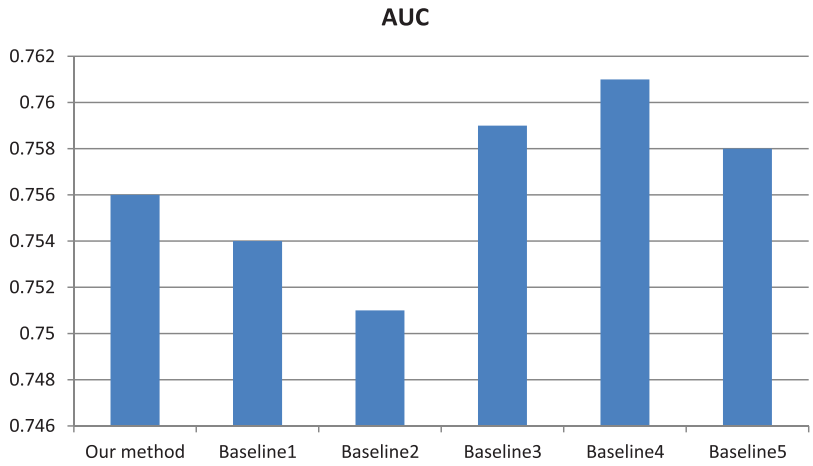

Figure 11: Comparison of AUC values by our method based on different groups of images. The AP values are experimental results based on original database and five other random divided groups of images, respectively

Figure 12: Input images and similar images retrieved using our method. The image at the top is the input image, while the images at the bottom are the similar images retrieved using our method. Special note: our method also retrieve a lot of unreasonable images, here we mainly select some reasonable ones for display

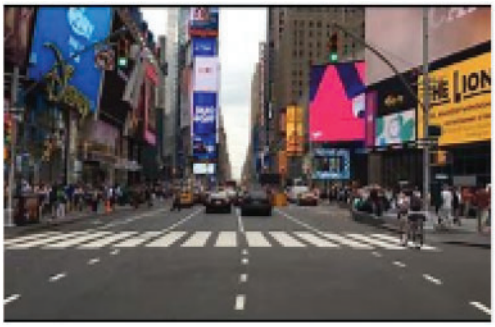

Figure 13: (Continued)

Figure 13: Input image and its similar images retrieved by our method. The image at the top is the input image and the other images are its similar to ours. Special note: our method also retrieves a lot of unreasonable images. Here, we mainly select some reasonable ones for display

Figure 14: Input image and its similar images retrieved by our method. The image at the top is the input image and the other images are its similar images retrieved by our method. Special note: our method also retrieve a lot of unreasonable images, here we mainly select some reasonable ones for display

Figure 15: Input image and its similar images retrieved by our method. The image at the top is the input image and the other images are similar to those retrieved by our method. Special note: our method also retrieve a lot of unreasonable images, here we mainly select some reasonable ones for display

On the whole, the above experiments are in line with our expectations. Both AP values and AUC values prove the effectiveness of our method. More importantly, ablation experiments have shown that several essential procedures are also indispensable. Based on relatively realistic experimental results, we analyze the reasons in depth. On the one hand, saliency detection and multi-level decomposition contribute to considering the image relationship from multiple dimensions. On the other hand, the cross mutual sparse coding strategy extracts the feature information deeper. In addition, the rationality of detail processing and strategy design also provide direct support.

Nevertheless, we also found some limitations in our proposed method. Even if we thoroughly consider the characteristics of an image very thoroughly, we cannot achieve very realistic retrieval results compared to some powerful machine-learning-based methods. After all, the reference image of our proposed method is just one image, and the essential information which could be extracted is limited. At the same time, image similarity judgment often has a certain degree of subjectivity. Although we invite a number of volunteers to judge, there will still be some errors. Therefore, our work still needs to be further optimized and improved.

4 Conclusion

In this paper, we proposed one refined sparse representation based similar category image retrieval method. Considering the saliency region and spatial information, this method fully extracts the image's essential information to the maximum extent possible. Sufficient experiments have proved that similar category images can be retrieved by the limited reference image instead of a learning instance-based machine learning model.

To be fair, our method has many shortcomings or limitations. After all, the information in an image is limited even if we make the most of it. For instance, if the semantic information of an image is too complex, our method is not very effective. Moreover, many factors of the reference image will directly affect the retrieval accuracy. Despite the fact that our method is not suitable for very realistic retrieval, we still provide an innovative strategy for retrieval generally.

Funding Statement: This research is sponsored by the National Natural Science Foundation of China (Grants: 62002200, 61772319), Shandong Natural Science Foundation of China (Grant: ZR2020QF012).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Swain, M. J., Ballard, D. H. (1992). *Indexing via color histograms*. Germany: Springer Berlin Heidelberg.
2. Carson, C., Thomas, M., Belongie, S., Hellerstein, J. M., Malik, J. (1998). Blobworld: A system for region-based image indexing and retrieval. *Third International Conference on Visual Information and Information Systems*, pp. 509–517. Berlin, Heidelberg.
3. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60(2),* 91–110. DOI 10.1023/B:VISI.0000029664.99615.94.
4. Sande, K., Gevers, T., Snoek, C. (2008). Evaluation of color descriptors for object and scene recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. Anchorage.
5. Bosch, A., Zisserman, A., Munoz, X., Zisserman, P. (2007). Representing shape with a spatial pyramid kernel. *ACM International Conference on Image and Video Retrieval*, pp. 401–408. University of Girona, Girona, Spain; University of Oxford, Oxford, UK.
6. Oliva, A. (2006). Building the gist of a scene: The role of global image features in recognition. In: *Visual perception (Part B), fundamentals of awareness, multi-sensory integration and high-order perception,* vol. 155, pp. 23–36. DOI 10.1016/S0079-6123(06)55002-2.
7. Glowacka, D., Teh, Y. W., Shawetaylor, J. (2016). Image retrieval with a Bayesian model of relevance feedback. arXiv:1603.09522.
8. Chen, T. (2018). An image retrieval method based on multi-instance learning and Bayesian classification. *Journal of Shenzhen Polytechnic, 17(3),* 7–11.
9. Kokare, M., Bhosle, N. (2020). Random forest-based active learning for content-based image retrieval. *International Journal of Intelligent Information and Database Systems, 13(1),* 72–88. DOI 10.1504/IJIIDS.2020.108223.
10. Lin, Z., Yang, Y., Cao, H. (2011). Large-scale image classification: Fast feature extraction and SVM training. *The 24th IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1689–1696. Colorado Springs, CO, USA.

11. Pasolli, E., Melgani, F., Tuia, D., Pacifici, F., Emery, W. J. (2014). Svm active learning approach for image classification using spatial information. *IEEE Transactions on Geoscience & Remote Sensing, 52(4),* 2217–2233. DOI 10.1109/TGRS.36.
12. Hoi, S., Rong, J., Zhu, J., Lyu, M. R. (2008). Semi-supervised SVM batch mode active learning for image retrieval. *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1–7. Anchorage, Alaska, USA.
13. Singh, S., Batra, S. (2020). An efficient bi-layer content based image retrieval system. *Multimedia Tools and Applications, 79(5)*. DOI 10.1007/s11042-019-08401-7.
14. Niu, D., Zhao, X., Lin, X., Zhang, C. (2020). A novel image retrieval method based on multi-features fusion. *Signal Processing Image Communication, 87(9),* 115911. DOI 10.1016/j.image.2020.115911.
15. Min, W., Mei, S., Li, Z., Jiang, S. (2020). A two-stage triplet network training framework for image retrieval. *IEEE Transactions on Multimedia, 22(12),* 3128–3138. DOI 10.1109/TMM.2020.2974326.
16. Kumar, N., Berg, A. C., Belhumeur, P. N., Nayar, S. K. (2009). Attribute and simile classifiers for face verification. *2009 IEEE 12th International Conference on Computer Vision*, pp. 365–372. Kyoto, Japan.
17. Chong, W., Blei, D. M., Li, F. F. (2009). Simultaneous image classification and annotation. *IEEE Conference on Computer Vision & Pattern Recognition*, pp. 1903–1910. Miami, Florida, USA.
18. Keysers, D., Deselaers, T., Gollan, C., Ney, H. (2007). Deformation models for image recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 29(8),* 1422–1435. DOI 10.1109/TPAMI.2007.1153.
19. He, K., Zhang, X., Ren, S., Sun, J. et al. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 37(9),* 1904–1916. DOI 10.1109/TPAMI.2015.2389824.
20. Jiang, W., Yi, Y., Mao, J., Huang, Z., Wei, X. (2016). Cnn-rnn: A unified framework for multi-label image classification. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2285–2294. Las Vegas, NV, USA.
21. Ren, S., He, K., Girshick, R., Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(6),* 1137–1149. DOI 10.1109/TPAMI.2016.2577031.
22. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S. et al. (2016). *SSD: Single shot multibox detector*. Springer: Springer, Cham.
23. Song, J. (2020). Binary generative adversarial networks for image retrieval. *International Journal of Computer Vision,* 1–22. DOI 10.1007/s11263-020-01305-2.
24. Staszewski, P., Jaworski, M., Cao, J., Rutkowski, L. (2021). A new approach to descriptors generation for image retrieval by analyzing activations of deep neural network layers. *IEEE Transactions on Neural Networks and Learning Systems*, 1–8. DOI 10.1109/TNNLS.2021.3084633.
25. Sun, K., Zhu, J. (2021). Learning representation of multi-scale object for fine-grained image retrieval. *2021 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1660–1664. Toronto, ON, Canada.
26. Chakraborty, S., Singh, S. K., Chakraborty, P. (2022). Cascaded asymmetric local pattern: A novel descriptor for unconstrained facial image recognition and retrieval. *Multimedia Tools and Applications, 78,* 25143–25162. DOI 10.1007/s11042-019-7707-0.
27. Dong, G., Yan, Y., Shen, C., Wang, H. (2021). Real-time high-performance semantic image segmentation of urban street scenes. *IEEE Transactions on Intelligent Transportation Systems, 22(6),* 3258–3274. DOI 10.1109/TITS.2020.2980426.
28. Wang, L., Qian, X., Zhang, Y., Shen, J., Cao, X. (2020). Enhancing sketch-based image retrieval by cnn semantic re-ranking. *IEEE Transactions on Cybernetics, 50(7),* 3330–3342. DOI 10.1109/TCYB.6221036.

29. Wu, H., Li, Y., Xiong, J., Bi, X., Zhang, L. et al. (2019). Weighted-learning-instance-based retrieval model using instance distance. *Machine Vision and Applications, 30,* 163–176. DOI 10.1007/s00138-018-0988-x.
30. Wu, H., An, D., Zhu, X., Zhang, Z., Hua, Z. (2021). Multi-source material image optimized selection based multi-option composition. *Image and Vision Computing, 107(3),* 104123. DOI 10.1016/j.imavis.2021.104123.
31. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M. H. (2013). Saliency detection via graph-based manifold ranking. *Computer Vision & Pattern Recognition*, pp. 3166–3173. Portland, OR, USA.
32. Zhou, D., Weston, J., Gretton, A., Bousquet, O., Schölkopf, B. et al. (2003). Ranking on data manifolds. *Neural Information Processing Systems,* 169–176.
33. Zhou, D., Bousquet, O., Lal, T. N., Weston, J., Olkopf, B. S. (2004). Learning with local and global consistency. *Advances in Neural Information Processing Systems, 16*(*3*).
34. Arora, S., Ge, R., Ma, T., Moitra, A. (2015). Simple, efficient, and neural algorithms for sparse coding. *Proceedings of the 28th Conference on Learning Theory*, pp. 113–149. Paris, France.
35. Yang, J., Kai, Y., Gong, Y., Huang, T. S. (2009). Linear spatial pyramid matching using sparse coding for image classification. *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1794–1801. Miami, Florida, USA.
36. Sun, X., Nasrabadi, N. M., Tran, T. D. (2019). Supervised multilayer sparse coding networks for image classification. *IEEE Transactions on Image Processing, 29,* 405–418.
37. Liu, B. D., Meng, J., Xie, W. Y., Shao, S., Li, Y. et al. (2019). Weighted spatial pyramid matching collaborative representation for remote-sensing-image scene classification. *Remote Sensing, 11*(*5*). DOI 10.3390/rs11050518.
38. Andrea, V., Andrew, Z. (2011). Image classification practical. http://www.robots.ox.ac.uk/∼vgg/share/practical-image-classification.html.
39. Lazebnik, S., Schmid, C., Ponce, J. (2006). Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2169–2178. New York, NY, USA.
40. James, H., Alexei, E. A. (2007). *Scene completion using millions of photographs*. New York, NY, USA: Association for Computing Machinery. DOI 10.1145/1276377.1276382.