Bagging Based Cross-Media Retrieval Algorithm

Gongwen Xu

Shandong Jianzhu University

Yu Zhang

Shandong Jianzhu University

Mingshan Yin

Shandong Jianzhu University

Wenzhong Hong

Shandong Jianzhu University

Ran Zou (✉ 12847@sdjzu.edu.cn)

Shandong Jianzhu University

Shanshan Wang

Shandong Jianzhu University

Research Article

Keywords: Cross-Media Retrieval, Bagging, Ensemble Learning

Posted Date: December 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1159198/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Bagging Based Cross-Media Retrieval Algorithm

Gongwen Xu¹, Yu Zhang¹, Mingshan Yin¹, Wenzhong Hong¹, Ran Zou^{2*}, Shanshan Wang³

¹ School of Business, Shandong Jianzhu University, Jinan, China

² School of Management Engineering, Shandong Jianzhu University, Jinan, China

³ School of Business, Shandong University of Political Science and Law, Jinan, China

Corresponding author: Ran Zou (e-mail: 12847@sdjzu.edu.cn).

ABSTRACT It is very challenging to propose a strong learning algorithm with high prediction accuracy of cross-media retrieval, while finding a weak learning algorithm which is slightly higher than that of random prediction is very easy. Inspired by this idea, we propose an imaginative Bagging based cross-media retrieval algorithm (called BCMR) in this paper. First, we utilize bootstrap sampling to carry out random sampling of the original training set. The amount of the sample abstracted by bootstrap is set to be same as the original dataset. Second, 50 bootstrap replicates are used for training 50 weak classifiers independently. We take advantage of homogenous individual classifiers and integrate eight different baseline methods in our experiments. Finally, we generate the final strong classifier from the 50 weak classifiers by the integration strategy of sample voting. We use collective wisdom to eliminate bad decisions so that the generalization ability of the integrated model could be greatly enhanced. Extensive experiments performed on three datasets show that BCMR can effectively improve the accuracy of cross-media retrieval.

INDEX TERMS Cross-Media Retrieval, Bagging, Ensemble Learning

I. INTRODUCTION

With the arrival of the Internet era, the representation form of the data is more flexible and versatile. As a result, the multimedia data emerges at a fast rate in live platforms and Internet, such as social networking sites and game sites. Media data in different forms of expression which can express the same information from their respective angles are beginning to combine to form complex relationships and organizational structures. The explosion of different types of data information stimulates the emergence of cross-media retrieval [1,2]. Cross-media retrieval breaks through the limitations of traditional single media retrieval methods [3,4], it can retrieve specific types of information that we need from data of great capacity for liquor magnanimity. It presents the data in a more colorful and comprehensive way comparing with the single media retrieval and satisfy the growing needs of customers for data retrieval. The retrieval of different types of data is shown in Fig. 1. When searching GPU on Amazon, we can get the different information about GPU, the text, image, audio and video. The performance and the price of the GPU can be found in the text, or the customers' voice message. The appearance of the GPU would be shown from pictures of different views. The usage, the installation and other more information could be seen in the videos. Cross-media retrieval refers to that the system automatically retrieves all media content related to the query subject when the user gives any media query, such as pictures. It is designed for scenes with different media types of query and retrieval results. People have diverse needs for online search. When one type of search results cannot meet the requirements, cross-media retrieval came into being.

FIGURE 1. The Concept about Cross-Media Retrieval

Cross-media retrieval is difficult and challenging because the data information is generally characterized by the following characteristics. The hybrid data of different modalities have complex organization structures, and most of them are unstructured or semi-structured, so it is difficult to be structured for storage and retrieval. The correlation between different modalities is usually implicit and not easy to be excavated. The existence of noise in different modalities data is very easy to destroy the one-to-one correspondence between the cross-media description, and misunderstands the semantics expressed by the data itself. There may be many descriptions of an object in the same modality. We need to use the description of other modalities to eliminate redundant descriptions and save storage space. Owing to the above features of cross-media data are difficult to handle, and it is expressed with the underlying features of different dimensions and different attributes. In the meanwhile, the characteristic isomerism and the incommensurability of different modalities data, which makes the correlation

measurement of the cross-media data become very difficult. There is natural gap between the low-level features and high-level semantics, so the semantic gap is the unavoidable issue in the field of cross-media retrieval. In order to solve these problems, researchers at home and abroad have proposed a variety of novel cross-media retrieval methods. Even the neural network [5,6] is used to train data in order to get an advanced method with high accuracy.

Motivated by the fact that integrating a weak learning algorithm is easier than finding a strong learning algorithm, an imaginative Bagging based cross-media retrieval algorithm (called BCMR) is proposed in this paper. Considering that a single training set is not sufficient to provide enough and effective information, we take advantage of bootstrap sampling [7] to process the training set, and form 50 bootstrap replicates. Therefore, the diversity of training data has been greatly enhanced. A single learner may acquire the sub-optimal results, so integrating multiple learners will use the wisdom of the group to get the best value to some extent. The primary purpose is to employ ensemble learning to cross-media retrieval in our paper.

The rest content was organized as follows. In Section 2, the existing cross-media retrieval methods from two main approaches, including similarity measure and common space learning, was introduced. In Section 3, we present the formal description of bootstrap sampling and Bagging. In Section 4, the experimental results of ensemble learning in cross-media retrieval from various aspects were brought out. In the end the conclusion of this paper was drawn in section 5.

II. RELATED WORK

In these days, the online multimedia data has a rapid development and attracts extensive attention. Researchers' exploration of cross-media retrieval is also in full swing. This retrieval mode has been enhanced by scholars in the fields of pattern recognition, probability statistics, and graph theory. This retrieval methods could be divided into two types of representation, binary and real-value[8,9]. There are some popular real-value representation, such as deep learning methods[10], dictionary learning methods[11,12] and graph correlation methods[13,14]. These methods have high accuracy and are widely used.

Deep learning plays a great role in cross-media retrieval in recent years, and deep neural network[15] has become popular in this field. Deep neural network could deeply mine the relation among different modes data. DCCA was proposed by Andrew et al. [16], which is a parameter model that needn't consider the training data when calculating the characteristic description. DCCA refers to Deep Canonical Correlation Analysis. There are two deep neural networks in DCCA, so in computing complexity, DCCA method is scalable. The isomorphic features of different modal data are described by learning the multiple nonlinear interchange. The generative adversarial network could mine the different modes' joint distribution, so Peng et al.[17] made use of generator and discriminator to capture the information among heterogeneous data and catch the semantic conformance of different modes.

Traditional dictionary learning methods consider sparse representation when retrieving in different data modes. New methods of dictionary learning are constantly emerging

recently. A novel method was carried out by Shang et al.[18]. This method projects heterogeneous data into an isomorphic subspace using the representation coefficients and it is effective in cross-media retrieval. There also some methods [19] extending single type media to cross-media scene, which are called coupled dictionary learning. Querying information will be mapped into other type space under sparse coefficient map. Considering task-driven, the modified dictionary learning was brought out by Bahrampour et al. [20]. In this method, the dictionaries and classifiers of different modes are obtained simultaneously.

There are two kinds methods being involved with graph, one is graph regularization, the other is graph-based methods. The first one often appears in semi-supervised learning and could express intra-pair and inter-pair similarity. A new method, joint representation learning (JRL) was carried out by Zhai et al.[21]. Besides of the original label information, this method could mine the pairwise correlation information fully. Peng et al. [22] designed a graph considering both different modalities samples and their patches, which can develop local information. Graph based methods design different independent graph for different modality. Tong et al.[23] conducted graphs and learned the task according to the supervision information and the graph constraints.

There are three contributions in this paper: (1) We put forward BCMR, which is the first approach to combine Bagging with cross-media retrieval. (2) We analyzed the effect of Bagging in different datasets, and thus the scope of application of Bagging in cross-media was also shown in our work. (3) By integrating different methods, we evaluate the stability of different algorithms in terms of the degree of accuracy improvement.

III. THE PROPOSED METHOD

A. BOOTSTRAP SAMPLING

Bootstrap [24] uses the original data to simulate the sampling statistical inference. It can be used to work on the distribution characteristics of a set of data statistic, especially for the problems such as interval estimation, hypothesis testing and so on, which are difficult to be exported by conventional methods. The basic idea is to make a resampling in the scope of the original data. The probability of each number being pumped every time in the original data is equal, and the sample is called the bootstrap samples [25]. After M time sampling, the probability that the sample is not always sampled is $(1 - \frac{1}{m})^m$. We can get the following expression by taking the limit.

$$\lim_{m \rightarrow \infty} \left(1 - \frac{1}{m}\right)^m \rightarrow \frac{1}{e} \approx 0.368 \quad (1)$$

B. FORMAL DESCRIPTION OF BAGGING

For the classification problem, based on its basic assumption that the predictor $\phi(x, L)$ predicts the class label $s \in \{1, \dots, S\}$. We draw L based on the distribution P . In the case of ensuring that Y , X come from P independent of L , correct classification's probability for L fixed is [26, 27]:

$$r(L) = P(Y = \phi(x, L)) = \sum_s P(\phi(x, L) = s | Y = s) P(Y = s) \quad (2)$$

We make

$$Q(s|x) = P_L(\phi(x, L) = s) \quad (3)$$

The probability of the whole of correct classification is

$$r = \sum_s E(Q(s|X)|Y = s) P(Y = s) = \sum_s \int Q(s|x) P(s|x) P_X(dx) \quad (4)$$

where $P_X(dx)$ is all of the x distribution.

Owing to $\phi_A(x) = \text{argmax}_i Q(i|x)$

$$r_A = \sum_s \int K(\text{argmax}_s Q(i|x) = s) P(s|x) P_X(dx) \quad (5)$$

where $K(\cdot)$ is the indicator function. Denote

$$C = \{x; \text{argmax}_s P(s|x) = \text{argmax}_s Q(s|x)\} \quad (6)$$

For $x \in C$

$$\sum_s K(\text{argmax}_s Q(i|x) = s) P(s|x) = \max_s P(s|x) \quad (7)$$

we can obtain that

$$r_A = \int x \in C \max_s P(s|x) P_X(dx) + \int x \in C' K(\phi_A(x) = s) P(s|x) P_X(dx) \quad (8)$$

where C represents the order-input input set, C' is the complementary set of C' . The highest available correct classification rate is as follows:

$$Q^*(x) = \text{argmax}_s P(s|x) \quad (9)$$

and has the correct classification rate

$$r^* = \int \max_s P(s|x) P_X(dx) \quad (10)$$

C. THE PROCESS OF INTEGRATION IN THE RETRIEVAL STAGE

Next, we will introduce the integration strategy and integration process in BCMR in detail. As shown in Fig. 2, we assume that testing set contains m image-text pairs. For 50 different bootstrap replicates, we use same baseline models to train data and get 50 weak classifiers. When calculating the similarity between texts and images in testing set, 50 $m * m$ dimensional distance matrices which called Distance_1 to Distance_50, are obtained respectively.

We sort the distance matrices of every row from the nearest to the farthest, that is to say, from left to right of every row, the similarity between samples is smaller. We transform the distance into the index corresponding to the sample, and get the matrices Sort_1 to Sort_50 in Fig. 2(b). Let's suppose that we extract all the n th rows of all weak learners and arrange them into Reset_n in Fig.2(c) in sequence, where $Cm.n$ represents the n th row of m th weak classifier. For every Reset_n, we take the sample voting strategy for each column, that is, the final result is the majority of votes. If there are equal votes, select the first sample. We mark the integration results of each Reset_n as Bn . Arrange all Bn in order and get the final integration result.

IV. EXPERIMENTS

In this section, we called Bagging add a baseline method as a BCMR method. A series of experiments were carried out and eight baseline methods were conducted to check the performance of BCMR, Bagging Cross-Media Retrieval. In the experiments, there were image retrieving text (I2T) task, text retrieving image (T2I) task and their average retrieval scores on three popular benchmark datasets: Labelme dataset, Wikipedia-CNN dataset and Pascal Sentence dataset.

A. EXPERIMENTAL SETTINGS

1) DATASETS

Labelme dataset [28] is an online image dataset invented by MIT's Laboratory for cross-media retrieval. There are 2688 images and their corresponding tags. 2016/672 image-text

pairs are used for training/testing. This dataset has 8 different categories (corresponding 8 unique outdoor scenes) and each image attribute one category. 512 dimensional GIST features are used to represent images and index vector of selected tags are used to represent texts.

FIGURE 2. The process of integration in the retrieval stage

Wikipedia-CNN dataset is selected from Wikipedia features contains 2866 image text pairs, which are divided into two sets randomly, 2173 for training and 693 for testing. There are 10 semantic categories for all the samples and the 2866 pairs are classified into 10 categories. The image features are from 4096 dimensional CNN features, while the text features are from 100 dimensional Latent Dirichlet Allocation.

Pascal Sentence dataset is a data set with 20 semantic categories, where there are 1000 image-text pairs. So each category contains 50 pairs, among which there are 30 pairs for training and 20 pairs for testing randomly. For the image, 4096 dimensional CNN features are extracted as its feature representation. For text, the BoW representation is obtained based on 300 word roots, and then its probability distribution in 100 themes is calculated by Latent Dirichlet Allocation as its feature representation.

The sample number of the above three datasets is listed in Table 1.

TABLE I
SAMPLE NUMBER OF THE DATASETS

Dataset	Training set	Testing set	category
Labelme	2016 pairs	672 pairs	8
Wikipedia-CNN	2173 pairs	693 pairs	10
Pascal Sentence	600 pairs	400 pairs	20

2) BASELINE METHODS

Partial Least Square (PLS) [29] is a novel statistical data analysis model. Partial least squares regression is especially suitable when the prediction matrix has more variables than the observation. A linear regression model is found by projecting prediction variables and observation variables into a new space. In the process of modeling, PLS combines the characteristics of PCA, CCA and LRA.

Semantic Matching (SM) [30] is a method that realizes the higher level representation of texts and images. Texts and images are projected into a same semantic subspace in order that the different modalities can be measured.

MDCR [31] is the most classical task driven method. It learns their mapping matrices for each subtask (I2T and T2I), which can ensure the optimization of I2T and T2I results.

JFSSL [32] uses feature projection to solve the similarity measurement problem of different modal data, carrying out l_{21} -norm constraints to deal with paired feature selection. In order to maintain the relationship intra-modality and inter-modality of multimodal data, graph regularization is used to the data projection.

CR-CMR is a very effective method in this field. While sparse representation is used in other methods, collaborative representation is applied here. This method combines the characteristics of both dictionary learning method and semantic mapping method.

JGRMDCR [33] is also a modality-dependent method which has high accuracy. It makes full use of the one-to-one correspondence between image-text pairs, and then excavates the similarity between intra-modality and inter-modality.

PL-ranking [34] concentrates on optimizing the ranked list. In order to improve the querying accuracy, it combines pairwise ranking loss constraint, listwise constraint and regularization.

GSS-SL [35] is a semi-supervised method. Firstly, the label graph constraints are used to create semantic labels for the untagged samples, so this method is a semi-supervised one. Then labels reflecting semantic information are used to bridge different modes of the data. This method attempts to divide labeled samples and unlabeled samples by JFSSL in different proportions, and verify its effectiveness from different aspects.

3) EVALUATION METRIC AND PARAMETER TUNING

TABLE II

MAP SCORES BASED ON SM ALGORITHM ON PASCAL SENTENCE DATASET

Sampling times	I2T	T2I	Average	running time(s)
original	0.472	0.466	0.469	2.355
10	0.475	0.479	0.477	28.312
30	0.487	0.511	0.499	93.248
50	0.492	0.521	0.507	156.338
80	0.495	0.529	0.512	255.304

100 0.498 0.531 0.515 324.131

We utilize MAP and PR curves which are common evaluation criteria for cross-media retrieval tasks in the following content.

As shown in Table 2, to explore the influence of sampling times on experimental results, we use SM algorithm as a baseline algorithm to perform various retrieval tasks on Pascal Sentence dataset. The first row's value is the experimental results of the original SM algorithm. We can clearly notice that as the number of sampling increase from 10 to 100, the MAP scores for I2T, T2I and their average value also increase, of which 10 to 50 increase very rapidly, and the growth from 50 to 100 was not obvious. The running time is increased according to the multiples of sampling times. Taking the MAP scores and running time into account, we choose the sampling number of 50 in this paper.

B. ACCURACY

1) RESULTS ON THE LABELME DATASET

In table 3, we list the MAP scores obtained by eight baseline methods and their corresponding method after integration on Labelme dataset. Obviously, the MAP scores after integration is much higher than that of the original methods in various cross-media retrieval tasks. In particular, Bagging+PL-ranking achieved the most obvious improvement for I2T task, T2I task and their average retrieval scores. According to the fact that Bagging algorithm can improve the retrieval performance of the unstable learning algorithm while the performance of the stable learning algorithm is not remarkable, we can learn that GSS-SL algorithm are more stable than other algorithm, and PL-ranking algorithm is the most unstable on Labelme dataset.

Fig. 3 shows the PR curve of eight baseline method and their corresponding methods after integration on Labelme dataset. It can be clearly see that methods after integration obtain higher accuracy at the most levels of recall.

Further analysis of MAP performance on the Labelme are shown in Fig 4.(a)(c)(e). It can be found that BCMR achieves better results than baseline methods for most class. In particular, BCMR gets higher MAP scores than corresponding baseline methods on 'street' and 'highway' class for I2T task. And obtains good performances on 'street', 'coast' and 'open country' class for T2I task and average retrieval scores.

TABLE III
MAP SCORES ON LABELME DATASET

methods	labelme		
	I2T	T2I	Average
PLS	0.536	0.507	0.522
Bagging+PLS	0.552	0.538	0.545
SM	0.807	0.818	0.813
Bagging+SM	0.820	0.848	0.834
MDCR	0.632	0.671	0.652
Bagging+MDCR	0.649	0.693	0.671
JFSSL	0.734	0.712	0.723
Bagging+JFSSL	0.755	0.738	0.747
CR-CMR	0.717	0.753	0.735

Bagging+CR-CMR	0.746	0.774	0.760
JGRMDCR	0.732	0.770	0.751
Bagging+JGRMDCR	0.760	0.781	0.771
PL-ranking	0.586	0.543	0.565
Bagging+PL-ranking	0.647	0.648	0.648
GSS-SL	0.744	0.777	0.761
Bagging+GSS-SL	0.770	0.779	0.775

2) RESULTS ON THE WIKIPEDIA-CNN DATASET
TABLE IV

MAP SCORES ON WIKIPEDIA-CNN DATASET			
methods	Wikipedia		
	I2T	T2I	Average
PLS	0.223	0.261	0.242
Bagging+PLS	0.245	0.306	0.276
SM	0.438	0.368	0.403
Bagging+SM	0.445	0.423	0.434
MDCR	0.422	0.382	0.402
Bagging+MDCR	0.435	0.387	0.411
JFSSL	0.369	0.323	0.346
Bagging+JFSSL	0.398	0.362	0.380
CR-CMR	0.408	0.395	0.402
Bagging+CR-CMR	0.410	0.411	0.411
JGRMDCR	0.433	0.404	0.419
Bagging+JGRMDCR	0.448	0.410	0.429
PL-ranking	0.333	0.365	0.349
Bagging+PL-ranking	0.365	0.398	0.382
GSS-SL	0.402	0.333	0.368
Bagging+GSS-SL	0.393	0.367	0.380

Table 4 we list the MAP scores obtained by eight baseline methods and their corresponding methods after integration on Wikipedia-CNN dataset. The higher accuracy also evaluate the effectiveness of BCMR. We can see that although PLS baseline method achieves the lowest accuracy in five baseline methods. Bagging+PLS attains greatly obvious improvement for I2T task, T2I task and their average retrieval scores. About 2.2%, 4.5% and 3.4% higher than baseline method. We are surprised to find that PL-ranking algorithm is much more stable on this dataset than Labelme dataset. And Fig. 5 shows the PR curve on Wikipedia-CNN dataset.

FIGURE 3. PR curves on Labelme dataset. (a)-(h) I2I task. (i)-(p) T2I task.

FIGURE 4. The MAP scores for each class on Labelme dataset and Wikipedia-CNN dataset. Further analysis of MAP performance on the Wikipedia-CNN are expressed in Fig 4.(b)(d)(f). Although BCMR obtains a slightly undesirable precision on a particular class, it gets higher MAP scores on most of classes on Wikipedia-CNN dataset.

3) RESULTS ON THE PASCAL SENTENCE

Although the number of Pascal Sentence datasets is more than the previous two datasets, and the number of samples is much smaller than them, we can see that it has achieved good performance in Table 5. The performance of BCMR on JGRMDCR baseline method is slightly inferior. Because Bagging algorithm has its own limitations, in other words, it

is difficult for Bagging algorithm to improve accuracy for relatively stable JGRMDCR algorithm. Because bootstrap sampling greatly increases the diversity of data, as shown in Fig. 6, Fig. 7, BCMR also performs well in two indicators of PR curves and MAP scores on Pascal Sentence dataset.

FIGURE 5. PR curves on Wikipedia-CNN dataset. (a)-(h) I2I task. (i)-(p) T2I task.

TABLE V

MAP SCORES ON PASCAL SENTENCE DATASET

methods	Pascals Sentence		
	I2T	T2I	Average
PLS	0.398	0.399	0.399
Bagging+PLS	0.424	0.433	0.429
SM	0.472	0.466	0.469
Bagging+SM	0.492	0.521	0.507
MDCR	0.447	0.450	0.449
Bagging+MDCR	0.469	0.476	0.473
JFSSL	0.449	0.473	0.461
Bagging+JFSSL	0.449	0.477	0.463
CR-CMR	0.471	0.480	0.476
Bagging+CR-CMR	0.485	0.510	0.498
JGRMDCR	0.483	0.484	0.484
Bagging+JGRMDCR	0.490	0.489	0.490
PL-ranking	0.396	0.391	0.394
Bagging+PL-ranking	0.421	0.408	0.415
GSS-SL	0.468	0.464	0.466
Bagging+GSS-SL	0.470	0.503	0.487

FIGURE 6. PR curves on Pascal Sentence dataset. (a)-(h) I2I task. (i)-(p) T2I task.

FIGURE 7. The MAP scores for each class on Pascal Sentence dataset.

V. CONCLUSION

In this paper, an imaginative approach called BCMR is proposed, which is the first cross-media retrieval algorithm using integration idea in the retrieval stage. The whole algorithm is divided into two stages: (1) Sampling. Bootstrap is adopted in this paper, in fact this is a randomly sampling with replacement method.

(2) Integration of retrieval stage. The sample voting method is used to integrate 50 weak classifiers' decisions. Based on the extensive experiments performed on the above three datasets, we can conclude that BCMR can effectively improve the accuracy of cross-media retrieval.

Ethical approval

There are no animal or human experiments in this paper.

Funding details (In case of Funding)

This work of this paper was partly under the help of the Shandong Province Key R&D Program (Soft Science Project) (2021RKY03056), the Shandong Science and Technology Research Project (Grant: 2019GGX101004).

Conflict of interest

The authors declared that they have no conflict of interest to this work.

Informed Consent

None.

References

- [1] Wang, K., He, R., Wang, W., Wang, L., and Tan, T., Learning Coupled Feature Spaces for Cross-Modal Matching, IEEE International Conference on Computer Vision, IEEE Computer Society, 2088-2095(2013).
- [2] Liu, Y., Zhang, H., Liu, L., Meng, L., Wang, Y., and Dong, X., Cross-media retrieval based on query modality and semi-supervised regularization, Journal of Advanced Computational Intelligence & Intelligent Informatics, 21(7), 1211-1220(2017).
- [3] Xu, G., Zhai, A., Wang, J., Zhang, Z., Li, X., Cross-Media Semantic Matching based on Sparse Representation. Technical Gazette, 26(6), 1707-1713(2019).
- [4] Sun, J., Liu, X., Wan, W., Li, J., Zhao, D., and Zhang, H., Video hashing based on appearance and attention features fusion via dbn, Neurocomputing, 213, 84-94(2016).
- [5] Vinyals, O., Toshev, A., Bengio, S., and Erhan, D., Show and tell: a neural image caption generator, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3156- 3164(2015).
- [6] Srivastava, N., and Salakhutdinov, R., Multimodal learning with Deep Boltzmann Machines, International Conference on Neural Information Processing Systems, 2222- 2230(2012).
- [7] Zhang, Y., and Oliver, D. S., Improving the ensemble estimate of the kalman gain by bootstrap sampling, Mathematical Geosciences, 42(3), 327-345(2010).
- [8] Xie, L., Zhu, L., Pan, P., and Lu, Y. Cross-modal self-taught hashing for large-scale image retrieval, Signal Processing, 124(C), 81-92 (2016).
- [9] Xu, G., Li, X., Lin, S., Zhang, Z., Zhai A., Combination subspace graph learning for cross-modal retrieval. Alexandria Engineering Journal. 59(3), 1333-1343(2020).
- [10] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., and Ng, A. Y., Multimodal Deep Learning, International Conference on Machine Learning, ICML, 689-696(2011).
- [11] Zhu, F., Shao, L., and Yu, M., Cross-modality submodular dictionary learning for information retrieval, ACM International Conference on Conference on Information and Knowledge Management (CIKM), 1479-1488(2014).
- [12] Li, X., Xu, G., Cao, Q., Zou, W., Xu, Y., Cong, P., Identification of Glioma Pseudoprogession Based on Gabor Dictionary and Sparse Representation Model, NeuroQuantology, 16(1), 43-51(2018).
- [13] Xu, G., Jia, G., Shi, L., Zhang, Z., Personalized Course Recommendation System Fusing with Knowledge Graph and Collaborative Filtering, Computational Intelligence and Neuroscience 8 (2021).
- [14] Xu, G., Li, X., Zhang, Z., Semantic Consistency Cross-Modal Retrieval With Semi-Supervised Graph Regularization, IEEE Access, 8, 14278-14288(2020)..
- [15] Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H., Generative adversarial text to image synthesis, International Conference on International Conference on Machine Learning, 1060-1069(2016).
- [16] Andrew, G., Arora, R., Bilmes, J., and Livescu, K., Deep canonical correlation analysis, International Conference on International Conference on Machine Learning, 3408C3415(2010).
- [17] Peng, Y., Huang, X., and Qi, J., Cross-media shared representation by hierarchical learning with multiple deep networks, International Joint Conference on Artificial Intelligence (IJCAI), 3846C3853(2016).
- [18] Shang, F., Zhang, H., Sun, J., Liu, L., and Zeng, H., Journal of Advanced Computational Intelligence and Intelligent Informatics (JACIII), 22(2), 280-289(2018).
- [19] Zhuang, Y., Wang, Y., Wu, F., Zhang, Y., and Lu, W., Supervised coupled dictionary learning with group structures for multi-modal retrieval, AAAI Conference on Artificial Intelligence, 1070C1076(2013).
- [20] Bahrampour, S., Nasrabadi, N. M., Ray, A., and Jenkins, W. K., Multimodal task-driven dictionary learning for image classification, IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 25(1), 24-38(2015).
- [21] Zhai, X., Peng, Y., and Xiao, J., Learning cross-media joint representation with sparse and semisupervised regularization, IEEE Transactions on Circuits & Systems for Video Technology, 24(6), 965-978(2014).
- [22] Peng, Y., Zhai, X., Zhao, Y., and Huang, X., Semi-supervised cross-media feature learning with unified patch graph regularization, IEEE Transactions on Circuits & Systems for Video Technology, 26(3), 583-596(2016).
- [23] Tong, H., He, J., Li, M., Zhang, C., and Ma, W., Graph based multimodality learning, ACM international conference on Multimedia (ACM MM), 862C871(2005).
- [24] Efron, B., Bootstrap methods: another look at the jackknife, Annals of Statistics, 7(1), 1-26(1979).
- [25] Parke, J., Holford, N. H., and Charles, B. G., A procedure for generating bootstrap samples for the validation of nonlinear mixed-effects population models, Comput. methods Progr. biomed, 59(1), 19-29(1999).
- [26] Breiman, L., Bagging predictors, Machine Learning, 24(2), 123-140(1996).
- [27] Liang, G., Zhu, X., and Zhang, C., An Empirical Study of Bagging Predictors for Different Learning Algorithms, AAAI Conference on Artificial Intelligence, 2, 627-49(2011).
- [28] Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T., Labelme: a database and web-based tool for image annotation, International Journal of Computer Vision, 77(1- 3), 157-173(2008).
- [29] Rosipal, R., and Kramer, N. Overview and recent advances in partial least squares. Subspace, Latent Structure and Feature Selection Techniques, 3940, 34-51(2006).
- [30] Rasiwasia, N., Pereira, J. C., Coviello, E., Doyle, G., Lanckriet, G. R. G., and Levy, R., A new approach to cross-modal multimedia retrieval, International Conference on Multimedia, 251-260(2010).
- [31] Wei, Y., Zhao, Y., Zhu, Z., Wei, S., Xiao, Y., and Feng, J., Modality-dependent cross-media retrieval, Acm Transactions on Intelligent Systems & Technology, 7(4), 57(2015).
- [32] Wang, K., He, R., Wang, L., Wang, W., and Tan, T., Joint feature selection and subspace learning for cross-modal retrieval. IEEE Transactions on Pattern Analysis & Machine Intelligence, 38(10), 2010(2016).
- [33] Yan, J., Zhang, H., Sun, J., Wang, Q., Guo, P., and Meng, L., Joint graph regularization based modality-dependent cross-media retrieval, Multimedia Tools & Applications(6), 1- 19(2017).
- [34] Zhang, L., Ma, B., Li, G., Huang, Q., and Tian, Q., PL-ranking: A Novel Ranking Method for Cross-Modal Retrieval. ACM on Multimedia Conference, 1355-1364(2016).
- [35] Zhang, L., Ma, B., Li, G., Huang, Q., and Tian, Q., Generalized semi-supervised and structured subspace learning for cross-modal retrieval. IEEE Transactions on Multimedia, 99, 1-1(2017).