International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT1952275

Content-Based Image Retrieval : A Comprehensive Study

Abhishek Swaroop[1], Aman[2], Amit Rawat[3], Ashwin Giri[4], Hardik Gothwal[5]
[1]Head of Department, Department of Information Technology, Bhagwan Parshuram Institute of Technology, New Delhi, Delhi, India
[2, 3,4,5]Student, Department of Information Technology & Engineering, GGSIPU, Bhagwan Parshuram Institute of Technology, Affiliated to GGSIPU, New Delhi, India

ABSTRACT

Learning efficient options illustrations and equivalency metric measures are imperative to the searching performance of a content-based image retrieval (CBIR) machine. Despite in depth analysis efforts for many years, it remains one amongst the foremost difficult open issues that significantly hinders the success of real-world CBIR systems. The key issue has been associated to the commonly known "linguistic gap" problem that exists between low-level image pixels captured by machines and high-level linguistics ideas perceived by humans. Among varied techniques, machine learning has been actively investigated as a potential direction to bridge the linguistics gap in the long run. Motivated by recent success of deep learning techniques for computer vision and other applications, In this paper, we'll conceive to address an open problem: if deep learning could be a hope for bridging the linguistics gap in CBIR and the way a lot of enhancements in CBIR tasks may be achieved by exploring the progressive deep learning methodologies for learning options illustrations and equivalency measures. Specifically, we'll investigate a framework of deep learning with application to CBIR tasks with an extensive set of empirical studies by examining a progressive deep learning technique (Convolutional Neural Networks) for CBIR tasks in varied settings. From our empirical studies, we found some encouraging results and summarized some vital insights for future analysis. CBIR tasks may be achieved by exploring the progressive deep learning techniques for learning options illustrations and equivalency measures.

Keywords : Content-Based Image Retrieval, Convolutional Neural Networks, Feature Illustration

I. INTRODUCTION

The search performance of a content-based image retrieval machine is crucially dependent on the options illustrations and equivalency measure.

Recent years have witnessed some vital advances of new techniques in machine learning. One vital breakthrough technique is understood as "deep learning", which incorporates a family of machine learning algorithms that conceive to model high-level abstractions in knowledge by using deep architectures composed of multiple non-linear transformations [5, 11]. In contrast to standard machine learning strategies that are usually employing "shallow" architectures, deep learning mimics the human brain that's organized in a very deep design and processes data through multiple stages of transformation and illustration. By exploring deep architectures to learn options at multiple level of abstracts from knowledge in an automated manner, deep learning strategies permit a system to learn complicated functions that directly map raw sensory input file to the output, while not counting on human-crafted options employing domain information. Several recent studies have reported encouraging results for applying deep learning techniques to a range of applications, in the domain

of speech recognition [16, 55], visual perception [26, 56], and natural language process [19, 34], among others.

Inspired by the success of deep learning, during this paper, we conceived to explore deep learning methodologies with application to CBIR assignments. Despite abundant analysis attention of applying deep learning for image classification and recognition in computer vision, there's still restricted quantity of attention specializing in the CBIR applications. In this research paper, we'll investigate deep learning strategies for learning feature representations from pictures and their similarity measures towards CBIR tasks.

II. RELATED WORK

Our analysis lies within the interaction of deep neural networks learning, content-based image retrieval and distance metrics learning. We'll briefly review every group of connected work below.

2.1 Content-Based Image Retrieval

Content-based image retrieval (CBIR) is one amongst the elemental analysis challenges extensively studied in multimedia system community for many years [30, 25, 45]. CBIR aims to go looking for pictures through analyzing their visual contents, and so image illustration is the very crux of CBIR. Over the past decades, a range of lowlevel feature descriptors have been proposed for image illustration [21], starting from global options, like color options [21], edge options [21], texture options [32], GIST [36, 37], and CENTRIST [49], and recent local feature representations, like the bag-of-words (BoW) models [44, 54, 50, 51] employing native feature descriptors (eg. Speeded-up robust features etc.). Standard CBIR approaches typically select rigid distance functions on some extracted low-level options for multimedia system similarity search, like Euclidean distance or cos similarity. However, the fixed rigid similarity/distance operation might not be continually optimum to the advanced visual image retrieval tasks because of the grand challenge of the linguistics gap between low-level visual options extracted by computers and high-level human perceptions.

Hence, recent years have witnessed a surge of active analysis efforts in the designing of varied distance/similarity measures on some low-level options by exploring machine learning techniques [35, 7, 6]. Among these techniques, some works have been centered on learning to hashing or compact codes [41, 35, 23, 57, 58]. For instance, Norouzi et al [35] proposed a mapping learning methodology for large scale multimedia system applications from high-dimensional knowledge to binary codes that preserve linguistics similarity. Jegou et al [23] adopted the fisher kernel to combine native descriptors and adopted a joint dimension reduction so as to scale back a picture to some dozen bytes while maintaining high accuracy. A different way to reinforce the feature illustration is distance metric learning (DML), as discussed thoroughly as follows.

2.2 Distance metric Learning

Distance metrics learning for image searching is intensively studied in each of multimedia system retrieval communities and machine learning [12, 2, 48, 29, 15, 47, 33, 46]. In the following, we'll briefly discuss various groups of existing work for distance metrics learning categorized by different customized training principles and settings. In terms standard training knowledge formats, most existing DML studies usually work with two sorts of knowledge (a.k.a. aspect information): pairwise constraints where must-link constraints and cannot-link constraints are given and triplet constraints that contains an identical pair and a dissimilar pair.

There have been studies that directly use the category labels for DML by following a typical machine learning theme, like the Large Margin Nearest Neighbor (LMNN) algorithmic program [48], that however isn't basically completely different. In terms of various learning approaches, distance metric learning techniques are generally classified into two groups: the global supervised approaches [2, 18] that learn a metric on a global setting by satisfying all the constraints at the same time, the native supervised approaches [48, 12] that learn a metric on the native sense by only satisfying the given native constraints from neighboring data. In terms of learning methodology, most existing DML studies usually use batch learning strategies which frequently assume the entire array of standard training knowledge should be provided before the learning task and train a model from scratch. In contrast to the batch learning strategies, so as to handle large-scale knowledge, online Decentralized Machine Learning algorithms are being recently studied. The key concept of distance metric learning is to learn an optimum metric that minimizes the gap between similar pictures and at the same time maximizes the gap between dissimilar pictures. In this analysis condition, another technique named similarity learning is closely associated with distance metric learning. As an example, Chechik et al. proposed an Online Algorithmic rule for scalable image similarity (OASIS) [7] for bettering image retrieval performance.

III. IMAGE DATASETS

Our empirical studies aim to guage the performance of the three feature generalization schemes based on completely different image datasets, including the overall image information "ImageNet", the item image information "Caltech256", the landmark image datasets "Oxford" and "Paris", and also the facial image dataset "Pubfig83LFW". We'll briefly introduce each of them as follows.

ImageNet: A large-scale dataset with over fifteen million labelled high-resolution pictures belonging to roughly 22,000 classes. the pictures were collected from the internet and labelled by human labelers by the help of Amazon's Mechanical Turk crowd-sourcing tool. Beginning in 2010, as a part of the Pascal Visual Object Challenge, an annual competition known as the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) has been held. ILSVRC uses a smaller set of ImageNet with roughly 1,000 pictures in each of 1,000 classes. In all, there are roughly 1.2 million standard training pictures, 50,000 validation pictures, and 150,000 testing pictures. ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012) set is the basis of training of the Deep Convolutional Neural Network (CNN) model in our framework.

Caltech256: contains 30,607 pictures of objects, that were obtained from Google image search and from PicSearch.com. Pictures were allotted to 257 classes and evaluated by humans so as to confirm image quality and relevancy.

Oxford: contains 5,063 high resolution pictures that were automatically downloaded from Flickr. It defines fifty five queries used for analysis, that consists of five for each of the eleven chosen Oxford landmarks. It's quite difficult because of substantial variations in scale, viewpoint and lighting conditions.

Paris: is analogously to the previous "Oxford" datasets. Its 6,392 pictures were obtained from Flickr, and there also are fifty five queries for analysis. Since it contains pictures of Paris it's thought of to be an autonomous dataset from "Oxford".

Pubfig83LFW: is an open-universe online facial image dataset [4], which mixes 2 commonly used face databases: PubFig83 [27] and LFW [20]. The eighty

three people from PubFig represent the demo pictures and standard training gallery, and every one of the remaining people from LFW represent the distractor gallery or background faces. All the faces from each individual in PubFig83 are at random divided into two-third standard training faces and one third demo faces. All overlapping people from LFW are manually removed, and also the left LFW dataset is employed as distractors to PubFig83. All the facial pictures are resized to 250×250, and only the facial pictures that could be detected by a series enterprise software system remained. In summary, the PubFig83+LFW dataset has 83 people with 8,720 faces for standard learning and 4,282 faces for testing and over 5,000 people from LFW with 12,066 faces for background and distractor faces.

Table 2: Image Retrieval Performance on Caltech256

		Euclidean				Various Metric Learning on BoW*				OASIS on Deep Feature		
	Features	BoW*	DF.FC1	DF.FC2	DF.FC3	OASIS	MCML	LEGO	LMNN	DF.FC1	DF.FC2	DF.FC3
10 classes	mAP	0.2300	0.6424	0.7695	0.7109	0.3300	0.2900	0.2700	0.2400	0.8617	0.9153	0.8512
	P@K=1	0.3700	0.8800	0.9200	0.8800	0.4300	0.3900	0.3900	0.3800	0.9360	0.9560	0.9440
	P@K=10	0.2700	0.7748	0.8624	0.7928	0.3800	0.3300	0.3200	0.2900	0.9084	0.9400	0.8944
	P@K=50	0.1800	0.3614	0.4188	0.3930	0.2300	0.2200	0.2000	0.1800	0.4457	0.4614	0.4498
20 classes	mAP	0.1400	0.4984	0.5609	0.5493	0.2100	0.1700	0.1600	0.1400	0.6962	0.7388	0.6399
	P@K=1	0.2500	0.7840	0.7960	0.8020	0.2900	0.2600	0.2600	0.2600	0.8320	0.8780	0.7860
	P@K=10	0.1800	0.6228	0.6772	0.6588	0.2400	0.2100	0.2000	0.1900	0.7768	0.8130	0.7186
	P@K=50	0.1200	0.2998	0.3315	0.3290	0.1500	0.1400	0.1300	0.1100	0.3905	0.4084	0.3764
50 classes	mAP	0.0900	0.4011	0.4624	0.4286	0.1200	-	0.0900	0.0800	0.5186	0.5410	0.4603
	P@K=1	0.1700	0.6792	0.7240	0.6912	0.2100	-	0.1800	0.1800	0.7288	0.7320	0.6840
	P@K=10	0.1300	0.5333	0.5913	0.5519	0.1600	-	0.1300	0.1200	0.6194	0.6394	0.5600
	P@K=50	0.0800	0.2509	0.2836	0.2700	0.1000	-	0.0800	0.0700	0.3155	0.3308	0.2980

The results marked with ⋆ are taken directly from the study in [7].

Table 1: Image Retrieval Performance on ImageNet

Feature	mAP	P@K=1	P@K=10	P@K=50	P@K=100	R@K=1	R@K=10	R@K=50	R@K=100
BoW.1200*	0.0007	0.0472	0.0234	0.0144	0.0118	0.0000	0.0002	0.0006	0.0009
BoW.4800*	0.0008	0.0487	0.0243	0.0148	0.0121	0.0000	0.0002	0.0006	0.0009
BoW.1K	0.0012	0.0243	0.0170	0.0124	0.0108	0.0000	0.0001	0.0005	0.0008
BoW.10K	0.0013	0.0212	0.0142	0.0106	0.0094	0.0000	0.0001	0.0004	0.0007
BoW.100K	0.0015	0.0216	0.0140	0.0110	0.0099	0.0000	0.0001	0.0004	0.0008
BoW.1M	0.0016	0.0286	0.0174	0.0132	0.0117	0.0000	0.0001	0.0005	0.0009
GIST	0.0002	0.0137	0.0080	0.0056	0.0049	0.0000	0.0001	0.0002	0.0004
DF.FC1	0.0748	0.4427	0.3650	0.2967	0.2637	0.0003	0.0028	0.0116	0.0205
DF.FC2	0.1218	0.4711	0.4105	0.3547	0.3254	0.0004	0.0032	0.0138	0.0254
DF.FC3	0.1007	0.4282	0.3726	0.3183	0.2898	0.0003	0.0029	0.0124	0.0226

* BoW.1200 and BoW.4800 are extracted by other groups, which are publicly available [3].

IV. EXPERIMENTS

In this section, we'll design an extensive set of experiments to guage the performance of deep learning techniques for CBIR tasks. Specifically, the first experiment is to look at how the deep CNN model performs for CBIR tasks on an equivalent dataset that was employed to train the model, and then the rest experiments aim to check the generalization of the pre-trained deep model to CBIR tasks on different new domains, which might be a lot different from the standard training knowledge used for coaching the initial CNN models. For performance analysis metrics, we've used use 3 conventional analysis measures commonly employed in CBIR tasks, namely the mean average preciseness (mAP), the preciseness at specific ranks ("P@K"), and the recall at specific ranks ("R@K").

4.1 Experiment on ImageNet

In this experiment, we'll aim to guage the CBIR performance by the use of methodology I. We'll assess the retrieval performance on the ILSRVC 2012 dataset. We'll use the 50,000 validation pictures as question set, and search on the 1.2-million standard training image set. We'll then compare methodology I with various bag-of-words (BoW) feature representations, that are commonly used for large-scale image retrieval. Among these BoW options, "BoW.1200" and "BoW.4800" are extracted by other groups. Table 1 shows the observations and results of our experiment. The Figure shows the results of various image retrieval queries on the ImageNet dataset.

Several observations are achieved from the results. Firstly, we could observe that this is an awfully difficult CBIR task. The most effective BoW feature illustration based on a codebook with the vocabulary size 1,000,000 could only accomplish the mAP of 0.0016, and also the performance of global GIST feature is even worse. The results of "BoW.1200" and "BoW.4800" are generated based on the options freed in other works, that is analogous to the other BoW representations generated by ourselves. Secondly, the activations based feature vector from the entirely connected layer FC1/FC2/FC3 accomplish significantly far better results, among which the "DF.FC2"(the last hidden layer) achieved the most effective performance with top-1 preciseness of 47.11%. Though the last output layer DF.FC3 is the classification output of the Image Web trained CNN model, that contains the most effective sematic data, it is apparantly not a decent feature illustration for CBIR tasks. By examining the typical assessment measures: preciseness and recall, we are able to find identical observations. The preciseness at specific ranks (P@K) = 1 of the last hidden layer "DF.FC2" is around 0.471 for K = 1. It suggests that the error rate of the closest neighbor classification with K = 1 is 0.53, that is incredibly near to the classification error rate of the ImageNet trained model (0.425). Finally, we note that our current experiments didn't add additional post processing step to enhance the CBIR performance, though some techniques (such as "geometric constraint based re ranking" [38] or "query expansion" [53]) typically could additionally boost the image retrieval performance, but that however is out of the scope of this study.

4.2 Experiments on Various CBIR Tasks

In this section, we aim to guage the performance of feature illustration schemes in Figure 1(b) on new diversified CBIR tasks. Specifically, we examine the performance of the models for feature representations on 3 completely different CBIR tasks: (i) object retrieval tasks employing the "Caltech256" dataset, that is an object-based image dataset, and different categories are quite distinct; (ii) landmark retrieval tasks employing the "Oxford" and "Paris" datasets, that contains landmark photos where all the photographs are captured in varied conditions (scale, viewpoint and lighting conditions); and (iii) facial image retrieval tasks employing the "Pubfig83LFW" dataset, that is difficult as intra category distinction sometimes might be even larger than inter category distinction.

V. CONCLUSION

Inspired by recent success of deep learning techniques, during this paper, we conceived to address the long- standing elementary feature illustration drawback in Content-based Picture Retrieving. We aim to guage if deep learning could be a hope for bridging the linguistics gap in CBIR for the long run, and the way a lot of empirical enhancements in CBIR tasks could be achieved by exploring the progressive deep learning techniques for learning equivalency metrics measures and options illustrations. Particularly, we investigated a

deep learning framework with applications in CBIR machine tasks with an extensive set of empirical studies by examining a progressive deep learning methodology (Convolutional Neural Networks) for CBIR tasks in varied settings. The results of the studies suggests that:

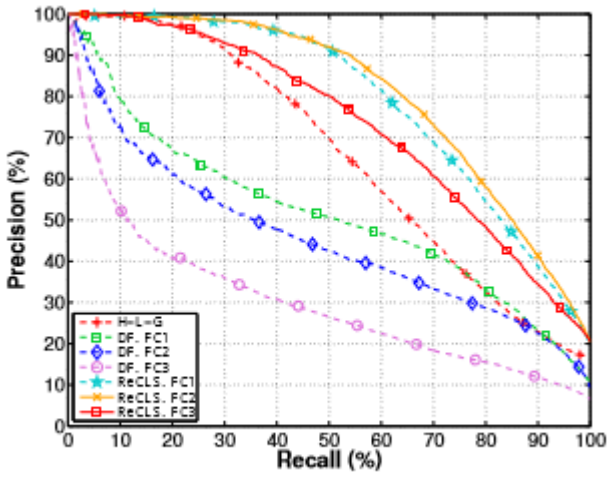

Figure 1 : Precision-recall result on the Pubfig83LFW dataset.

(i) deep Convolutional Neural Networks model pre-trained on a large dataset can be directly used for options extraction in new CBIR tasks, the promising results on Caltech256 dataset demonstrate that the pre trained model are capable to capture high linguistics data within the raw pixels; (ii) The options extracted by pre-trained CNN model might or might not be better than the standard hand- crafted options, however with the right set of feature refining schemes, the deep learning feature illustrations consistently outdo traditional hand-crafted options on all datasets; (iii) When being applied for feature representation in a brand new domain, we've found results that suggest that similarity learning could additionally boost the search performance of direct options output of pre-trained deep learning models; and (iv) Finally, by retraining the deep models with classification or similarity learning objective on the new domain, we've found that the retrieval performance can be boosted significantly which is far better than the enhancements generated by "shallow" similarity learning. Despite encouraging results achieved, we believe this can be simply a starting for deep learning with application to CBIR tasks, and there are still several open challenges. In future work, we are going to investigate additional advanced deep learning techniques and assess additional different and diversified datasets for more comprehensive empirical studies thus as to provide additional insights for bringing the linguistics gap of multimedia system data retrieval in the future.

VI. REFERENCES

[1]. D. H. Ackley, G. E. Hinton, and T. J. Sejnowski. A learning algorithm for boltzmann machines*. Cognitive science,9(1):147–169,1985.

[2]. A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall. Learning distance functions using equivalence relations. In ICML,pages11–18,2003.

[3]. H. Bay, T. Tuytelaars, and L. J. V. Gool. Surf: Speeded up robust features. In ECCV (1), pages 404–417, 2006. 4B. C. Becker and E. G. Ortiz. Evaluating open-universe face identification on the web. In CVPR Workshops, pages 904–911,2013 . 5]Y.Bengio, A.C.Courville, and P. Vincent. Unsupervised feature learning and deep learning: A review and new perspectives.CoRR,abs/1206.5538,2012.

[4]. H. Chang and D.-Y. Yeung. Kernel-based distance metric learning for content-based image retrieval. Image andVisionComputing,25(5):695–703,2007.

[5]. G. Chechik, V. Sharma, U. Shalit, and S. Bengio. Large scale online learning of image similarity through ranking. Journal of Machine Learning Research, 11:1109–1135, 2010.

[6]. D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In NIPS, pages 2852–2860,2012.

[7]. K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. Online passive-aggressive algorithms. Journal of Machine Learning Research, 7:551–585, 2006.

[8]. J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, Q. V. Le, M. Z. Mao, M. Ranzato, A. W. Senior, P. A. Tucker, K. Yang, and A. Y. Ng. Large scale distributed deep networks. In NIPS, pages 1232–1240, 2012. 11L. Deng. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing, 3:e2, 2014. 12C. Domeniconi, J. Peng, and D. Gunopulos. Locally adaptive metric nearest-neighbor classification. IEEE Trans. Pattern Anal. Mach. Intell., 24(9):1281–1285, 2002. 13J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. CoRR, abs/1310.1531,2013.

[9]. R. B. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. CoRR, abs/1311.2524, 2013. 15M. Guillaumin, J. J. Verbeek, and C. Schmid. Is that you? metric learning approaches for face identification. In ICCV,pages498–505,2009.

[10]. G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Processing Magazine, IEEE, 29(6):82–97,2012.

[11]. G. E. Hinton, S. Osindero, and Y. W. Teh. A fast learning algorithm for deep belief nets. Neural Computation,18(7):1527–1554,2006.

[12]. S. C. H. Hoi, W. Liu, M. R. Lyu, and W.-Y. Ma. Learning distance metrics with contextual constraints for image retrieval. In CVPR (2), pages 2072–2078, 2006. 19E. H. Huang, R. Socher, C. D. Manning, and A. Y. Ng. Improving word representations via global context and multiple word prototypes. In ACL (1), pages 873–882, 2012.

[13]. G. B. Huang, M. Ramesh, T. Berg, and E. Learned- Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[14]. A. K. Jain and A. Vailaya. Image retrieval using color and shape. Pattern Recognition, 29(8):1233–1244, 1996. 22 P. Jain, B. Kulis, I. S. Dhillon, and K. Grauman. Online metric learning and fast similarity search. In NIPS, pages 761–768, 2008.

[15]. H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid. Aggregating local image descriptors into compact codes. IEEE Trans. Pattern Anal. Mach. Intell., 34(9):1704–1716, 2012.

[16]. R. Jin, S. Wang, and Y. Zhou. Regularized distance metric learning: Theory and algorithm. In NIPS, pages 862–870, 2009.

[17]. Y. Jing and S. Baluja. Visualrank: Applying pagerank to large-scale image search. IEEE Trans. Pattern Anal. Mach. Intell., 30(11):1877–1890, 2008.

[18]. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In NIPS, pages 1106–1114, 2012.

[19]. N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In ICCV, pages 365–372, 2009.

[20]. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.

[21]. J.-E. Lee, R. Jin, and A. K. Jain. Rank-based distance metric learning: An application to image retrieval. In CVPR, 2008.

[22]. M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content- based multimedia information retrieval: State of the art and challenges. TOMCCAP, 2(1):1–19, 2006.

[23]. D. G. Lowe. Object recognition from local scale- invariant features. In ICCV, pages 1150–1157, 1999.

[24]. B. S. Manjunath and W.-Y. Ma. Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell., 18(8):837–842, 1996.

[25]. A. S. Mian, Y. Hu, R. Hartley, and R. A. Owens. Image set based face recognition using self-regularized non-negative coding and adaptive distance metric learning. IEEE Transactions on Image Processing, 22(12):5252– 5262, 2013.

[26]. T. Mikolov, W. tau Yih, and G. Zweig. Linguistic regularities in continuous space word representations. In HLT-NAACL, pages 746–751, 2013.

[27]. M. Norouzi, D. J. Fleet, and R. Salakhutdinov. Hamming distance metric learning. In NIPS, pages 1070– 1078, 2012.

[28]. A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision, 42(3):145–175, 2001.

[29]. A. Oliva and A. Torralba. Scene-centered description from spatial envelope properties. In Biologically Motivated Computer Vision, pages 263–272, 2002.

[30]. J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In CVPR, 2007.

[31]. A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. CoRR, abs/1403.6382, 2014.

[32]. R. Salakhutdinov and G. E. Hinton. Deep boltzmann machines. In AISTATS, pages 448–455, 2009.

[33]. R. Salakhutdinov and G. E. Hinton. Semantic hashing. Int. J. Approx. Reasoning, 50(7):969–978, 2009.

[34]. R. Salakhutdinov, A. Mnih, and G. E. Hinton. Restricted boltzmann machines for collaborative filtering. In ICML, pages 791–798, 2007.

[35]. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. CoRR, abs/1312.6229, 2013.

[36]. J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman. Discovering objects and their localization in images. In ICCV, pages 370–377, 2005.

[37]. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell., 22(12):1349–1380, 2000.

[38]. D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. He, and C. Miao. Learning to name faces: a multimodal learning scheme for search-based face annotation. In SIGIR, pages 443–452, 2013.

[39]. Z. Wang, Y. Hu, and L.-T. Chia. Learning image-to-class distance metric for image classification. ACM TIST,4(2):34, 2013.

[40]. K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In NIPS, 2005.

[41]. J. Wu and J. M. Rehg. Centrist: A visual descriptor for scene categorization. IEEE Trans. Pattern Anal. Mach. Intell., 33(8):1489–1501, 2011.

[42]. L. Wu and S. C. H. Hoi. Enhancing bag-of-words models with semantics-preserving metric learning. IEEE MultiMedia, 18(1):24–37, 2011.

[43]. L. Wu, S. C. H. Hoi, and N. Yu. Semantics-preserving bag-of-words models and applications. IEEE Transactions on Image Processing, 19(7):1908–1920, 2010.

[44]. P. Wu, S. C. H. Hoi, H. Xia, P. Zhao, D. Wang, and C. Miao. Online multimodal deep similarity learning with application to image retrieval. In ACM Multimedia, pages 153–162, 2013.

[45]. H. Xie, Y. Zhang, J. Tan, L. Guo, and J. Li. Contextual query expansion for image retrieval. IEEE Transactions on Multimedia, 16(4):1104–1114, 2014.

[46]. J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo. Evaluating bag-of-visual-words representations in scene classification. In Multimedia Information Retrieval, pages 197–206, 2007.

[47]. D. Yu, M. L. Seltzer, J. Li, J.-T. Huang, and F. Seide. Feature learning in deep neural networks - a study on speech recognition tasks. CoRR, abs/1301.3605, 2013.

[48]. M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. CoRR, abs/1311.2901, 2013.

[49]. L. Zhang, Y. Zhang, X. Gu, J. Tang, and Q. Tian. Scalable similarity search with topology preserving hashing. IEEE Transactions on Image Processing, 23(7):3025–3039, 2014.

[50]. Y. Zhang, L. Zhang, and Q. Tian. A prior-free weighting scheme for binary code ranking. IEEE Transactions on Multimedia, 16(4):1127–1139, 2014.

Cite this article as :

Abhishek Swaroop, Aman, Amit Rawat, Ashwin Giri, Hardik Gothwal, "Content-Based Image Retrieval : A Comprehensive Study", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5, Issue 2, pp.1073-1081, March-April-2019. Available at doi : https://doi.org/10.32628/CSEIT1952275
Journal URL : http://ijsrcseit.com/CSEIT1952275