OPEN ACCESS

Global Journal of Research in Engineering & Computer Sciences

Volume 01| Issue 02 | Nov - Dec | 2021

Journal homepage: https://gjrpublication.com/journals/

Original Research Article

Content Based Image Retrieval Using Deep Learning Techniques Review

*Showkat A Dar[1], Selvarani M[1]&M Arulselvi[1]

[1]Department of Computer Science and Engineering, Annamalai University

Submission Date: 31 Oct. 2021 |**Published Date:** 20 Dec. 2021

***Corresponding author:** Showkat Ahmad Dar

Abstract

A content-based image retrieval (CBIR) system works on the low-level visual features of a user input query image, which makes it difficult for the users to formulate the query and also does not give satisfactory retrieval results. In the past image annotation was proposed as the best possible system for CBIR which works on the principle of automatically assigning keywords to images that help image retrieval users to query images based on these keywords. Image annotation is often regarded as the problem of image classification where images are represented by some low-level features and the mapping between low-level features and high-level concepts (class labels) is done by supervised learning algorithms. In a CBIR system learning of effective feature representations and similarity measures is very important for the retrieval performance. Semantic gap has been the key challenge for this problem. A semantic gap exists between low-level image pixels captured by machines and the high-level semantics perceived by humans. The recent successes of deep learning techniques especially Convolutional Neural Networks (CNN) has been reviewed with CBIR.

Keywords: Content Based Image Retrieval (CBIR), visual representation, indexing, similarity measurement, spatial context, search re-ranking, intention gap and semantic gap.

INTRODUCTION

In the last three years, object classification and detection capabilities have dramatically improved due to advances in deep learning and convolutional networks. One encouraging news is that most of this progress is not just the result of more powerful hardware, larger datasets and bigger models, but mainly a consequence of new ideas, algorithms and improved network architectures. No new data sources were used, for example, by the top entries in the ILSVRC 2014 competition besides the classification dataset of the same competition for detection purposes (Zhou et al., 2011). Our Google Net submission to ILSVRC 2014 actually uses 12 times fewer parameters than the winning architecture of while being significantly more accurate. On the object detection, the biggest gains have not come from naive application of bigger and bigger deep networks, but from the synergy of deep architectures and classical computer vision, like CNN Algorithm.

Another notable factor is that with the ongoing traction of mobile and embedded computing, the efficiency of our algorithms especially their power and memory use gains importance. It is noteworthy that the considerations leading to the design of the deep architecture presented in this paper included this factor rather than having a sheer fixation on accuracy numbers. For most of the experiments, the models were designed to keep a computational budget of 1.5 billion multiply adds at inference time, so that the they do not end up to be a purely academic curiosity, but could be put to real world use, even on large datasets, at a reasonable cost. In this paper, an efficient deep neural network architecture will be focused for computer vision, code named Inception, which derives its name from the Network in network paper in conjunction with the famous “we need to go deeper” internet. In general, one can view the Inception model as a logical culmination of while taking inspiration and guidance from the theoretical work (Chum et al., n.d.). The benefits of the architecture are experimentally verified on the ILSVRC 2014 classification and detection challenges, where it significantly outperforms the current state of the art.

Content-Based Image Retrieval

Content-based image retrieval (CBIR) for decades has been one of the most researched fields of computer vision. CBIR aims to search for images through analyzing their visual contents, and thus image representation is the crux of CBIR. In the past there has been a variety of proposed low-level feature descriptors for image representation, ranging from global features like color features, edge features, texture features, GIST and CENTRIST, and recent local feature representations, such as the bag-of-words (Bow) models using local feature descriptors (SIFT, SURF). Conventional CBIR approaches usually choose rigid distance functions on some extracted low-level features for multimedia similarity search, such as Euclidean distance or cosine similarity. Therefore, in the recent past there have been a surge of active research efforts in the design of various distance/similarity measures on some low-level features by exploring machine learning techniques. Among these techniques, some works have focused on learning to hashing or compact codes(Philbin et al., 2008) .

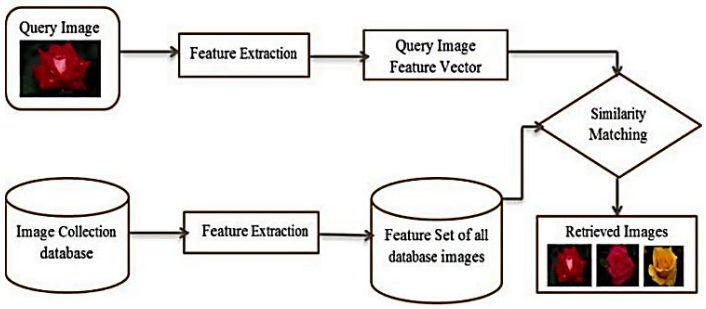

Fig-1: Block diagram of CBIR system

Deep Learning

Deep learning refers to a class of machine learning techniques, where many layers of information processing stages in hierarchical architectures are exploited for pattern classification and for feature or representation learning (Chum et al., 2011). It lies in the intersections of several research areas, including neural networks, graphical modeling, optimization, pattern recognition, and signal processing, etc. can be adopted the deep supervised back propagation convolutional network for digit recognition. In the recent past, it has become a valuable research topic in the fields of both computer vision and machine learning where deep learning achieves state-of-the art results for a variety of tasks. The deep convolutional neural networks (CNNs) came out first in the image classification task of Image Net classification with deep convolutional neural networks. The model was trained on more than one million images, and has achieved a winning top-5 test error rate of 15.3% over 1, 000 classes. (Jegou et al., 2008).After that, some recent works got better results by improving CNN models. The top-5 test error rate decreased to 13.24% in by training the model to simultaneously classify, locate and detect objects. Besides image classification, the object detection task can also benefit from the CNN model, as reported in. Over the past several years, a rich family of deep learning techniques has been proposed and extensively studied, e.g., Deep Belief Network (DBN), Boltzmann Machines (BM), Restricted Boltzmann Machines (RBM), Deep Boltzmann Machine (DBM), Deep Neural Networks (DNN), etc. Among various techniques, the deep convolutional neural networks, which is a discriminative deep architecture and belongs to the DNN category, has found state-of-the-art performance on various tasks and competitions in computer vision and image recognition. Specifically, the CNN model consists of several convolutional layers and pooling layers, which are stacked up (Zhang et al., 2011) . The convolutional layer shares many weights, and the pooling layer sub-samples the output of the convolutional layer and reduces the data rate from the layer below. The weight sharing in the convolutional layer, together with appropriately chosen pooling schemes, endows the CNN with some invariance properties (e.g., translation invariance).

Deep Convolution Neural Network

Convolutional neural network (CNN) is a type of feed-forward artificial neural network where the individual neurons are tiled in such a way that they respond to overlapping regions in the visual field" (Rui et al., 1998). They are biologically inspired invariant of Multilayer Perceptron (MLP) which is designed for the purpose of minimal preprocessing. These models are widely used in image and video recognition. When CNNs are used for image recognition, they look at small portions of the input image called receptive fields with the help of multiple layers of small neuron collections which the model (Zhong Wu et al., 2010).The results we get from this collection are tiled in order for them to overlap such that a better representation of the original image is obtained; every such layer repeats this process (Ali & Sharma, 2017). This is the reason they are able if the input image is translated in any way. The outputs of neuron clusters are combined by local or global pooling layers which may be included in convolutional networks. Inspired by biological process, convolutional networks also contain various combinations of fully connected layers and convolutional layers, with point-wise nonlinearity applied at the end of or after each layer (Saritha et al., 2019), (Wan et al., 2014). The

convolution operation is used on small regions so as to avoid the situation when if all the layers are fully connected billions of parameters will exist. Convolutional networks use shared weights in the convolutional layers difficult to which hand-engineered feature. The layers in CNN are

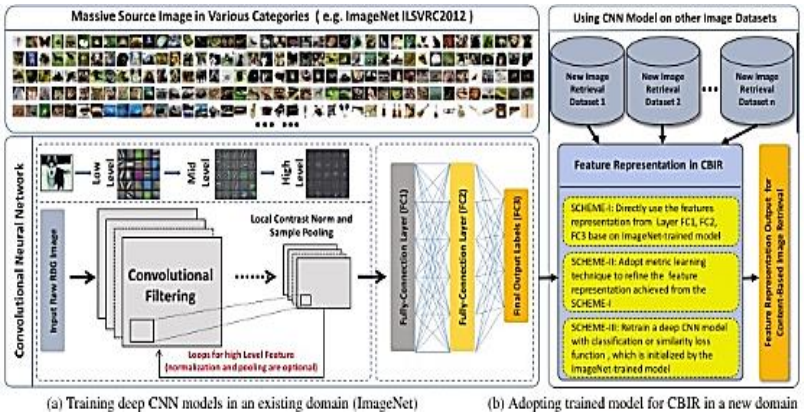

Fig-2: Framework of Deep Learning for CBIR

Convolution Layer

Feature map is obtained by repeatedly applying a function across sub-regions of the entire image, mainly by convolution of the input image with a linear filter, adding a bias term and then applying a non-linear function (Alzu'bi et al., 2015). The input feature maps are indexed by the leading dimensions, whereas the pixel coordinates is referred by the other two. When we combine it all as at layer m the weight that connects each pixel of the k-th feature map with the pixel of the i^{th} layer .

Max Pooling Layer

Max-pooling a form of non-linear down-sampling is an important concept of CNNs (Nistér & Stewénius, 2006). The input image is partitioned into a group of non-overlapping rectangles and a maximum value is given for each such sub-region.

Full Model: LeNet

The LeNet families of models have sparse, convolutional layers and max-pooling concepts as its core. The alternating convolution and max-pooling layers compose the lower-layers of the model." The upper-layers however are fully-connected and correspond to a traditional Multi-layer Perceptron which is a combination of hidden layer and logistic regression. The input to the first fully-connected layer is the set of all features maps at the layer below" (Philbin et al., 2007).

Deep Convolutional Network in CBIR

CBIR systems use visual features such as color, image edge, texture, and suitability of names in input images with images in the database (Zhang et al., 2011). CBIR system is established to overcome the problem like manual annotation and language indistinctness. The knowledge of computer vision is used for image retrieval based on some low-level features like color, shape, and texture. The most important parts of the CBIR system are the computational complexity and the retrieval accuracy (Ali & Sharma, 2017). The convolutional neural network (CNN) in image classification and detection, this paper proposes a simple and effective hybrid model of deep convolutional network and auto encoder network. This model uses the CNN network to extract the high-level semantic features of the image, then uses the depth auto encoder network to reduce the dimension of the extracted image features, and compresses the features into a 128-bit vector representation. Nearest Neighbor Search (ANN) is an effective strategy for large-scale image retrieval (Zhou et al., 2017). Content-based Image Retrieval (CBIR). The recent success of deep learning researches brings a hope for bridging the semantic gap. Many researchers have attempted to explore deep learning techniques with applying to CBRN tasks (Wang et al., 2015). The retrieval performance of a content-based image retrieval system crucially depends on the feature representation and similarity measurements. The ultimate aim of the proposed method is to provide an efficient algorithm to deal with the above mentioned problem definition(Lin et al., n.d.). In a CBIR system learning of effective feature representations and similarity measures is very important for the retrieval performance. Semantic gap has been the key challenge for this problem. A semantic gap exists between low-level image pixels captured by machines and the high-level semantics perceived by humans (Smeulders et al., 2000).The Support Vector Machine (SVM) has been used to support the learning process to reduce the semantic gap between the user and the CBIR system. SVM can classify the data into relevance training set and Gabor Filtering will extract the feature from the given image dataset. It can also

improve the performance of CBIR (Liu et al., 2007). The efforts of CBIR have been concentrated on reducing the semantic gap that exists between low-level image features represented by digital machines and the profusion of high-level human perception used to perceive images(Low, 2004). The effect of the convolutional network depth on it accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth using architecture with very small (3 × 3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers (Sivic & Zisserman, 2003). We propose a deep convolutional neural network architecture codenamed Inception that achieves the new state of the art for classification and detection in the Image Net Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The main hallmark of this architecture is the improved utilization of the computing resources inside the network (Su et al., 2003)(Li et al., n.d.).

Conclusion

For the small dataset with 1000 images the accuracy rate would be 98.6% but with a large data set (> 10000 images) the accuracy would be 96% without losing the time complexity requirement. The content features extraction seems to be reliable compared to the existing algorithms, the DBN generates a huge data set for learning features and provides a good classification to handle the finding of the efficient content extraction.

The framework has been implemented and extensively evaluated in different scenarios. In future enhancement similar mention can be forward to real time extraction.

References

1. Ali, A., & Sharma, S. (2017). Content based image retrieval using feature extraction with machine learning. *Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems, ICICCS 2017*, *2018-January*, 1048–1053. https://doi.org/10.1109/ICCONS.2017.8250625
2. Alzu'bi, A., Amira, A., & Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation*, *32*, 20–54. https://doi.org/10.1016/J.JVCIR.2015.07.012
3. Chum, O., Mikulík, A., Perdoch, M., & Matas, J. (2011). Total recall II: Query expansion revisited. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 889–896. https://doi.org/10.1109/CVPR.2011.5995601
4. Chum, O., Philbin, J., Sivic, J., Isard, M., & Zisserman, A. (n.d.). *Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval*.
5. Jegou, H., Douze, M., & Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *5302 LNCS*(PART 1), 304–317. https://doi.org/10.1007/978-3-540-88682-2_24
6. Li, X., Bertini, M., Li, X., Uricchio, T., Uricchio, ; T, Ballan, L., Bertini, M., & Bimbo, A. Del. (n.d.). 4 Socializing the Semantic Gap: A Comparative Survey on Image Tag Assignment, Refinement, and Retrieval. *ACM Computer Surveys*, *49*(14), 39. https://doi.org/10.1145/2906152
7. Lin, Z., Ding, G., Hu, M., & Wang, J. (n.d.). *Semantics-Preserving Hashing for Cross-View Retrieval*.
8. Liu, Y., Zhang, D., Lu, G., & Ma, W.-Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, *40*, 262–282. https://doi.org/10.1016/j.patcog.2006.04.045
9. Low, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 91–110. https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf
10. Nistér, D., & Stewénius, H. (2006). Scalable recognition with a vocabulary tree. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2*, 2161–2168. https://doi.org/10.1109/CVPR.2006.264
11. Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2007). Object retrieval with large vocabularies and fast spatial matching. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2007.383172
12. Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2008). Lost in quantization: Improving particular object retrieval in large scale image databases. *26th IEEE Conference onComputer Vision and Pattern Recognition, CVPR*. https://doi.org/10.1109/CVPR.2008.4587635
13. Rui, Y., Huang, T. S., Ortega, M., & Mehrotra, S. (1998). Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, *8*(5), 644–655. https://doi.org/10.1109/76.718510
14. Saritha, R. R., Paul, V., & Kumar, P. G. (2019). Content based image retrieval using deep learning process. *Cluster Computing*, *22*(February), 4187–4200. https://doi.org/10.1007/s10586-018-1731-0

15. Sivic, J., & Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. *Proceedings of the IEEE International Conference on Computer Vision*, *2*, 1470–1477. https://doi.org/10.1109/ICCV.2003.1238663
16. Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-Based Image Retrieval at the End of the Early Years. *Undefined*, *22*(12), 1349–1380. https://doi.org/10.1109/34.895972
17. Su, Z., Zhang, H., Li, S., & Ma, S. (2003). Relevance Feedback in Content-Based Image Retrieval: Bayesian Framework, Feature Subspaces, and Progressive Learning. *IEEE Transactions on Image Processing*, *12*(8), 924–937. https://doi.org/10.1109/TIP.2003.815254
18. Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., & Li, J. (2014). Institutional Knowledge at Singapore Management University Deep learning for content-based image retrieval : A comprehensive study Chinese Academy of Sciences. *Singapore Management University*, *7*(3), 157–166.
19. Wang, Y., Li, Q., Lan, T., & Chen, J. (2015). A comparison of Content based image retrieval systems. *Proceedings - 17th IEEE International Conference on Computational Science and Engineering, CSE 2014, Jointly with 13th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2014, 13th International Symposium on Pervasive Systems*, , 669–673. https://doi.org/10.1109/CSE.2014.143
20. Zhang, Y., Jia, Z., & Chen, T. (2011). Image retrieval with geometry-preserving visual phrases. *Undefined*, 809–816. https://doi.org/10.1109/CVPR.2011.5995528
21. Zhong Wu, Qifa Ke, Isard, M., & Jian Sun. (2010). Bundling features for large scale partial-duplicate web image search. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, *5*(3), 25–32. https://doi.org/10.1109/cvpr.2009.5206566
22. Zhou, W., Li, H., Lu, Y., & Tian, Q. (2011). Large scale image search with geometric coding. *MM'11 - Proceedings of the 2011 ACM Multimedia Conference and Co-Located Workshops*, 1349–1352. https://doi.org/10.1145/2072298.2072012
23. Zhou, W., Li, H., & Tian, Q. (2017). *Recent Advance in Content-based Image Retrieval: A Literature Survey*. http://www.gazopa.com/