Chinese Traditional Medicine Journal

ISSN : 2693 6356
2022 | Vol 5 | Issue 5

Effective Use of Convolutional Neural Networks for Deep Learning in CBIR

Lakshmeelavanya Alluri[1], Hemalatha Dendukuri[2]
Assistant Professor[1,2],
SRKR pharmacy College, Bhimavaram, Andhra Pradesh, India.

Abstract

Research into content-based image retrieval (CBIR) has attracted a great lot of attention. Users have a hard time coming up with the input since a CBIR system uses low-level observable characteristics, and the system itself doesn't provide satisfactory retrieval results. To boost the efficiency of the retrieval process in a CBIR system, it is crucial to investigate effective feature representations and suitable similarity metrics. The fundamental problem has been the semantic mismatch between the low-level picture pixels and the high-level semantics that people comprehend. Machine learning (ML) is one approach that has been investigated as a potential means of closing the semantic gap. In this research, we are motivated by the recent achievements of deep learning techniques in computer vision applications to take on an advanced deep learning technique called Convolutional Neural Network (CNN) in order to investigate its effects on the study of feature representations and similarity measures. In this research, we investigated how CNNs may be used to address classification and retrieval challenges. We decided to adapt the deep architecture for image retrieval by employing transfer learning. By calculating the Euclidean distance between each picture's feature vectors and our query image, we can determine which images are most similar to it in the dataset. These feature vectors were extracted from the retraining of the proposed CNN model's last-but-one fully connected layer.

INTRODUCTION

The popularity of search engines like Bing Image Search has skyrocketed in recent years. With the proliferation of image search engines like Gazopa's (Private Company) CBIR search engine, Imense Image Search Portal (Private Company), and Like.com (Private Company), picture retrieval has become a formidable challenge. The problems, restrictions, and time commitments inherent with metadata-based systems have increased curiosity in CBIR. Existing technology makes it easy to search through textual information, but it is impractical for humans to manually describe each image in a database, especially for very large databases or for images that will be generated automatically, like those from surveillance cameras. There are further problems, such as the possibility of missing photos whose descriptions use alternative comparable words. The miscategorization issue may be avoided using systems that classify photos according to semantic classes, such as "tiger" as a subclass of "animal," but the user will have to put in more work to find the images that might be tigers but are instead classified as "animals." To solve the image retrieval issue, or the challenge of searching for specific digital pictures inside vast databases, content-based image retrieval (CBIR) applies techniques for image collecting, preprocessing, analysis, representation, and comprehension. In contrast to more conventional methods, the CBIR system is built on the principles of concept-based image indexing (CBII) [1].The effectiveness of a CBIR system's retrieval function relies heavily on the accuracy with which features are represented and similarity measurements are taken. Despite the many methods proposed, this is still a difficult undertaking because of the disconnect between the image's pixels and the high-level semantics people experience. ML is a promising strategy for addressing this issue over the long run. When it comes to classification tasks and feature research, deep learning is a family of ML algorithms where several layers of data processing stages in hierarchical patterns are used [2]. The field of picture categorization has seen remarkable progress because to deep learning frameworks. There is a discrepancy between how photographs are categorized and how they are ranked in terms of similarity. Black boots, white boots, and dark-gray boots are all boots for the purposes of image classification, however when rating comparable photos, we often score the dark-gray boot higher than the white boot if the query image is a black boot. For data with a grid-like structure, such as picture data (a 2D grid of pixels), CNNs [2] are a specialized ANN. Simply said, convolutional neural networks (CNNs) are ANNs in which at least one layer replaces matrix multiplication with convolution. Convolution allows for parameter sharing, equivariant representations, and sparse interactions, all of which might aid in the development of a better ML system. This challenge is well-suited for the use of neural networks, and CNNs in particular are well-known for their capacity to learn forms, textures, and colors.

Using a state-of-the-art deep learning system—CNNs for examining feature representations from image data—we conducted research on a deep learning architecture for CBIR systems. In general, we retrain the previously learned CNN model using our data. Next, the trained network is put to use to do two things: categorize objects into their correct classes and conduct a nearest-neighbors analysis to return the most comparable and relevant photos to the input image [3-4].

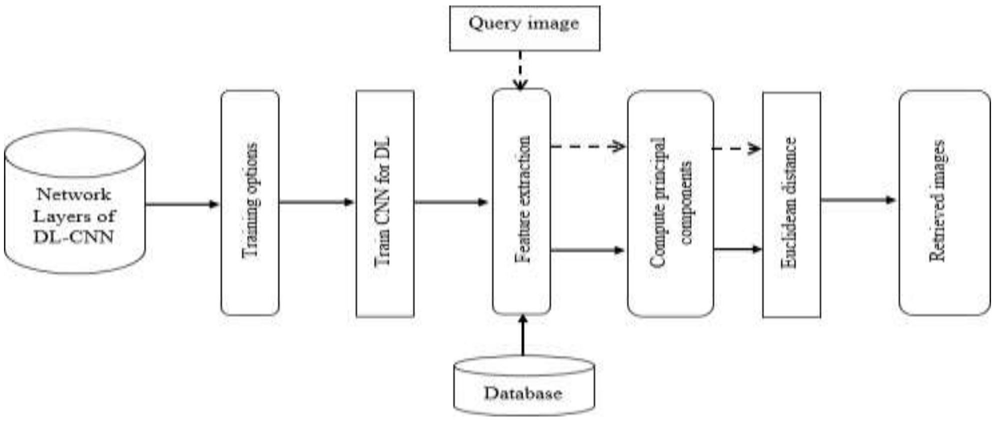
I. RELATED WORK

The 1.2 million pictures in the ImageNet dataset were classified into 1000 categories using a deep CNN trained by Krizhevsky et al. [5]. The authors developed an eight-layer network, the first five of which were convolutional layers and the last three of which were completely linked. This network has been trained using two GTX 580 GPUs, each with 3GB of memory, since the memory limitations of a single GPU would prevent it from being trained with a larger dataset. Top-1 error rates of 37.5% and top-5 error rates of 17.0% were attained by the authors when they employed characteristics derived from the 7th layer to retrieve related images. However, Babenko et al. [6] recommended to compress the features using dimensionality reduction approach and achieved a decent performance due to the large dimensionality of CNN features and the inefficiency of similarity calculation between two 4096-dimensional vectors. For hash learning, researchers have turned to deep models. The revolutionary performance of image retrieval using deep learning was shown by Xia et al. [7], who introduced a supervised hashing approach to investigate binary hash codes. They have utilized a matrix decomposition approach to examine the codes being used to represent the data in a preliminary processing stage. However, in the event of enormous data, this step is essential due to the increased storage and processing needs. Using hashing-based approaches that project the high-dimensional features to low-dimensional feature space and construct the binary hash codes, Lin et al. [8] suggested an easy and effective supervised learning model for rapid image retrieval system. This method uses binary pattern matching techniques or Hamming distance calculation, both of which enhance search efficiency while drastically decreasing computing time. The authors say that whereas it takes 109.767ms to compute the Euclidean distance between two 4096-dimensional feature vectors, it only takes 0.113ms to compute the Hamming distance between two 128-bit binary codes. Adding more layers to the network and more neurons to each layer is the simplest technique to improve the performance of Deep Neural Networks (DNNs). Szegedy et al. [3] developed a deep CNN architecture called Inception that outperformed previous methods on the ImageNet dataset for both classification and detection. The efficiency with which this strategy utilizes networked computer resources is its major metric. The writers have expanded the network's reach and depth. The architectural tradeoffs were made with the Hebbian principle in mind. This architecture is useful for dramatically expanding the number of neurons at each stage without introducing a corresponding increase in computing complexity. Increases in both the breadth of each step and the total number of steps are now possible because to more efficient use of computer resources. using the goals of addressing clothing style categorization and related clothes retrieval, Chen et al. [9] investigated Deep Learning using CNNs. Transfer learning is used to simplify training by refining pre-trained structures on huge datasets. The model is built around a collection of deep networks, each of which was trained on a small subset of data due to the massive size of deep networks' parameters. On three apparel datasets, our strategy outperformed previous methods that use ML algorithms with shallow structure, with an 18% increase in accuracy seen on the biggest dataset including 80,000 photos.In their study, Khosla and Venkataraman [10] use a strategy wherein additional CNNs are trained on the shoe dataset and then used to categorize input shoe images into suitable shoe classes and to conduct the closest neighbors assessment to return K most similar shoes to the given input shoe picture. The closest matches to the input picture were returned using Caffe as the neural network architecture and the Euclidean distance measure. The average score was 4.12 out of 5, and the retrieval accuracy was 75.6%, both of which were attained using this method of determining the Euclidean distance between the features vectors of the pictures. For fine-grained image categorization using pretrained neural networks, Iliukovich-Strakovskaia et al. [11] proposed a 'Two Flow Model' in which the input picture is processed in two distinct ways. In the first process, the picture is transformed into a low-dimensional feature vector space using certain conventional dimensionality reduction algorithms, and then the most useful features are selected. The second route processes the picture using a previously trained CNN, which uses the features from the global pooling layer to refine the feature selection process. At last, a nonlinear classifier is fitted using a combined set of characteristics extracted from both flows. Using Inception_BN and Inception_21k as examples of pretrained deep neural networks, this method employs Random Forest for both feature selection and the final step of a nonlinear classifier. By comparison, Inception_21k achieved 69.3% accuracy on the global pooling layer, whereas Inception_BN deep neural network accuracy ranged from 55.5% to 68.0%.

II. PROPOS ED IMPLEMENTATION

This section describes proposed methodology which employs DConvNet for CBIR system.

Working of CNN can be explained as follows: A 2-D convolutional layer applies sliding filters to the input. The layer convolves the input by moving the filters along the input vertically and horizontally and computing the dot product of the weights and the input, and then adding a bias term. A Re LU layer performs a threshold operation to each element of the input, where any value less than zero is set to zero. A max pooling layer performs down-sampling by dividing the input into rectangular pooling regions and computing the ma ximu m of each region. A fully connected layer multiplies the input by a weight matrix and then adds a bias vector.



3.1. DL-CNNFig. 1. Proposed DConvNet for CBIR system

eigenvalue
decomposition of every
covariancematrix,
which gives the list of
eigenvalues or

According to the facts, training and testing of DL-CNN involves in allowing every source image via a succession of convolution layers by a kernel or filter, rectified linear unit (Re LU), max pooling, fully connected layer and utilize SoftMax layer
with classification layer to categorize the objects with probabilistic values ranging from [0,1]. Figure
1 discloses the architecture of DL-CNN that is utilized in proposed methodology for CBIR system for enhanced feature representation of word image over conventional retrieval systems.

3.2. Principal component analysis

Principal component analysis is an approach of machine learning which is utilized to reduce the dimensionality. It utilizes simple operations of matrices from statistics and linear algebra to compute a projection of source data into the similar count or lesser dimensions. PCA can be thought of a projection approach where data with *m*-columns or features are projected into a subspace by *m* or even lesser columns while preserving the most vital part of source data. Let *I*be a source image matrix with a size of $n * m$ and results in *J* which is a projection of *I*. The primary step is to compute the value of mean for every column. Next, the values in every column are centered by subtracting the value of mean column. Now, covariance of the centered matrix is computed. At last, compute theeigenvectors. These eigenvectors constitute the directions or components for the reduced subspace of *J*, whereas the peak amplitudes for the directions are represented by these eigenvectors. Now, these vectors can be sorted by the eigenvalues in descending order to render a ranking of elements or a xes of the new subspace for *I*. Genera lly, *k* eigenvectors will be selected which are referred principal components or features

3.3. Euclidean distance

To evaluate distances between query word image I_q and retrieved word images I_r, a metric must be defined. We need a measurement method to tell how the query and retrieved word images are similar (bit per bit). Therefore, we want a similarity measure where the distance value will be the number of similar bits in the considered images.

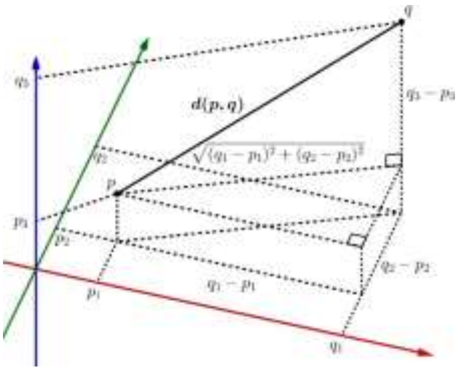


Fig. 2. Illustration of Euclidean distance

Fig. 3. Retrieved dog images using DConvNetCBIR system

III. RESULTS AND DISCUSSION

In this section we discussed the simulation results of CBIR system. The proposed algorithm has been tested with few databases and displayed the outputs in the below figures. Fig. 3 shows that retrieving images using proposed CBIR scheme. Similarly, proposed retrieval system has been shown in fig. 4 and 5 with different classification images. As a measure of performance, we have used two widely used metrics of Precision and Recall. Precision is a measure of ability of CBIR algorithm to retrieve only relevant images, while Recall decides the ability of CBIR algorithm to retrieve all relevant images as defined by eq. (1) and eq. (2) respectively.

Fig. 4. Retrieved car images using DConvNetCBIR system

Fig. 5. Retrieved bird images using DConvNetCBIR system

IV. CONCLUSIONS

Using DConvNet and PCA with pairwise hamming distance, this paper demonstrated a powerful CBIR system. The suggested CBIR system showed higher performance in simulations, with more relevant pictures being retrieved. Additionally, mAP and mAR are used to illustrate how the proposed CBIR system's performance compares to that of previously described CBIR systems.

REFERENCES

[1] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "Asurvey of content-based image retrieval with high-level semantics," Pattern recognition, vol.40, no. 1, pp. 262–282, 2007.

[2] Y. Le Cun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[3] C. Szegedy,W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Go ing deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[4] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818– 2826.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural in formation processing systems, 2012, pp. 1097–1105.

[6] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval,"in European conference on computer vision. Springer, 2014, pp. 584–599.

[7] R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan, "Supervised hashing for image retrieval via image representation learning." in AAAI, vol. 1, 2014, p. 2.

[8] K. Lin, H.-F. Yang, J.-H. Hsiao, and C.-S. Chen, "Deep learning of binary hash codes for fast image retrieval," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 27–35.

[9] J.-C. Chen and C.-F. Liu, " Visual-based deep learning for clothing from large database," in Proceedings of the ASE Big Data & Social Informatics 2015. ACM, 2015, p. 42.

[10] N. Khosla and V. Venkataraman, " Building image-based shoe search using convolutional neural networks," CS231n Course Project Reports, 2015.

[11] A. Iliu kovich-Strakovskaia, A. Dral, and E. Dral, " Using pre-trained models for fine-grained image classification in fashion field," 2016.

[12] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition." in Icml, vol. 32, 2014, pp. 647–655.

[13] G. Shrivakshan, C. Chandrasekar et al., "A comparison of various edge detection techniques used in image processing," IJCSI International Journal of Computer Science Issues, vol. 9, no. 5, pp. 272–276, 2012.

[14] A. Maurya and R. Tiwari, "A novel method of image restoration by using different types of filtering techniques," International Journal of Engineering Science and Innovative Technology (IJESIT) Vo lume, vol. 3, 2014.

[15] R. Kandwal, A. Kumar, and S. Bhargava, "Review: existing image segmentation techniques," International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 4, 2014.

[16] K. Roy and J. Mukherjee, "Image similarity measure using color histogram, color coherence vector, and sobel method," International Journal of Science and Research (IJSR), vol. 2, no. 1, pp. 538–543, 2013.

[17] J. Shlens, "Train your own image classifier with Inception in Tensor- Flow," https://research.googleblog.com/2016/03/train-your-ownimageclassifier-with.html, 2016.

[18] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. Le- Cun, "Over feat: Integrated recognition, localization and detection using convolutional networks," arXiv preprint arXiv:1312.6229, 2013.

[19] P. Wu, S. C. Hoi, H. Xia, P. Zhao, D. Wang, and C. Miao, "Online multimodal deep similarity learning with application to image retrieval," in Proceedings of the 21st ACM international conference on Multimedia. ACM,2013, pp. 153–162.

[20] S. Liu, Z. Song, G. Liu, C. Xu, H. Lu, and S. Yan, "Street-to shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 3330–3337.

[21] K. Yamaguchi, M. H. Kiapour, L. E. Ortiz, and T. L. Berg, "Retrieving similar styles to parse clothing," IEEE transactions on pattern analysis and machine intelligence, vol. 37, no. 5, pp. 1028–1040, 2015.

[22] J. Wan, P. Wu, S. C. Hoi, P. Zhao, X. Gao, D. Wang, Y. Zhang, and J. Li, " Online learning to rank for content-based image retrieval," 2015.