Turkish Journal of Computer and Mathematics Education Vol.11 No.03 (2020), 1579- 1584
Research Article

Deep Learning Convolutional Neural Networks for Content Based Image Retrieval

Ramakrishna Pasula[1], Ramesh Banoth[2], Ramesh Gugulothu[3]
[1,2]Assistant Professor, [3]Associate Professor, [1,2,3]Department of ECE
[1,2,3]Siddhartha Institute of Technology and Sciences, Telangana, India.

Abstract
Content Based Image Retrieval (CBIR) has received an excellent deal of interest within the research community. A CBIR system operates on the visible features at low-level of a user's input image that makes it troublesome for the users to devise the input and additionally doesn't offer adequate retrieval results. In CBIR system, the study of the useful representation of features and appropriate similarity metrics is extremely necessary for improving the performance of retrieval task. Semantic gap has been the main issue which occurs between image pixels at low-level and semantics at high-level interpreted by humans. Among varied methods, machine learning (ML) has been explored as a feasible way to reduce the semantic gap. Inspired by the current success of deep learning methods for computer vision applications, in this paper, we aim to confront an advance deep learning method, known as Convolutional Neural Network (CNN), for studying feature representations and similarity measures. In this paper, we explored the applications of CNNs towards solving classification and retrieval problems. For retrieval of similar images, we agreed on using transfer learning to apply the deep architecture to our problem. Extracting the last-but-one fully connected layer from the retraining of proposed CNN model served as the feature vectors for each image, computing Euclidean distances between these feature vectors and that of our query image to return the closest matches in the dataset.

I. INTRODUCTION

Recent years there is a rapid growth in searching engines such as Bing image search: Microsoft's CBIR engine (Public Company), Google's CBIR system, note: does not work on all images (Public Company), CBIR search engine, by Gazopa (Private Company), Imense Image Search Portal (Private Company) and Like.com (Private Company), image retrieval has become a challenging task. The interest in CBIR has grown because of the retrieval issues, limitations and time consumption in metadata-based systems. We can search the textual information very easily by the existing technology, but this searching methods requires humans to describe each images manually in the database, which is not possible practically for very huge databases or for the images which will be generated automatically, e.g. images generated from surveillance cameras. It has more drawbacks that there is a chance to miss images that use different equivalent word in the description of images. The systems based on categorizing images in semantic classes like “tiger” as a subclass of “animal” can debar the miscatergorization problem, but it will requires more effort by a use to identify the images that might be “tigers” , but all of them are categorized only as an “animal”. Content-based image retrieval (CBIR) is a application of methods of acquisition, pre-processing, analyzing, representation and also understanding images to the image retrieval problem, that is the problem of exploring for digital images from large databases. The CBIR system is opposed to traditional approaches, which is known on concept-based approaches i.e., concept-based image indexing (CBII) [1].Representation of features and similarity measurements are critical for the retrieval performance of a CBIR system. Various approaches have been suggested, but even then, it remains as a challenging task due to the semantic gap present between the image pixels and high-level semantics perceived by humans. One favorable approach is ML that aims to solve this problem in the long-term. Deep learning represents a category of ML approaches where several layers of data processing steps in hierarchical layouts are utilized for classification task and study of features [2]. Deep learning frameworks have attained great achievements in image classification. However, the ranking of similar images is inconsistent with the classification of images. For classification of images, "black boots," "white boots" and "dark-gray boots" are all boots, but for ranking of similar images, if a query image is a "black boot," we conventionally want to rank the "dark gray boot" higher than the "white boot." CNNs [2] are a specific type of ANN for handling data that features a grid-like topology like, image data, which is a 2D grid of pixels. CNNs are merely ANNs that involves the use of convolution instead of conventional matrix multiplication operation in a minimum of one in all their layers. Convolution supports three essential concepts that can facilitate in improving a ML system: parameter sharing, equivariant representations, and sparse interactions. CNNs are eminent for their potential to learn shapes, textures, and colors, making this problem suitable for the application of neural networks.

In this, we investigated an architecture of deep learning for CBIR systems by applying an advanced deep learning system, that is, CNNs for studying feature representations from picture data. Overall, our approach is to retrain the pre-trained CNN model, that is, on our dataset. Then, the trained network is used to perform two

tasks: classify objects into its appropriate classes and perform a nearest-neighbors analysis to return the most similar and most relevant images to the input image [3-4].

II. RELATED WORK

Krizhevsky et al. [5] trained a deep CNN to classify ImageNet dataset consisting of 1.2 million images into 1000 different classes. The authors worked on a network containing eight layers, where first five were convolutional layers, and last three were fully connected layers. Since a single GTX 580 GPU with 3GB memory bounds the maximum network size for training, therefore, this network has been trained on two GTX 580 3GB GPUs. The authors used the features extracted from 7th layer to fetch similar pictures and achieved the top-1 error rate of 37.5% and top-5 error rates of 17.0%. However, because of the high dimensionality of CNN features and inefficiency of similarity computation between two 4096-dimensional vectors, Babenko et al. [6] suggested to compress the features using dimensionality reduction method and attained a good performance. Deep models have been used for hash learning. Xia et al. [7] proposed a supervised hashing method to study binary hash codes to retrieve images using deep learning and revealed the revolutionary performance of retrieval on datasets that are publicly available. In a pre-processing step, they have used a matrix decomposition algorithm for studying the codes to represent the data. But this stage is critical in case of large data as it consumes storage and requires more computational time. Lin et al. [8] proposed a straightforward and efficient supervised learning model for fast image retrieval system using hashing-based methods that project the high dimensional features to low-dimensional feature space and produce the binary hash codes. This approach used binary pattern matching methods or Hamming distance calculation that greatly reduces the computational time and also optimizes the search efficiency. The authors have claimed that Euclidean distance computation between two 4096-dimensional feature vectors requires 109.767ms while Hamming distance computation between two 128 bits binary codes require 0.113ms, thus reducing the time complexity. The easiest way of enhancing the performance of Deep Neural Networks (DNNs) is by increasing the number of layers in the network as well as the number of neurons in each layer. Szegedy et al. [3] presented a deep CNN architecture, Inception, that achieved the state-of-the-art performance for image classification and image detection tasks in the ImageNet dataset. The primary indicator of this model is the effective use of computing resources in the network. The authors have increased the width and depth of the network. The architectural decisions are based on the Hebbian principle to optimize quality. This structure helps to increase the number of neurons at each step remarkably without increasing computational complexity in later steps. The improved usage of computational resources permits the increment of the width of each step and the number of steps without getting into computational problems. Chen et al. [9] explored Deep Learning with CNNs with an aim of solving clothing style classification and similar clothing retrieval. To lower the complexity of training, transfer learning is used by fine tuning pre-trained structures on large datasets. Since the parameters are enormous for any deep network, the model is designed to use multiple deep networks trained with a sub-dataset. Compared with the existing approaches that use ML algorithms with shallow structure, this method provided more likely outcomes on three clothing datasets, particularly on the large dataset with 80,000 images where an improvement of 18% in accuracy was recognized.The approach of Khosla and Venkataraman [10] research is to train other CNNs on the shoe dataset and then use these trained networks to classify input shoe image into appropriate shoe class and perform the nearest neighbors evaluation to return K most similar shoes to the given input shoe image. The authors used Caffe as neural network architecture and Euclidean distance metric to return the closest matches to the input image. This approach of computing Euclidean distance between the features vectors of the images has achieved 75.6% precision on retrieval process and an average score of 4.12 out of 5. Iliukovich-Strakovskaia et al. [11] suggested a 'Two Flow Model' for fine-grained image classification based on pretrained neural networks where the given input image goes through several processing flows. In the first flow, the image which is considered as a feature vector of raw pixels is reduced to a low-dimensional feature vector space using some standard dimensionality reduction methods and then feature selection stage is used to choose the most informative features. In the second flow, the image goes through pre-trained CNN, and the features from global pooling layer are then utilized in the next processing stage to select the most informative features. Finally, the features extracted from both the flows are merged to fit a nonlinear classifier. In this approach, pretrained deep neural networks such as Inception_BN and Inception_21k are used and Random Forest was used at feature selection stage and the last stage of nonlinear classifier. Using this model, the accuracy of using Inception_BN deep neural network varied between 55% and 68% depending on the layer used for features while Inception_21k gave 69.3% accuracy on global pooling layer.

III. PROPOSED 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 ReLU 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 maximum of each region. A fully connected layer multiplies the input by a weight matrix and then adds a bias vector.

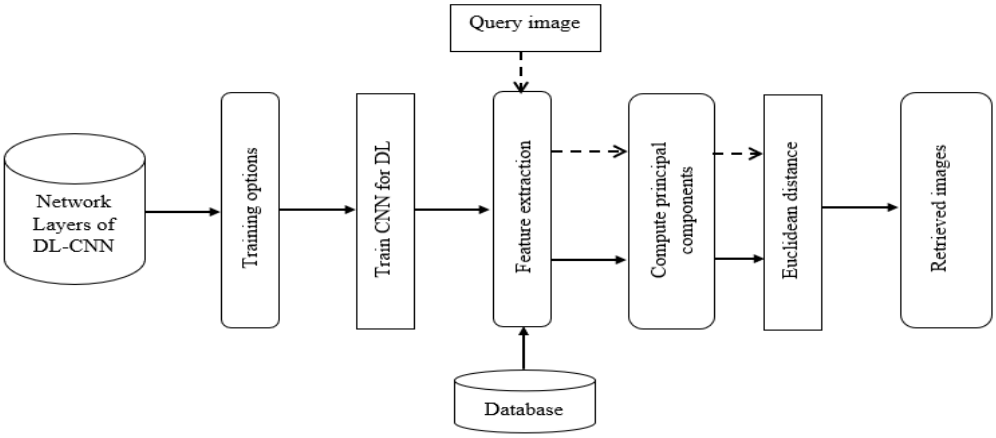

Fig. 1. Proposed DConvNet for CBIR system

3.1. DL-CNN

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 (ReLU), 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 the eigenvalue decomposition of every covariance matrix, which gives the list of eigenvalues or eigenvectors. 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 axes of the new subspace for *I*. Generally, *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.

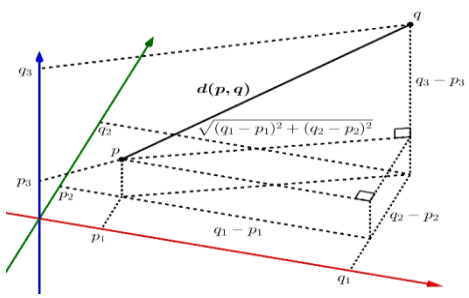

Fig. 2. Illustration of Euclidean distance

IV. 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.

$$P = \frac{Total\ numebr\ \ of\ relevant\ \ images\ \ retrieved}{Total\ \ numebr\ \ of\ retrieved\ \ images} \quad (1)$$

$$R = \frac{numebr\ \ of\ relevant\ \ images\ \ retrieved}{numebr\ \ of\ relevant\ \ images\ \ in\ database} \quad (2)$$

Fig. 3. Retrieved dog images using DConvNet CBIR system

Fig. 4. Retrieved car images using DConvNet CBIR system

Fig. 5. Retrieved bird images using DConvNet CBIR system

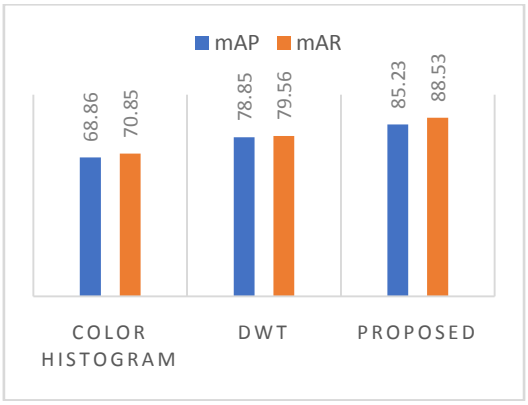

Fig. 6.Performance of mAP and mAR with proposed and existing CBIR systems

V. CONCLUSIONS

This article presented an efficient CBIR system using DConvNet and PCA with pairwise hamming distance. Simulation results disclosed that proposed CBIR system obtained superior performance by retrieving more relevant images. Further, the performance evaluation of proposed CBIR system is demonstrated using mAP and mAR and compared with the existing CBIR systems presented in the literature.

REFERENCES

[1] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," Pattern recognition, vol. 40, no. 1, pp. 262–282, 2007.
[2] Y. LeCun, 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, "Going 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 information 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. Iliukovich-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) Volume, 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.