Indexing 100M Images with Deep Features and MI-File

Giuseppe Amato, Fabrizio Falchi, Claudio Gennaro, and Fausto Rabitti

ISTI-CNR
via G. Moruzzi, 1 - 56124 Pisa, Italy
`<name>.<surname>@isti.cnr.it`

Abstract. In the context of the Multimedia Commons initiative, we extracted and indexed deep features of about 100M images uploaded on Flickr between 2004 and 2014 and published under a Creative Commons commercial or noncommercial license. The extracted features and an online demo built using the MI-File approximated data structure are both publicly available. The online CBIR system demonstrates the effectiveness of the deep features and the efficiency of the indexing approach.

Keywords: Deep Features, MI-File, Content-Based Image Retrieval, Similarity Search

1 Introduction

Deep Convolutional Neural Networks (DCNNs) have recently shown impressive performance on a number of multimedia information retrieval tasks [6, 9, 4]. In particular, the activation of the DCNN hidden layers has been also used in the context of transfer learning and conten-based image retrieval [3, 8]. In fact, Deep Learning methods are "representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level" [7]. These representations can be successfully used as features in generic recognition or visual similarity search tasks.

In this paper we present a public online Content-Based Image Retrieval system indexing about 100M images. The dataset is the YFCC100M which is the largest Creative Commons image dataset available today. The deep features were extracted using a public available DCNN using the Caffe[5] framwork and can be downloaded from http://www.deepfeatures.org. The 4,096-dimensional features vectors were indexed using MI-File[1], a permutation-based approximated data structure. The online demo is available at http://mifile.deepfeatures.org. A screenshot of the web bases interface can be seen in Figure 1

2 The YFCC100M Deep Features Dataset

The Yahoo Flickr Creative Commons 100 Million (YFCC100M) dataset[1] was created in 2014 as part of the Yahoo Webscope program. YFCC100M consists of 99.2 million photos and 0.8 million videos uploaded to Flikcr between 2004 and 2014 published under a Creative Commons commercial or noncommercial license. More information about the dataset can be found in the recent article in Communications of the ACM [10].

For extracting deep features we used the Caffe [5] deep learning framework. In particular we used the neural network Hybrid-CNN whose model and weights are public available in the Caffe Model Zoo[2]. The Hybrid-CNN was trained on 1,183 categories (205 scene categories from Places Database and 978 object categories from the train data of ILSVRC2012 (ImageNet) with 3.6 million images [11]. The architecture is the same as Caffe reference network. The deep features we have extracted are the activation of the fc6 layer.

We have made them public available at http://www.deepfeatures.org and they will be soon included in the Multimedia Commons initiative corpus. The Multimedia Commons initiative[3] is an effort to develop and share sets of computed features and ground-truths for the YFCC100M.

3 MI-File

Recently, permutation based indexes have attracted interest in the area of similarity search. The basic idea of permutation based indexes is that data objects are represented as appropriately generated permutations of a set of pivots (or reference objects). Similarity queries are executed by searching for data objects whose permutation representation is similar to that of the query. This, of course assumes that similar objects are represented by similar permutations of the pivots.

One of the most promising permutation based approach is the The Metric Inverted File (MI-File) [1]. It uses an inverted file to store relationships between permutations. It also uses some approximations and optimizations to improve both efficiency and effectiveness. The basic idea is that entries (the lexicon) of the inverted file are the pivots P. The posting list associated with an entry $p_i \in P$ is a list of pairs $(o, \Pi_o^{-1}(i))$, $o \in C$, i.e. a list where each object o of the dataset C is associated with the position of the pivot p_i in Π_o.

As already mentioned, in [1] it was observed that truncated permutations can be used without huge lost of effectiveness. MI-File allows truncating the permutation of both data and query objects independently. We denote with l_x the length of the permutation used for indexing and with l_s the one used for searching (i.e. the length of the query permutation).

1 http://bit.ly/yfcc100md
2 http://github.com/BVLC/caffe/wiki/Model-Zoo
3 http://multimediacommons.wordpress.com/

Fig. 1. Screen shot of the on line image content based search engine

The MI-File also uses a strategy to read just a small portion of the accessed posting lists, containing the most promising objects, further reducing the search cost. The most promising data objects in a posting list, associated with a pivot p_i for a query q, are those whose position of the pivot p_i, in their associated permutation, is closer to the position of p_i in the permutation associated with q. That is, the promising objects are the objects o, in the posting list, having a small $|\Pi_o^{-1}(i) - \Pi_q^{-1}(i)|$. To control this, a parameter is used to specify a threshold on the maximum allowed position difference (mpd) among pivots in data and query objects. Provided that entries in posting lists are maintained sorted according to the position of the associated pivot, small values of mpd imply accessing just a small portion of the posting lists.

Finally, in order to improve effectiveness of the approximate search, when the MI-File execute a k-NN query, it first retrieves $k \cdot amp$ objects using the inverted file, then selects, from these, the best k objects according to the original distance. The factor $amp \geq 1$, is used to specify the size of the set of candidate objects to be retrieved using the permutation based technique, which will be reordered according to the original distance, to retrieve the best k objects.

The MI-File search algorithm computes incrementally a relaxed version of the Footrule Distance with location parameter l between the query and data objects retrieved from the read portions of the accessed posting lists.

4 Conclusion and Future Work

In this work, we presented an online CBIR system which indexes, using MI-File, a dataset of deep features extracted from 100M images that are part of the well-known and public available YFCC100M dataset. This system demonstrate the effectiveness of the features extracted and the efficiency of the MI-File indexing approach.

In the future, we plan to release results obtained sequentially scanning the entire set of deep features in order to measure effectiveness of approximate indexing approaches. We hope that our deep features corpus will become the new reference for content-based image retrieval on a large scale updating our previous CoPhIR[2] dataset. CoPhIR[4] also consists of about 100M images taken from Flickr. However, it also contains copyrighted images and deep features for the images are not available.

References

1. Amato, G., Gennaro, C., Savino, P.: Mi-file: using inverted files for scalable approximate similarity search. Multimedia tools and applications 71(3), 1333–1362 (2014)
2. Bolettieri, P., Esuli, A., Falchi, F., Lucchese, C., Perego, R., Piccioli, T., Rabitti, F.: Cophir: a test collection for content-based image retrieval. arXiv preprint arXiv:0905.4627 (2009)
3. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. arXiv preprint arXiv:1310.1531 (2013)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015)
5. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv:1408.5093 (2014)
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. pp. 1097–1105 (2012)
7. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521(7553), 436–444 (2015)
8. Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 806–813 (2014)
9. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
10. Thomee, B., Elizalde, B., Shamma, D.A., Ni, K., Friedland, G., Poland, D., Borth, D., Li, L.J.: Yfcc100m: The new data in multimedia research. Communications of the ACM 59(2), 64–73 (2016)
11. Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., Oliva, A.: Learning deep features for scene recognition using places database. In: Advances in neural information processing systems. pp. 487–495 (2014)

[4] http://cophir.isti.cnr.it/