Deep learning feature based medical image retrieval for large-scale datasets

Nandinee Fariah Haq[1] nandinee@ece.ubc.ca
Mehdi Moradi[2] mmoradi@us.ibm.com
Z. Jane Wang[1] zjanew@ece.ubc.ca

[1]University of British Columbia, Vancouver, Canada
[2]IBM Almaden Research Center, San Jose, USA

Abstract

Automated content based methods for retrieval of similar images serve as efficient management tools for handling large amounts of data. In medical imaging, they can also be used as a tool in clinical decision support systems. In this work we propose a medical image retrieval framework for extracting similar images from large-scale databases. The framework consists of a deep learning based feature extractor, an algorithm to build a graph of images, and a graph clustering method to extract similar images from this large image graph given a query image. We evaluated the performance of the method on CheXpert dataset, where given a query image the proposed approach extracted images with similar disease labels with a retrieval precision of 73%. The application of the framework evaluated on the CheXpert dataset indicates its potential as a large-scale image retrieval tool.

1 Motivation

Content-based medical image retrieval has been an active field of research for the past few years. In content-based retrieval systems the images are first represented in terms of high dimensional features, and similarity between images are measured from distances between these features. The selection of feature space is very important in medical image retrieval systems since the performance of the whole system depends on the extracted features. The success of deep-learning based approaches in other clinical applications indicate deep learning based features can add value to image retrieval techniques. To extract image-representative codes or features, the current deep-learning based retrieval approaches mostly use networks pre-trained on other image databases [1, 2]. However medical images can be very different from natural images and hence pre-trained networks might not be the proper feature extractor for medical settings. Although the framework reported in [3] used a custom network trained on medical image data, its performance was inferior to the state of the art methods. Moreover, while retrieving similar images, the literature-based methods compare every image features in the database to extract the closest matches [4, 5], which is computationally expensive when there are a large number of images in the database. Hence a prior clustering technique that can extract similar image subspace beforehand can be highly beneficial for image retrieval techniques since a query image can then be only compared with its most similar subset of images. In this work, we report a deep-learning and graph-clustering based medical image retrieval framework. The images were represented by features extracted from a deep-learning based network trained on the medical images. To implement an efficient search engine for medical image retrieval task the database was also divided offline into clusters of most similar images. During online retrieval the search space was reduced to the closest cluster space and only a subset of the database images were tested for similarity, hence reducing the computational overhead significantly.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

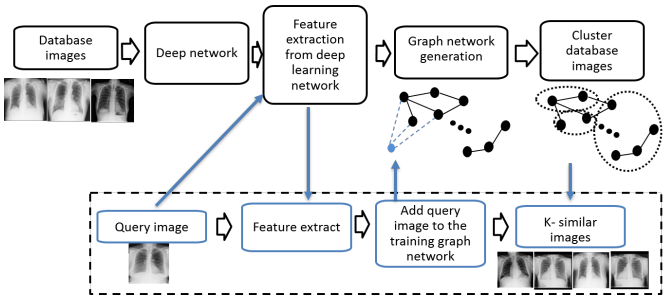

Figure 1: Framework

2 Method

We formulated the medical image retrieval from large dataset as a deep-leaning based cluster extraction problem that consists of three parts- image code generation,image graph formulation, and similar image cluster formation. The framework is shown in Fig. 1.

2.1 Image code generation

We start with a database of images, $D = \{I_f\}$. A deep convolutional neural network is then trained on the augmented data to generate disease likelihood from the X-ray images. In this work we considered a model proposed in [6] which consists of densely connected convolutional layer blocks, known as DenseNet. The initial weights for other layers are assigned from a model pretrained on ImageNet dataset [7]. To handle the class imbalance present in the medical image datasets used here, we optimized the weighted binary cross-entropy function as a loss function during training. The weighted were defined as the ratio of positive and negative samples over the total number of samples. After training the network, the 1024-dimensional feature from the second-last layer was extracted for each image. The normalized feature vector, i_f was then used as the image-representative code for the database images and the codes were saved with the associated images in the database, $D = \{I_f, i_f\}$.

2.2 Image graph formulation

The next step of the framework consists of generating a graph of similar images. The database can be thought of a graph of images where similar images are strongly connected with each other whereas images that are different are either loosely connected or unconnected. From the database, we generated a graph of images $\mathcal{G}$ where each image is represented by a node in the graph and similar images share edges between them. The edges were calculated from image code-similarity and the similarity between samples was defined as the cosine similarity between their representative codes, i_f.

2.3 Similar image cluster formation

The next step of the framework consists of finding similar image clusters from the database of images. From the graph $\mathcal{G}$, we extracted similar image clusters by applying a modularity based graph

Algorithm 1 Top-K similar image retrieval for a query image, I

Input : Graph $\mathcal{G}$, Edge generating function function E, Database $\{I_f, i_f, C(I_f)\}$, Query image I, Number of images to retrieve K
Add I to $\mathcal{G}$ with edges $E(I, I_f)$
Set region $R = I_f$
while $|R| < K$ **do**
    Select cluster c, associated database image set I_c s.t. maximum weighted modularity is achieved when R is assigned to c^{th}-cluster
    for $j = 1, 2, .., min(|I_c|, K)$ **do**
        Select I_u s.t. $I_u = \arg\max_{I_j} \Delta q(I, I_j); I_j \in I_c, I_j \notin R$ and $C(I_j) = c$
        Add I_u to R
    end for
end while

Table 1: Performence of the image retrieval framework on top-50 retrieved images.

Dataset	ACG	ACG_p	Precision
CheXpert-5	0.65	0.54	78%
CheXpert	0.65	0.37	73%

Table 2: Comparison of the proposed approach with literature-based methods.

Method	Avg. Gain
Lan *et al.*[10]	0.31
Chen *et al.*[5]	0.42
Proposed Framework	0.57

clustering method proposed in [8, 9], that has the ability to find clusters from graphs without any prior knowledge regarding the number and sizes of clusters. The database is then updated with the clustering label associated with each image, $D = \{I_f, i_f, C(I_f)\}$. While a new query image is presented, at first its image-representative code is extracted from the trained network and the query image is placed in the graph, $\mathcal{G}$ where the edges between the query image and the database images are generated based on their code similarity. Then top-K most similar images were extracted by implementing a region-growing approach with the query image as the seed point. The region growing approach is shown is Algorithm. 1. We start with a region that only includes the query image I. To assign database images to the region of the query image we first look for the closest cluster in terms of the weighted-modularity, q [9] and confine our search space only to that cluster. The image-nodes were added successively to the region, R for which the maximum weighted-modularity is achieved until K images are retrieved from the database.

3 Experiments and results

For this work we used the recently published CheXpert dataset [11], which contains 223,648 chest X-ray images from 64,740 unique patients with thirteen disease labels. We also report the retrieval performance based on most prevalent diseases in the dataset [11], and the results are reported as CheXpert-5.

We trained the deep network model with a batch size of 32 for 20 epochs with Adam optimizer. The initial learning rate was 10^{-3} and the learning rate was reduced by a factor of 10 at plateau. Since the CheXpert dataset has uncertain labels for a lot of samples, while updating the gradient during training we ignored the uncertain labels . During the retrieval the uncertain labels were counted as positive labels. The retrieval performance is reported in Table. 1. ACG is defined as $ACG = \sum_i^K r_i/K$ where r_i is the graded relevance of the image retrieved at position i and is defined as the ratio of common labels between the retrieved image and the query image. ACG_p is calculated using the same formulation but using the common positive labels only. Precision is defined as the percentage of relevant images retrieved over the total number of retrieved images. The relevance of each retrieved image is calculated based on the common disease labels between the query image and the retrieved images. Table.2 reports the performance of the proposed method with two recent literature-based approaches in terms of average gain as reported in [5]. As can be seen from Table. 1, when all thirteen disease labels were used the framework was able to retrieve images where 73% of the retrieved images' disease labels matched exactly with the query image , and the percentage was 78% when the five most prevalent disease labels were used. However, there are disease labels which shows similar appearances on X-ray images (for example- pneumonia and consolidation). To report the performance here we matched the labels exactly, so if a pneumonia positive sample was used as a query image and a consolidation positive image was retrieved, we counted it as a non-relevant sample. In future we plan to analyse the retrieved images considering their appearances in the X-ray images. However the preliminary performance of the framework indicates its potential as a large-scale medical image retrieval tool.

4 Conclusion

In this work we present a large-scale medical image retrieval framework and reported the performance of the framework on the largest available chest X-ray image dataset. Although developed for chest X-ray image retrieval, the proposed approach can generally be applied to other medical image based retrieval tasks.

References

[1] Y. Anavi *et al.*, "A comparative study for chest radiograph image retrieval using binary texture and deep learning classification," in *37th EMBC*. IEEE, 2015, pp. 2940–2943.

[2] A. Shah *et al.*, "Deeply learnt hashing forests for content based image retrieval in prostate MR images," in *Medical Imaging*, vol. 9784. SPIE, 2016, p. 978414.

[3] X. Liu *et al.*, "Generating binary tags for fast medical image retrieval based on convolutional nets and Radon transform," in *IJCNN*. IEEE, 2016, pp. 2872–2878.

[4] S. Conjeti, A. G. Roy, A. Katouzian, and N. Navab, "Hashing with residual networks for image retrieval," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 541–549.

[5] Z. Chen, R. Cai, J. Lu, J. Feng, and J. Zhou, "Order-sensitive deep hashing for multimorbidity medical image retrieval," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 620–628.

[6] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li *et al.*, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 248–255.

[8] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol. 70, no. 6, p. 066111, 2004.

[9] N. F. Haq, M. Moradi, and Z. J. Wang, "Community structure detection from networks with weighted modularity," *Pattern Recognition Letters*, vol. 122, pp. 14–22, 2019.

[10] R. Lan, S. Zhong, Z. Liu, Z. Shi, and X. Luo, "A simple texture feature for retrieval of medical images," *Multimedia Tools and Applications*, vol. 77, no. 9, pp. 10 853–10 866, 2018.

[11] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," *arXiv preprint arXiv:1901.07031*, 2019.