An Efficient Radiographic Image Retrieval System Using Convolutional Neural Network

Manish Chowdhury*, Samuel Rota Bulò†, Rodrigo Moreno*, Malay Kumar Kundu‡ and Örjan Smedby*

* KTH, School of Technology and Health, Hälsovägen 11c, SE-141 57 Huddinge, Sweden
† FBK-irst, via Sommarive, 18, I-38123 Povo, Trento, Italy
‡ Machine Intelligence Unit, Indian Statistical Institute, Kolkata-108, India

Abstract—**Content-Based Medical Image Retrieval (CBMIR) is an important research field in the context of medical data management. In this paper we propose a novel CBMIR system for the automatic retrieval of radiographic images. Our approach employs a Convolutional Neural Network (CNN) to obtain high-level image representations that enable a coarse retrieval of images that are in correspondence to a query image. The retrieved set of images is refined via a non-parametric estimation of putative classes for the query image, which are used to filter out potential outliers in favour of more relevant images belonging to those classes. The refined set of images is finally re-ranked using Edge Histogram Descriptor, *i.e.* a low-level edge-based image descriptor that allows to capture finer similarities between the retrieved set of images and the query image. To improve the computational efficiency of the system, we employ dimensionality reduction via Principal Component Analysis (PCA). Experiments were carried out to evaluate the effectiveness of the proposed system on medical data from the "Image Retrieval in Medical Applications" (IRMA) benchmark database. The obtained results show the effectiveness of the proposed CBMIR system in the field of medical image retrieval.**

I. INTRODUCTION

Over the last few decades, medical imaging has become an increasingly active research area, resulting in a rapid development of new technologies and instrumentations, which play a pivotal role in a large number of healthcare applications. To support the physicians in their clinical diagnosis and treatments, an increasing number of medical image modalities are used. In large hospitals, several terabytes of digital medical images are generated and stored every year in so-called Picture Archiving and Communication Systems (PACS) [1], for diagnosis of diseases, research and education [2]. For precise diagnosis and treatment planning, medical professionals often have to browse through similar content images in these archives. This introduces the need of novel, intelligent techniques to efficiently search through large collections of medical images [3], [4].

An open challenge for classification/retrieval of medical images is the inter- versus intra-class variability problem [3]. Several Content-Based Medical Image Retrieval (CBMIR) prototypes have been proposed to address these problems using representative features for the images' content [4], [1], [2]. In general, the proposed systems can be categorized into two different classes: task/modality dependent and independent CBMIR systems. The CBMIR systems from the first category are developed based on a specific organ, imaging modality or diagnostic study [5], [6], [7]. These systems are ineffective for other medical applications. As for the second category, there have been few attempts to develop task/modality independent CBMIR systems: MedGIFT [8], KmED [9], and Image Retrieval in Medical Applications (IRMA) [10], [11] among many others.

Different CBMIR system that have been proposed employ classifiers such as Support-Vector Machines (SVM), Random Forests (RF), to coarsely classify the query image and restrict the retrieved set to images belonging to the same class. This reduces the overall cost of computation without impacting much on the final accuracy [12], [3], [13], provided that the classifier delivers fast predictions and the pre-classification accuracy is good enough. A drawback of this approach is that errors occurring at the pre-classification stage might severely jeopardize the whole retrieval process.

In this paper, we propose a novel CBMIR system for radiographic images that mitigates the drawbacks mentioned above. Our approach is based on two-stages, each employing a different type of feature descriptors to perform the retrieval task. In the first stage, a set of images from the (possibly) large collection of images is retrieved by comparing their high-level feature signature that is computed by means of a pre-trained Convolutional Neural Network (CNN). A relevance score is then computed to rank the classes represented in the retrieved set and identify candidate classes for the query image. This relevance score is computed by taking into account the position of the represented classes in the result set.

In the second stage, images belonging to the classes estimated in the previous stage are retrieved from the database, this time considering Edge Histogram Descriptor (EHD) as image descriptor in order to capture similarities at a finer granularity level. The proposed system is computationally efficient and the image collection can be easily enlarged since we do not rely on pre-trained classifiers, but we rather use a simple K-NN procedure for classification. Accordingly, no training procedure has to be run offline on the collection of images and new images can be added to the collection with limited effort. In order to improve the efficiency of the method, Principal Component Analysis (PCA) has been used in order to reduce the dimensionality of the feature space in both stages.

The rest of the paper is organized as follows: In Sec. II, we describe the image descriptors adopted in our system, namely CNN features for the first stage and EHD for the second one.

Sec. III describes the proposed approach to CBMIR. Sec. V provides an experimental evaluation of the proposed algorithm and, finally, we draw conclusions in Sec. VI.

II. IMAGE DESCRIPTORS

The performance of an image retrieval system inherently depends on the effectiveness of the feature descriptor representing the content of the image. Global descriptors can be used to coarsely determine the similarity of images, while local cues can be exploited for a finer-grained comparison. This is the principle that underlies our two-stage retrieval approach and turns out to be particularly effective for the retrieval of radiographic images. Besides the granularity of the descriptor another important aspect to take into account is the dimensionality of the representation, in particular if one aims to build an efficient CBMIR system. Compact feature representations can in general be constructed and evaluated in a more efficient way, thus being more appealing for real-world applications like radiographic image retrieval, where the collections to be searched are large [2], [4].

In this work we make use of CNN-based features as global image descriptors that will be employed in the first stage to perform a coarse retrieval, whereas EHD is used as local descriptor to obtain a finer-grained ranking of the retrieved images in the second stage. CNNs denote a family of feed-forward, neural networks that propagate the input signal through several computational blocks/layers, which compute convolutional features, apply non-linear activation functions, and perform pooling operations to reduce the dimensionality of the representation. CNNs have been shown to be very effective in many computer vision tasks in particular by employing very deep architectures [14]. The strength of CNNs derives from learning representations for the data at different levels of abstraction. For image data, the lower levels of abstraction might describe the differently orientated edges in the image; middle levels might describe parts of an object, while high layers refer to larger object parts and even the object classes.

Few recent studies can be found in the medical field that uses deep architecture methods [15], [16], [17]. Models trained on the non-medical dataset Imagenet have been successfully used for classification of medical images [15]. Using the same intuition, we use the Berkeley Caffe reference model imagenet-caffe-alex [18] as feature-extractor for radiographic images. We use the output of the "*fc7*" layer, *i.e.* the fully-connected layer before the output layer, as descriptor. The "*fc7*" layer has 4096 units (yielding a 4096-dimensional descriptor) and provides a high-level description of the image, which can be used for the first-stage, coarse retrieval task.

In the second, finer-grained retrieval stage of the proposed system, the X-ray images are summarized in terms of a popular descriptor namely Edge Histogram Descriptor (EHD) from the Moving Picture Experts Group-7 (MPEG-7). This descriptor captures the distribution of edges and provides a texture signature that is useful for image-to-image matching even when the underlying texture is not homogeneous [19]. EHD represents the distribution of local edge in an image by dividing the image into 4×4 sub-images and generates a histogram from the edges contained in each of these sub-images. Edges in the image are categorized into five types: vertical, horizontal, $45°$ diagonal, $135°$ diagonal and non-directional edges. Finally, a histogram with 16×5 bins is constructed, yielding a 80-dimensional descriptor.

Since the dimensions of the feature vectors of the two different stages (coarse retrieval stage: 4096 and fine retrieval stage: 80) are large, we applied PCA, to reduce the computational complexity by removing less relevant dimensions. In the rest of the paper, we refer the 4096-dimensional descriptor obtained from the CNN as *CNN-descriptor*, while we refer the other one as *EHD-descriptor*.

III. PROPOSED CBIR APPROACH

In this section we describe the proposed approach for content-based retrieval of radiographic images, which is organized into two stages.

A. First stage

Consider a database consisting of n labeled images $\mathcal{D} = \{(I_1, y_1), \ldots, (I_n, y_n)\}$, where $y_j \in \mathcal{Y}$ denotes the class labels (from a finite set $\mathcal{Y}$) of the jth image I_j, and let $f_j^{\text{CNN}} \in \mathbb{R}^{4096}$ be the CNN-descriptor of image I_j. Given a query image denoted by I_0, our retrieval system finds the most related images in the dataset by employing the CNN-descriptor as image representation and by using the Euclidean distance to compute the (dis)similarity to the query image. Specifically, let $\pi = (\pi_1, \ldots, \pi_m)$ be a m-permutation of n satisfying $d(f_0^{\text{CNN}}, f_{\pi_i}^{\text{CNN}}) \le d(f_0^{\text{CNN}}, f_{\pi_j}^{\text{CNN}})$ for all $1 \le i \le j \le m$ (obtained via sorting), where $d(\cdot, \cdot)$ denotes the Euclidean distance. Intuitively, π_j is the index of the image that is ranked at the jth position and the relation ensures that images ranked closer to position 1 have a larger similarity (*i.e.* smaller distance) to the query image. Moreover, m represents the size of the retrieved set of images (also referred to as scope size). From another point of view, the images indexed by π, which form the retrieved set, represent the m nearest neighbors in $\mathcal{D}$ of the query image.

Let $\mathcal{Y}_0 \subseteq \mathcal{Y}$ be the subset of image classes that are represented within the retrieved set of images for the given query image I_0. We compute for each $y \in \mathcal{Y}_0$ the *class relevance score* S_y^π with respect to π as the following function of the ranking position of images belonging to class y:

$$S_y^\pi = \sum_{j=1}^{m} \delta_{yy_{\pi_j}} \left(1 - \frac{j-1}{m}\right), \qquad (1)$$

where $\delta_{y\hat{y}}$ is the Kronecker delta yielding 1 if $y = \hat{y}$ and 0 otherwise, and y_{π_j} is the class of the image ranked in the jth position according to π.

The class relevance score described above is used to determine a subset of classes in $\mathcal{Y}_0$ that will be considered as relevant for the query image. Assume $k = |\mathcal{Y}_0|$ and let $(y_1, \ldots, y_k)$ be a permutation of the elements in $\mathcal{Y}_0$ satisfying $S_{y_i}^\pi \ge S_{y_j}^\pi$ for all $1 \le i \le j \le k$, *i.e.* classes in $y_j \in \mathcal{Y}_0$ having a better relevance score will have a lower index j. We

determine the relevant classes for the query image by finding the largest index $1 \leq j^* \leq k$ that satisfies $\sum_{i=1}^{j^*} S_i^\pi < \alpha$, where $\alpha = \sum_{i=0}^{\lfloor \frac{m}{2} \rfloor} \left(1 - \frac{i}{m}\right)$ is the hypothetical relevance score of a class that covers the first $\lfloor \frac{m}{2} \rfloor$ positions in the ranking. The relevant classes are finally given by $\mathcal{Y}_0^r = \bigcup_{i=1}^{j^*} \{y_i\}$ and will be used in the final stage to refine the set of retrieved image.

B. Second stage

The second stage of the retrieval process works on a subset of the original dataset $\mathcal{D}$, which comprises only images with a class belonging to the set of relevant classes $\mathcal{Y}_0^r$ described above. Let $\mathcal{I}^r$ be the indices of the images in $\mathcal{D}$ with a class label in $\mathcal{Y}_0^r$, i.e. $\mathcal{I}^r = \{j : y_j \in \mathcal{Y}_0^r, 1 \leq j \leq n\}$. Moreover, in this stage, we switch to the EHD-descriptor for the images in order to capture finer-grained similarities to the query image. Accordingly, we denote by f_j^{EHD} the EHD descriptor of image I_j, where $j \in \mathcal{I}^r \cup \{0\}$. Akin to the first stage, we find a m-permutation $\hat{\pi} = (\hat{\pi}_1, \ldots, \hat{\pi}_m)$ of $\mathcal{I}^r$ satisfying the relation $d(f_0^{\text{EHD}}, f_{\hat{\pi}_i}^{\text{EHD}}) \leq d(f_0^{\text{EHD}}, f_{\hat{\pi}_j}^{\text{EHD}})$ for all $1 \leq i \leq j \leq m$.[1] The final ordered set of retrieved images is given by $\mathcal{R} = (I_{\pi_j})_{j=1}^m$.

C. Dimensionality reduction

Instead of using the CNN- and EHD-descriptor in their original size, we employ PCA as an unsupervised dimensionality reduction technique. This provides us with two linear transformations T^{CNN} and T^{EHD} for the CNN- and EHD-descriptors, respectively. We can then replace f_j^{CNN} and f_j^{EHD} with their projected counterparts $\hat{f}_j^{\text{CNN}} = T^{\text{CNN}}(f_j^{\text{CNN}})$ and $\hat{f}_j^{\text{EHD}} = T^{\text{EHD}}(f_j^{\text{EHD}})$, respectively.

IV. RELATED WORKS

We provide here a brief overview of related works. In [20], the authors used 768 features for image representation and a K-NN classifier for classification. Here, each image is split in 16 equal sub-blocks and each sub-block is described using 48 features. To describe the image texture the authors have used Haralick's coefficients extracted from the gray-level co-occurrence matrix (16 features = 4 coefficients 4 directions), the box-counted fractal dimension (1 feature), and the Gabor wavelets (24 features = 2 coefficients 3 scales 4 orientations). In addition they used features derived from gray level statistical measures: different estimations of the first order (mean, median and mode), second order (variance and L2 norm), third and forth order (skewness and kurtosis) moments. In [21], the authors have used Non-Subsampled Contourlet Transform (NSCT) and Fuzzy C-Means (FCM) clustering to construct image descriptors, and used Least Square-Support Vector Machine (LS-SVM) and Earth Mover's Distance (EMD) to classify the images. Collins *et al.* have designed the CBMIR system by using the SIFT-PCA features and a L1-distance-based similarity measure [22]. Here, the

authors have used SIFT features, yielding a 128-dimensional histogram of local gradient directions for 8 orientations in 16 tiles. Additionally, they have included 4 extra parameters of the 2 spatial coordinates of the SIFT keypoint, the scale parameter, and the dominant-orientation parameter. Each SIFT feature vector is centered and normalized using Z-Score transform before applying k-means clustering. In [23], the authors have studied the performance of simple statistical features (mean, standard deviation, skew, energy and entropy) with the ED measure. Recently, Camlica *et al.* have proposed a CBMIR system with Local Binary Pattern (LBP) and SVM [24].

In all these state-of-the-art approaches, only handcrafted features have been used for image representation, whereas our approach includes both handcrafted (EHD) and learned (CNN) features and is able to significantly improve the results over the competitors as we will show in the next section.

V. EXPERIMENTAL RESULTS

In this section we provide experimental quantitative and qualitative evidence of the validity of the proposed CBMIR system for radiographic images.

A. Setup

We evaluated our approach on the medical image collection from the ImageCLEF2009 medical benchmark [3]. The X-ray images are noisy with irregular brightness and contrast, and sometimes contain dominant visual artifacts such as artificial limbs and X-ray frame borders. The entire dataset contains $14,410$ images divided into 193 distinct categories, where the number of images in each category varies largely (from 1 to 2314 images). We have restricted the analysis to the 31 categories having at least 50 images and we randomly selected 50 images for each category, yielding a dataset of 1550 images organized into 31 classes. The images of this database suffer from the inter-class versus intra-class variability problem. Images from this dataset are organized into semantic categories in a way to reflect the human perception of image similarity. The performance on the dataset was evaluated by considering 10 random images from each class as the query image (in total 310 queries) and by measuring the average precision and recall,[2] or the average accuracy (which coincides with the average precision), depending on the experiment. The experiments were conducted on a Dell Precision T7810 PC with 32GB RAM and the proposed approach has been implemented using MATLAB R2016a. All experiments were run with scope size $m = 20$, where not differently stated.

B. Quantitative analysis

In order to assess the effectiveness of the proposed combination of high-level CNN-based features for the coarse retrieval of images in the first stage and low-level EHD-based features for the finer-grained refinement in the second stage (we refer this setting as CNN→EHD), we report the results also for the following variants of our algorithm: only the first stage is run

---

[1] we implicitly assume $m \geq |\mathcal{I}^r|$, but if this is not the case one could either shrink the scope size in the second stage, or augment the resulting set with the best ranked images in the first stage, with a class label not in $\mathcal{Y}_0^r$.

[2] precision = $\frac{\text{\# of retrieved relevant images}}{\text{scope size}}$, recall = $\frac{\text{\# of retrieved non-relevant images}}{\text{\# of relevant images in the database}}$.
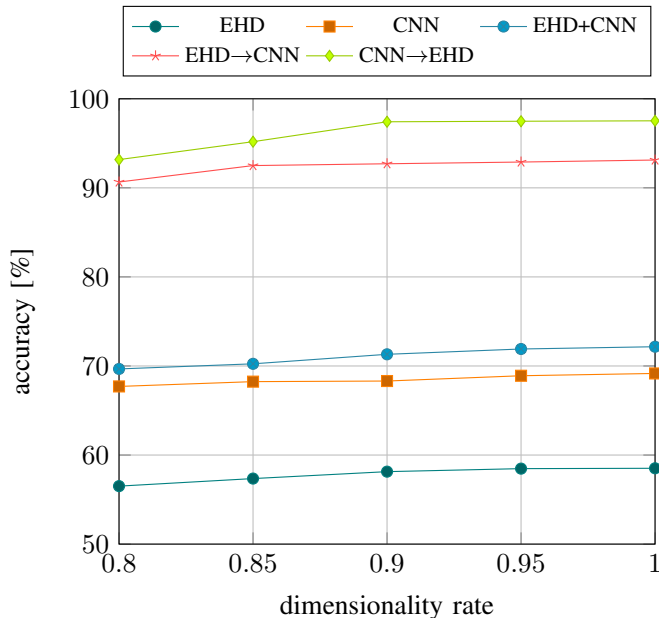
Fig. 1. Experimental results obtained in terms of accuracy on a benchmark dataset by different variants of the proposed method (see Sec. V for details) using different levels of dimensionality reduction. The dimensionality rate indicates the share of dimensions that are preserved (*i.e.* 1 means that we preserve the original dimensionality). The results are averaged over 10 random queries per class.

with the EHD-descriptor (EHD), only the first stage is run with the CNN-descriptor (CNN), only the first stage is run by concatenating the EHD- and CNN-descriptors (EHD+CNN), the first stage is run using the EHD-descriptor and the second one using the CNN-descriptor (EHD→CNN).

A first experiment evaluates the effect of the dimensionality reduction procedure on the performance of the system at different compression levels. Fig. 1 shows the accuracy obtained with the different variants of our algorithm mentioned in the previous paragraph, at different dimensionality rates (0.8, 0.85, 0.9, 0.95 and 1). The dimensionality rate is 1 when no dimensionality reduction occurs and diminishes as the reduction increases. The experiment shows that there is little impact on the accuracy of the different tested approaches if we apply the dimensionality reduction. In particular we can safely have a 10% reduction of the dimensionality of the descriptors. If we compare the single approaches we tested, we can see that the method we propose (CNN→EHD), which uses CNN-descriptors in the first stage and EHD-descriptors in the second stage, is the most effective, because the image representations that we obtain from the convolutional neural network better discriminate the classes of images in the dataset, even though no specific training has been conducted on the radiographic images. This renders the class-based selection of images in the second stage more effective. If we invert the role of the EHD- and CNN- descriptors (*i.e.* EHD→CNN) we experience a noticeable drop in the performance, because the class discrimination in the first stage becomes less reliable. We can also see that if we skip the second stage the drop

becomes significantly larger, which is expected since no class information is used.

In Table. I we compare the results obtained by our method with 10% dimensionality reduction against other state-of-the-art approaches. We implemented in Matlab the methods proposed in [20], [21], [22], [23], [24] and applied them to the same dataset. As shown, the proposed method performs significantly better than the competitors. This is probably due to the use of high-level, data-driven CNN-descriptors in the first stage, while the competitors rely mainly on hand-crafted mid- or low-level features.

TABLE I
COMPARISON OF THE ACCURACY OF THE PROPOSED METHOD AGAINST OTHER EXISTING CBMIR SYSTEMS ON A BENCHMARK DATABASE.

Mechanism	Accuracy
Florea *et al.*, 2006 [20]	73.21%
Chowdhury *at al.*, 2012 [21]	76.00%
Collins *et al.*, 2013 [22]	70.12%
Ayyachamy *et al.*, 2013 [23]	60.16%
Camlica *et al.*, 2015 [24]	82.01%
Proposed Approach	**97.79%**

In Fig. 2 we report the precision/recall plots of the different variant's of our method, where we vary the scope size ($m = 10, 20, 30, 40$). The results are consistent with the ones showed on Fig. 1, *i.e.* the proposed approach is the most effective also at different scope sizes. Note that the use of small scope sizes leads to lower recall because it becomes more difficult to retrieve all the relevant images when their number exceeds the scope size. In particular, since we have 50 potential relevant images the maximum recall that we can expect with a scope size of *e.g.* 20 is 0.4.

C. Qualitative analysis

Some qualitative results are shown in Fig. 3. Specifically, we report the outcome of 4 queries (denoted with letters a,b,c,d) in both stages (identified by the number 1 and 2). For each block of images, the isolated, top-left image is the query image and the green/red boxes indicate relevant/irrelevant images, respectively.

In Fig. 3(a1) we give the set of retrieved images after the first stage of our algorithm from a query image belonging to the "hip" category. As can be seen, 8 out of 20 images are correctly retrieved and have an appearance that is similar to the query image. Moreover, most of the truly relevant images have a large relevance score, since they cover the first ranking positions. Accordingly, one can filter out the misclassified images at the second stage that do not belong to the relevant classes in $\mathcal{Y}_0^r$. Indeed, after applying the second stage of the algorithm as discussed in the previous section the quality of the retrieved set improves significantly, as shown in Fig. 3(a2). There are nevertheless still two wrong retrieved images, whose class was regarded as relevant after the first stage (*i.e.* their class was in $\mathcal{Y}_0^r$), having an EHD-descriptor similar to the query image.
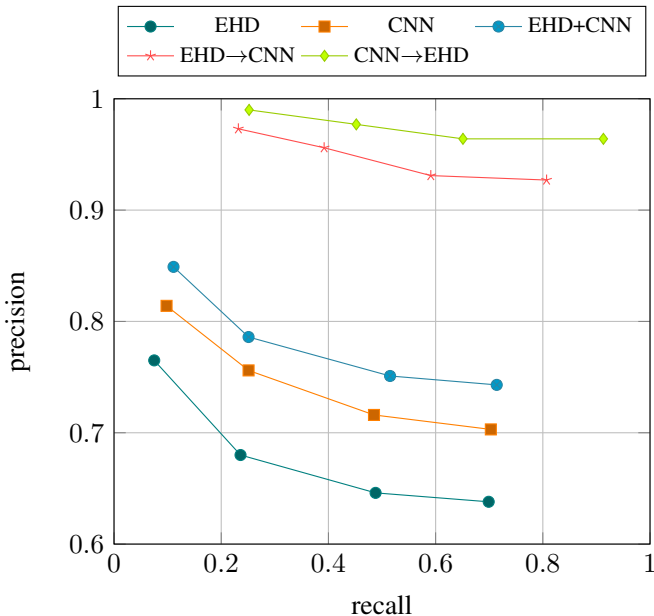
Fig. 2. Precision/recall curves obtained by different variants of the proposed approach (see Sec. V for details) on a benchmark dataset, by varying the scope size $(10, 20, 30, 40$ left-to-right). The results are averaged over 10 random queries per class.

Another retrieval example is given in Fig. 3(b1). Here, 11 out of 20 images are correctly retrieved from the query image (top-left) after the first stage, while 9 images are from different classes. Again, the second stage allows to filter out the outliers, thus improving the quality of the retrieved set, and obtains 100% accuracy, as shown in Fig. 3(b2).

An additional example of successfull retrieval is given in Fig. 3(c1,c2) with a query image from the "foot" category. Here, 16 out of 20 images are from the right category, while 4 images are wrong, but again the second stage pushes the final accuracy to 100%.

Finally, Fig. 3(d1,d2) show a case where our method fails. Although the query image is from the category "hand", the "foot" category is deemed as the most relevant class in the first stage, jeopardizing the final result in the second stage. Notice that some images from the correct class that were retrieved in the first stage (see Fig. 3(d1)) have been lost in the subsequent stage due to the wrong estimation of the class of the query image (see Fig. 3(d2)).

VI. DISCUSSION

In this paper, we have proposed an efficient CBMIR system for radiographic images based on a novel two-stage retrieval approach. The aim of the first stage is to provide a coarse-grained retrieval of the images mainly aimed at identifying a set of putative classes to which the query image belongs to. For this reason, we employ high-level CNN-based features. The estimated set of relevant classes is used in the second stage to filter our potential outliers and focus the retrieval only on those images having a class label that was considered relevant in the previous stage. In the second stage, the restricted set

of images is ranked based on low-level EHD features, which allow for a finer-grained evaluation of the similarity to the query image. The experiments we conducted have shown the effectiveness of the proposed approach.

It is worth to point out that we extracted CNN features from a network that has been pre-trained on a standard object recognition dataset (ImageNet) instead of training an ad-hoc network on radiological images. The reason for this decision is that this step requires a very large amount of medical data that is currently not available. A similar procedure was also applied by Greenspan *et al.* [15], [25] with good results. This inspired our approach that uses CNN-based features for image representation.

Ongoing research includes the use of our system on different types of medical images, such as endoscopic, colonoscopy and ultrasound images. We are also interested in incorporating a human interaction mechanism to make the system more robust. Furthermore, we will test the system on databases of images with annotated diagnostics for assessing its potential for performing computer aided diagnosis.

REFERENCES

[1] H. Greenspan and A. T. Pinhas, "Medical image categorization and retrieval for PACS using the GMM-KL framework," *IEEE T Inf. Technol. Biomed.*, vol. 11, no. 2, pp. 190–202, 2007.

[2] A. Kumar, J. Kim, W. Cai, M. Fulham, and D. Feng, "Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data," *J Digit Imaging*, vol. 26, no. 6, pp. 1025–1039, 2013.

[3] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level using patch-based visual words," *IEEE T Med. Imag.*, vol. 30, no. 3, pp. 733–746, 2011.

[4] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: Current status and future directions," *J Digit Imaging*, vol. 2, no. 24, pp. 208–222, 2011.

[5] D.-M. Kwak, B.-S. Kim, O.-K. Yoon, C.-H. Park, J.-U. Won, and K.-H. Park, "Content-based ultrasound image retrieval using a coarse to fine approach," *Ann. NY Acad. Sci.*, vol. 980, pp. 212–224, 2002.

[6] W. Cai, D. D. Feng, and R. Fulton, "Content based retrieval of dynamic PET functional images," *IEEE T Inf. Technol. Biomed.*, vol. 4, no. 2, pp. 152–158, 2000.

[7] L. Zheng, A. W. Wetzel, J. Gilbertson, and M. J. Becich, "Design and analysis of a content-based pathology image retrieval system," *IEEE T Inf. Technol. Biomed.*, vol. 7, no. 4, pp. 249–255, 2003.

[8] H. Muller, A. Rosser, J. P. Vallee, and A. Geissbuhler, "Comparing feature sets for content-based medical information retrieval," in *Proc. SPIE Med. Img.*, vol. 5351, 2004, pp. 99–109.

[9] W. W. Chu, C. C. Hsu, A. F. Cardenas, and R. K. Taira, "Knowledge-based image retrieval with spatial and temporal constructs," *IEEE T Knowl. Data Eng.*, vol. 10, no. 6, pp. 872–888, 1998.
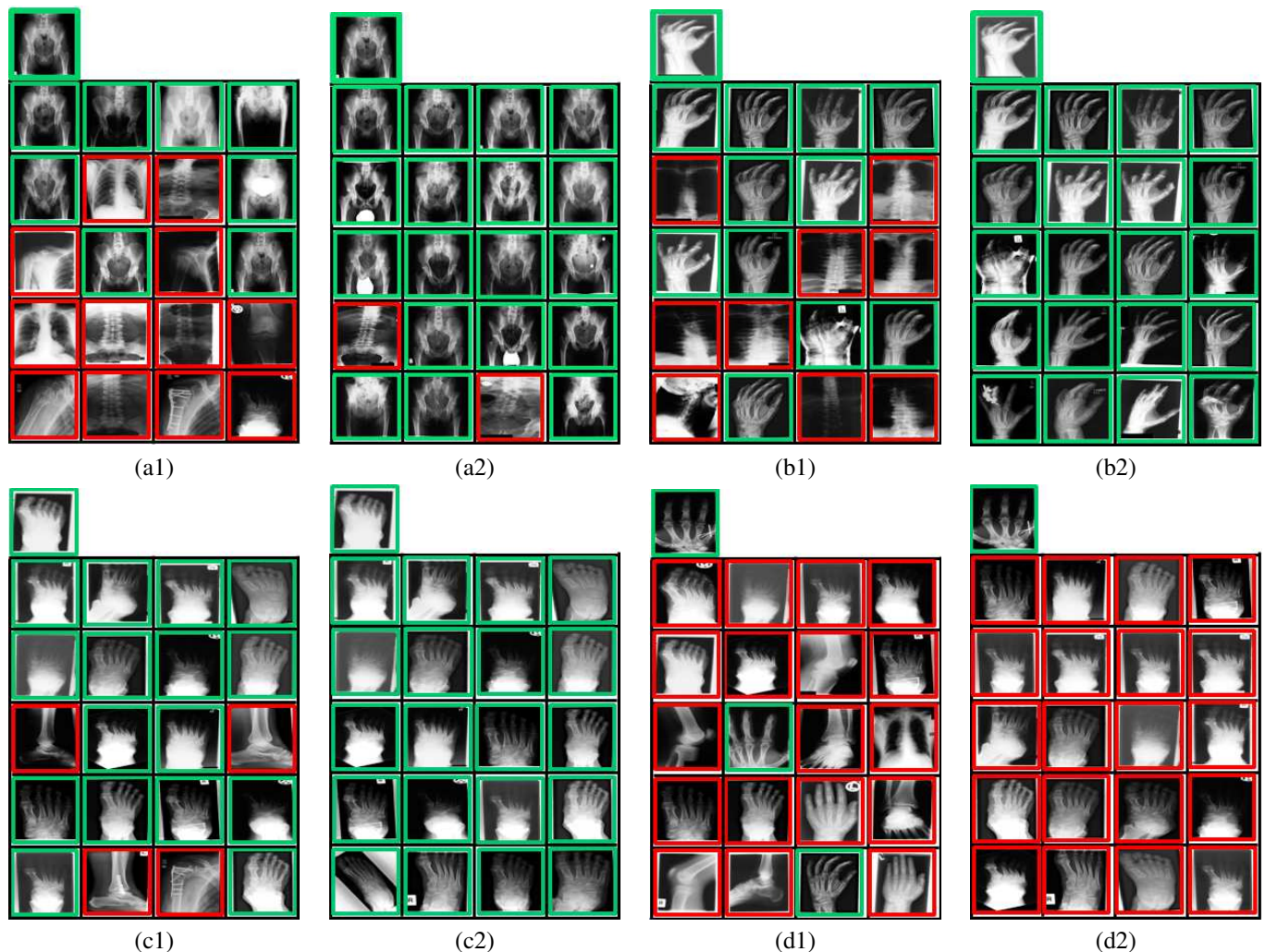
Fig. 3. Examples of retrieval results (a,b,c,d). The isolated, top-left image in each block of images represents the query image. For each query we report the result after the first stage (indexed by 1) and the second stage (indexed by 2). All queries excepting the last one (d), show successful queries. More details are given in Sec. V.

[10] T. M. Lehmann, M. O. Guld, C. Thies, B. Fischer, D. Keysers, and M. Kohnen, "Content-based image retrieval in medical applications for picture archiving and communication systems," in *Proc. SPIE Med. Imag.*, vol. 5033, 2003, pp. 109–117.

[11] E. Ahn *et al.*, "X-ray image classification using domain transferred convolutional neural networks and local sparse spatial pyramid," in *Proc. of ISBI*, 2016, pp. 855–858.

[12] B. C. Ko, S. H. Kim, and J.-Y. Nam, "X-ray image classification using random forests with local wavelet-based cs-local binary patterns," *J Digit Imaging*, vol. 24, no. 6, pp. 1141–1151, 2011.

[13] M. Rahman, S. K. Antani, and G. R. Thoma, "A learning-based similarity fusion and filtering approach for biomedical image retrieval using SVM classification and relevance feedback," *IEEE T Inf. Technol. Biomed.*, vol. 15, no. 4, pp. 640–646, 2011.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[15] Y. Bar *et al.*, "Chest pathology detection using deep learning with non-medical training," in *Proc. of ISBI*, 2015, pp. 294–297.

[16] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, K. Cha, and M. A. Helvie, "Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis," vol. 9785, 2016, pp. 97 850Y–97 850Y–7.

[17] J. Cho, K. Lee, E. Shin, G. Choy, and S. Do, "How much data is needed to train a medical image deep learning system to achieve necessary high accuracy?" *arXiv:1511.06348v2*, 2016.

[18] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. of MM*, 2014, pp. 675–678.

[19] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE T Circuits Syst. Video Tech.*, vol. 11, no. 6, pp. 1996–2002, 2001.

[20] F. Florea, E. Barbu, A. Rogozan, and A. Bensrhair, "Using texture-based symbolic features for medical image representation," in *Proc. of ICPR*, vol. 2, 2006, pp. 946–949.

[21] M. Chowdhury, S. Das, and M. K. Kundu, "Effective classification of radiographic medical images using ls-svm and nsct based retrieval system," in *Proc. of 5th Int. Conf. on Computers And Devices for Communication (CODEC-2012)*, 2012.

[22] J. Collins and K. Okada, "Learning metrics for content-based medical image retrieval," in *Proc. of EMBC*, 2013, pp. 3363–3366.

[23] S. Ayyachamy and V. S. Manivannan, "Distance measures for medical image retrieval," *Int. J Imag. Syst. Tech.*, vol. 23, no. 1, pp. 9–21, 2013.

[24] Z. Camlica, H. R. Tizhoosh, and F. Khalvati, "Medical image classification via SVM using LBP features from saliency-based folded data," in *Proc. of ICMLA*, 2015, pp. 128–132.

[25] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "A comparative study for chest radiograph image retrieval using binary texture and deep learning classification," in *Proc. of EMBC*, 2015, pp. 2940–2943.