Journal of Medical Systems (2018) 42:24
https://doi.org/10.1007/s10916-017-0875-4

IMAGE & SIGNAL PROCESSING

Medical Image Retrieval with Compact Binary Codes Generated in Frequency Domain Using Highly Reactive Convolutional Features

Jamil Ahmad[1] · Khan Muhammad[1] · Sung Wook Baik[1]

Received: 21 August 2017 / Accepted: 22 November 2017

Abstract
Efficient retrieval of relevant medical cases using semantically similar medical images from large scale repositories can assist medical experts in timely decision making and diagnosis. However, the ever-increasing volume of images hinder performance of image retrieval systems. Recently, features from deep convolutional neural networks (CNN) have yielded state-of-the-art performance in image retrieval. Further, locality sensitive hashing based approaches have become popular for their ability to allow efficient retrieval in large scale datasets. In this paper, we present a highly efficient method to compress selective convolutional features into sequence of bits using Fast Fourier Transform (FFT). Firstly, highly reactive convolutional feature maps from a pre-trained CNN are identified for medical images based on their neuronal responses using optimal subset selection algorithm. Then, layer-wise global mean activations of the selected feature maps are transformed into compact binary codes using binarization of its Fourier spectrum. The acquired hash codes are highly discriminative and can be obtained efficiently from the original feature vectors without any training. The proposed framework has been evaluated on two large datasets of radiology and endoscopy images. Experimental evaluations reveal that the proposed method significantly outperforms other features extraction and hashing schemes in both effectiveness and efficiency.

Keywords Image retrieval · Convolutional neural network · Hash codes · Fourier transform · Feature selection

Introduction

Medical image repositories have always been a precious resource for medical experts to review previous cases for suggesting diagnosis to similar cases, or teach medical students about certain cases of interest [1, 2]. Efficient retrieval of relevant information is an essential task of content-based medical image retrieval (CBMIR) systems [3]. Typical CBMIR systems have to deal with huge amounts of data even in a local scale database. These data repositories not only grow in volume but also in complexity over time, which pose several challenges to the current CBMIR systems such as efficient indexing and accurate retrieval of semantically relevant contents from huge datasets.

Pair-wise image matching is the core element of any CBMIR system where a query image is matched with an image in the dataset to determine similarity between them [4, 5]. Traditional systems relied on low-level features extraction from medical images by analyzing their colors, textures, shapes, and spatial layout. Some of the most widely used features extraction approaches include bag-of-visual-words (BoVW) [6, 7] based on scale invariant features transform (SIFT) [8] keypoints, histograms of oriented gradients (HoG) [9], vectors of locally aggregated descriptors (VLAD) [10], fisher vectors (FV) [11, 12], GIST [13] or CENTRIST [14]. Each of these features extraction algorithms analyze contents at local scales and attempt to represent visual contents in such a way that pair-wise image matching can be achieved through simple distance measures between their respective feature vectors. However, low-level features often fail to effectively represent semantic concepts in images and hence retrieval using these features frequently lead to unsatisfactory results, particularly in large and complex datasets [15].

In recent years, deep learning based methods have achieved state-of-the-art performance in a variety of computer vision tasks including, image classification

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Sung Wook Baik
sbaik@sejong.ac.kr

[1] Digital Contents Research Institute, Sejong University, Seoul, Republic of Korea

Published online: 19 December 2017

[15–17], object detection, image segmentation [18, 19], and image retrieval [20–24]. The major advantage of these approaches is their ability to automatically learn features from raw data [25]. Convolutional neural networks are very powerful deep learning architectures which can automatically learn discriminative features from raw data and have been successfully applied to image retrieval scenarios [21, 26–28]. It consists of several layers which perform abstraction of the raw data at various scales. It has been found that activations from various convolution and fully connected (FC) layers can serve as features for effectively representing images. However, these high dimensional features become inefficient in retrieval from very large datasets. To overcome this issue, the high dimensional features are often converted to compact binary codes which serve as hash codes. These hash codes with locality sensitivity property can significantly reduce the search space, allowing very efficient retrieval using approximate nearest neighbor (ANN) search techniques in large datasets. Several methods have been proposed for learning hash codes which can be categorized into two groups: 1) learning-based methods and 2) data-independent approaches. Each of these approaches either use supervised or unsupervised training methods to learn and project high dimensional features to low-dimensional binary hash codes while preserving locality sensitivity.

In this paper, we present an efficient algorithm to select optimal convolutional features for medical image representation. We also introduced a novel and computationally efficient approach to transform these features into compact binary codes. The optimal subset of features discriminatively represent medical images eliminating unused convolution features, thereby improving features extraction efficiency. Further, the binary codes constructed in the Fourier feature space serve as hash codes for efficient retrieval. The proposed method is evaluated on large datasets of medical radiographs and endoscopy images. Major contributions in this work are as follows:

a) We present an efficient method to select optimal subsets of convolutional features based on their neuronal responses for representing medical images using pre-trained CNNs
b) A novel method is introduced to transform convolutional features to compact binary codes using Fourier decomposition which can serve as hash codes for efficient retrieval in large datasets. The proposed method do not require any training and can be directly applied for transforming features to short binary codes.

The rest of the paper is organized as: Section 2 presents related work in the field of medical image retrieval especially utilizing CNN features and hash codes. The proposed method is illustrated in Section 3. Experimental results are discussed in Section 4. The paper is concluded in Section 5 with references to future research directions.

Literature Review

In recent years, neuronal activation features from deep CNNs have shown tremendous success in computer vision applications. Particularly in image classification and retrieval, features from the last convolution layer or the FC layers have been extensively investigated and used [29, 30]. Features from the deep convolutional layers have recently witnessed more favor due to their ease of use and natural interpretation. These features preserve their spatial layout in the image, and each value correspond to the neuronal activation in local receptive field. Several pooling approaches have been proposed to derive a global representation of the accumulated local features. For instance, Azizpour et al. [30] used max pooling of the final convolutional layer for image retrieval and achieved good performance. Babenko et al. [21] introduced sum-pooled convolutional (SPoC) features where they accumulated sum of activation values per feature map to derive an effective representation for images. Their technique did not use computationally complex high-dimensional embedding of convolutional feature vectors and performed remarkably well in image retrieval. They also showed that sum pooling achieved better performance than max pooling. In another work, Gong et al. [26] used VLAD encoding on deep features extracted from small patches of the input image at multiple scales. Individual encodings are pooled in an orderless manner to construct a generic representation. It achieved state-of-the-art performance on challenging dataset, however, their method was computationally expensive which is not favored for large scale datasets. In other similar work, Mohedano et al. [31] used the bag-of-words model to aggregate local convolutional feature vectors to represent images using inverted indices. All these methods attempt to aggregate local deep features to construct an effective global representation.

In the field of medical image retrieval, deep features have also been used. Srinivas et al. [32] presented a dictionary learning based medical image retrieval framework. Medical images were grouped into predefined categories using sparse representations. Query image was compared with each image cluster to determine relevant images. In [33], the authors used kernels from the first convolutional layer of a pre-trained CNN and clustered them on the basis of their sensitivity to colors and textures, and then used them to extract features for classifying medical images. Their results indicate that such

primitive features can be used to represent medical images like endoscopy due to the presence of low-level features. In a similar approach, Zhang et al. [15] first decomposed medical images into multiple scales and then non-negative sparse coding with fisher discriminative analysis was used for representation [34]. The multi-scale images helped their method to capture diverse visual details which eventually lead to improved classification performance. Ahmad et al. [3] introduced saliency-injected neural code (SiNC) approach where the authors integrated deep features from the FC layers of a fine-tuned CNN for the whole image and salient component in the image. They used a weighted aggregation scheme to combine both sets of features in order to derive a representation for medical images, focusing on visual attention. Further, they also showed that SiNC features can be conveniently transformed to short binary codes for efficient retrieval of radiographs from large datasets.

Hash based image retrieval approaches have been extensively investigated for large scale datasets. Numerous approaches have been proposed in recent years which can be categorized as data independent methods and learning-based methods. Data independent methods include locality sensitive hashing (LSH) [35], spectral hashing (SH) [36], spherical hashing (SpH) [37], Kernelized LSH (KLSH) [38], principal component analysis based hashing (PCAH) [39], PCA with random rotation (PCA-RR) and Iterative Quantization (ITQ) [40], Density sensitive hashing (DSH) [39], Circulant binary embedding with optimization (CBE-opt) [41] and Compact Quantization (CQ) [42] etc. LSH uses random projections for generating hash tables, requiring a lot of memory for deriving reasonable codes. DSH improves LSH by considering geometrical structure of the data to avoid purely random projections. KLSH utilizes kernels to learn data projections, making it a highly computational expensive scheme. Some of these methods like SH, SpH, CBE-opt requires a lot of time to train or transform features to binary codes. Others like PCA-H and KLSH yield hash codes with low retrieval performance. Deep learning-based methods include Deep Hashing (DH) [43], simultaneous feature learning and hashing [44], and deep semantic ranking based hashing [45], to name a few. These methods have recently been developed to utilize the learning ability of these algorithms to learn data projections. However, these methods require a lot of data, time and computational resources to train, which may not always be available. Each of these approaches attempt to transform high dimensional feature vectors to low-dimensional binary representations. Feature vectors of relevant images are placed near to each other in the low-dimensional hamming space in order to facilitate ANN search approaches. Though these methods perform effectively in a variety of image retrieval scenarios, some of these methods require computationally expensive training procedures. Furthermore, they may require heavy computations for transforming features to binary codes. There is need to devise efficient ways to derive hash codes from high-dimensional deep convolutional features.

Materials and Methods

Feature learning followed by their transformation to compact hash codes has been the focus of intense research in the information retrieval community in recent years. Both of these tasks are highly essential for realizing efficient access to visual data in large image repositories. The proposed framework involves two major modules as shown in Fig. 1. The first module selects optimal subset of convolutional feature maps from a pre-trained CNN by analyzing their neuronal responses on the target images. Global mean activations from the selected feature maps form the representational feature vector. The second module then transforms that feature vector into compact binary code, which can serve as hash codes. Details of both modules are provided in the subsequent sections.

Optimal Feature Selection

Selection of appropriate features is an essential task in any image representation scheme. It allows reduction in feature dimensions by discarding irrelevant features, and improves features extraction efficiency. In a pre-trained CNN, convolutional feature maps from a deep convolutional layer are capable of representing a huge variety of objects, based on the type of data it is trained on (e.g. ImageNet [46]). In such a case, utilizing all the features for a particular class of images (e.g. medical images) may not be necessary. Therefore, feature selection techniques can be used to determine optimal subset of feature maps to adequately represent medical images. For this purpose, we present an optimal feature selection algorithm as listed in Algorithm 1. A set of training images are forward propagated through the pre-trained CNN (VGG-16 [47]) to obtain convolutional activations from "pool6" layer. The obtained tensor contains $6 \times 6 \times 512$ activations. We then compute global mean of each of the 512 feature maps to obtain 512-d feature vector. After that, we construct a null utilization index (NUI) to store indices of those feature maps which generated zero activation values for the training images as shown in Fig. 2. Null activation are highlighted in the NUI map with yellow points. Yellow colored columns in the NUI correspond to those feature maps which consistently

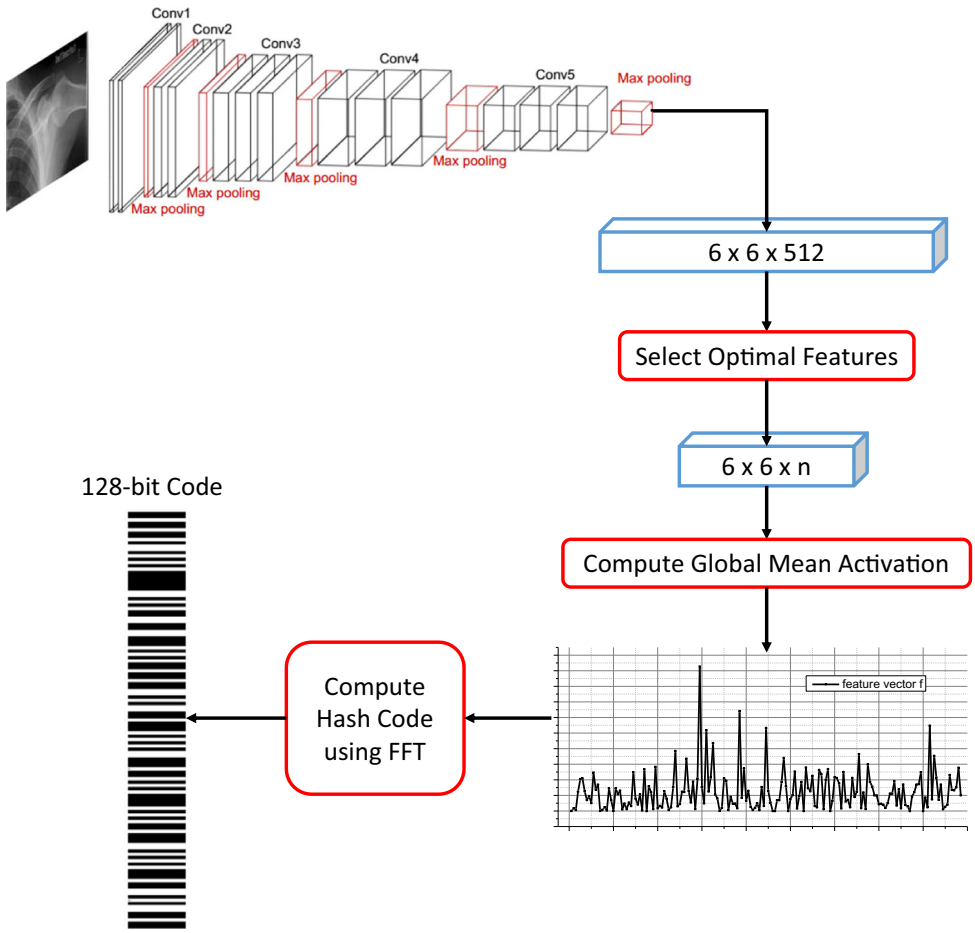

Fig. 1 Proposed features extraction framework

generate null activations for the target medical images (e.g. radiographs). We consider these feature maps as irrelevant and eliminate them from the features extraction phase. Percentage frequencies are computed for each feature maps from the NUI map and those features whose utilization is less than a threshold t are

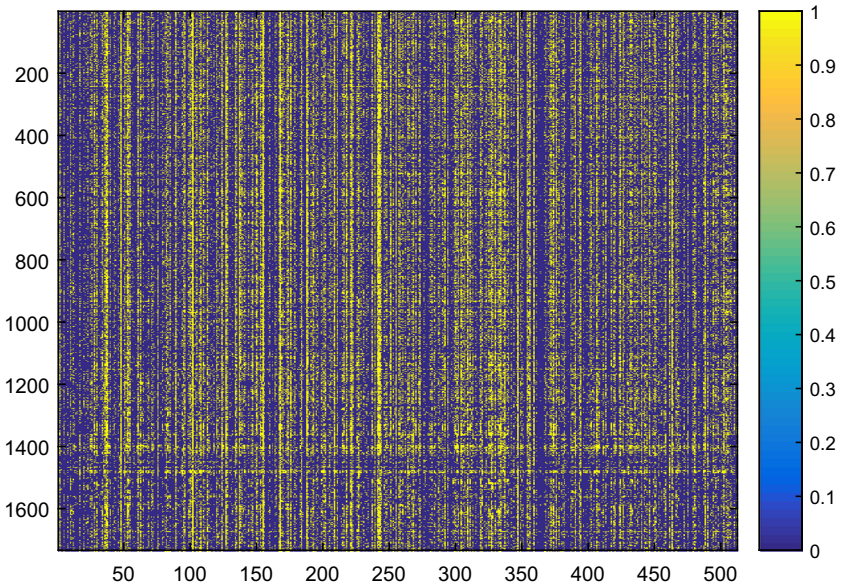

Fig. 2 Null utilization index (NUI) map

removed. Activation maps of the selected features are shown in Fig. 3 where the stronger activations on parts of the medical images can be seen. Sub-figures (a) and (b) contain responses of 16 different feature maps on similar images, whereas, (c) and (d) show activations of the same feature maps on different images. On similar images, the activation consistency can be seen in almost all of the feature maps, where stronger activations are generated on similar parts in both images. At the top of each feature map, the mean activation value is shown. To

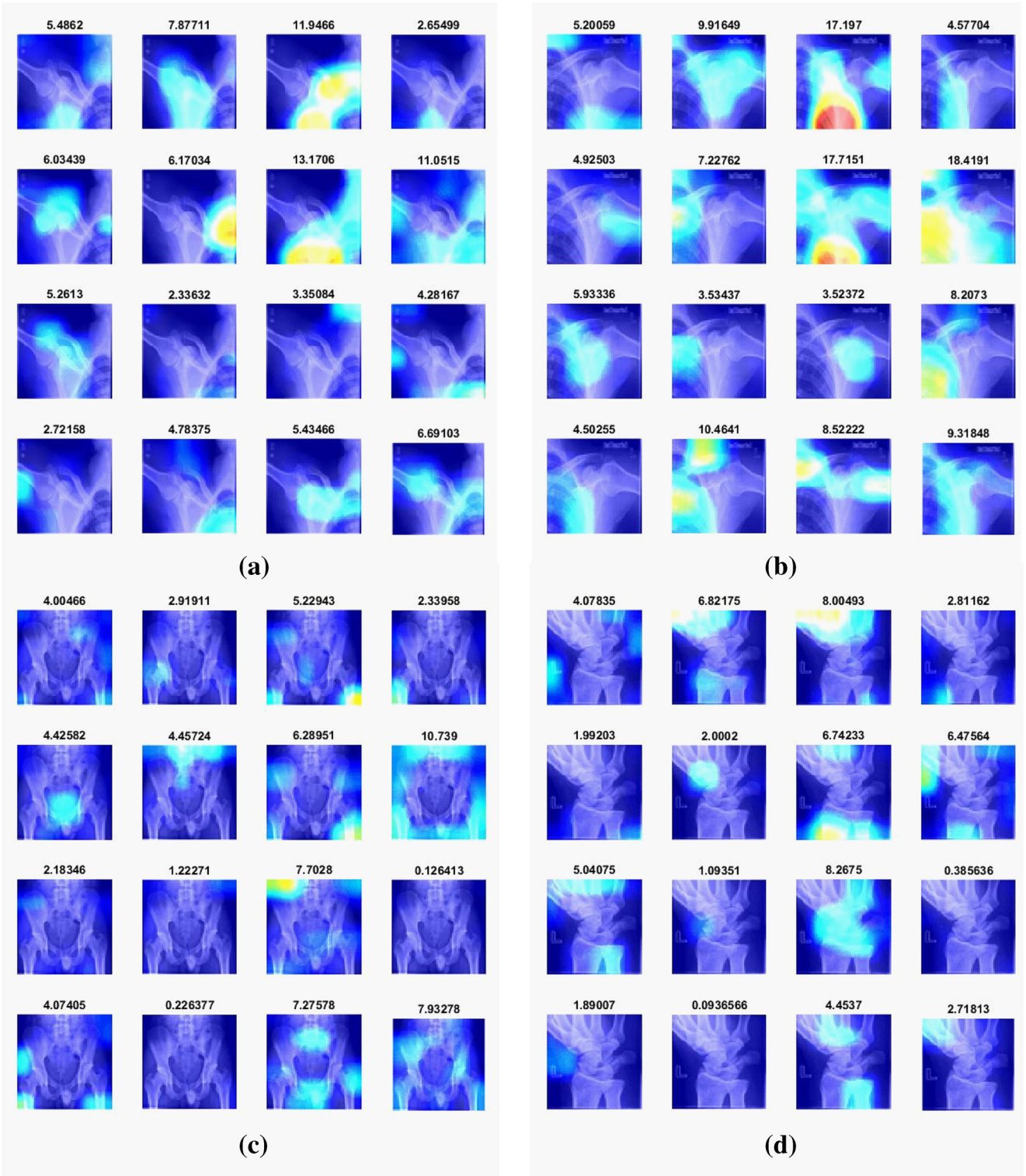

Fig. 3 Activations in the selected feature maps

derive the feature vector from the selected maps, we compute their global mean. For various values of t, we obtained several subsets of features. Optimal subset was selected after experimental evaluation.

Algorithm 1: Optimal Feature Selection

Input:

Set of training images (TS)

Output

Optimal feature Set (F_S)

Preparation:

1. Load the pre-trained CNN
2. Initialize Null Utilization Index (NUI) map with size $TS \times F_N$ to 0.

Steps:

1. **for each** image TSi **in** TS
   a. Forward propagate TSi through CNN
   b. Extract $h \times w \times F_N$ feature maps from "pool6"
   c. Compute mean of each feature map Fi to obtain F_N values
   d. Mark feature maps with 1 whose $Fmi = 0$ in the NUI for TSi.

   end for
2. Compute frequencies of null activations F_NA_i for each Fi
3. Compute frequency percentages P_F_NAi for each Fi
4. Return all Fi as Fs whose $P\_F\_NAi < t$

Generation of Compact Binary Codes using Fourier Decomposition

Activations from the deep convolution layers or the FC layers often yield high dimensional feature vectors. Usually these features are directly used to compute distances between comparing images, which yield state-of-the-art performance in many cases. However, direct feature matching becomes infeasible in case of very large datasets. Hence, more efficient methods are required such as locality sensitive hashing. These techniques attempt to derive compact binary codes from the high-dimensional feature vectors in such a way that the distance between the actual features correlates with the distance between the binary codes. In other words, sets of images which lie close to each other in the high dimensional feature space should also remain close to each other in the low-dimensional hamming space. This characteristic (known as locality sensitivity) allows direct access to the particular location in the feature space where potentially relevant images may exist. It significantly reduces the search space and ensures efficient access to data in large datasets by effectively avoiding exhaustive searching. Several hashing methods exist in literature, which yield state-of-the-art performance. In this paper, we introduced an efficient method to convert the selected features to short sequences of binary digits.

Given an input image I, we computed the convolutional features from the selected maps to obtain n-dimensional feature vector f. We treat the selected feature vector f as a one-dimensional signal (as shown in Fig. 4a and compute its Fast Fourier Transform (FFT) using (1) to obtain the frequency domain representation of f. The resulting representation is complex in nature whose real components are acquired using (2). Consequently, both positive and negative frequency components are obtained. We then normalize these values using the DC component of the spectrum. The Fourier transform of the input feature vector f is computed as:

$$FT_k = \sum_{j=0}^{n-1} f_j e^{-i2\pi kj/n},\ k = 0, \ldots, n-1 \tag{1}$$

where FT is the Fourier transform of the feature vector f, n represents the dimensions of f, and k represents the frequency components of the transform. The real part of FT is obtained as:

$$FT_k^R = Real(FT_k) \tag{2}$$

Once the real part of the transform is obtained, it is normalized by the DC component of the signal using the following:

$$FN_k^R = \frac{1}{FT_0^R}\left(FT_k^R\right) \tag{3}$$

where FN_k^R is the normalized FT of f and FT_0^R represents the DC component of the signal. After this normalization, we obtain the normalized frequencies in the FT where both positive and negative components can be seen as shown in Fig. 4b. In order to derive binary representation, we transformed the positive frequencies to bits with values 1 and the rest with values 0 as:

$$H_k = \begin{cases} 1, & FN_k^R > 0 \\ 0, & Otherwise \end{cases} \quad (4)$$

where H_k is the obtained hash code with k bits. Sample hash code is provided in Fig. 4c. The computed hash code's length depends upon the number of frequency components we choose from the FT. The maximum length that can be set for the hash code is equal to the length of the FT or the dimensions of the feature vector f. Each bit of this binary code, correspond to the presence of a particular frequency component in f. A bit valued 1 corresponds to the presence of that particular frequency component, whereas a zero-valued bit would indicate the absence of the frequency component. Mostly the low-frequency content is meaningful because only low frequencies exist in the signal. Hence, we select the low n-frequencies to construct the binary representation. It is important to mention here, that each single bit in the code represent a frequency component or a variation in activation values of multiple convolutional features. Hence, activation patterns are effectively modeled using the binary code which leads to an effective representation of the original feature vector. Further, it is already known that the original signal can be approximately reconstructed from the selected frequency components, the binary code can also be used to reconstruct f in some form. If the high activation values can be restored from the binary representation, then it would guarantee high retrieval performance. Locality sensitivity of hash codes is essential for allowing efficient retrieval from large scale datasets. It refers to the correlation between the Euclidean distance between convolutional features and the hamming distance between their corresponding hash codes. Figure 5 exhibits the locality sensitivity aspect of FFT-based hash codes by highlighting this correlation. The red colored dots represent the relevant

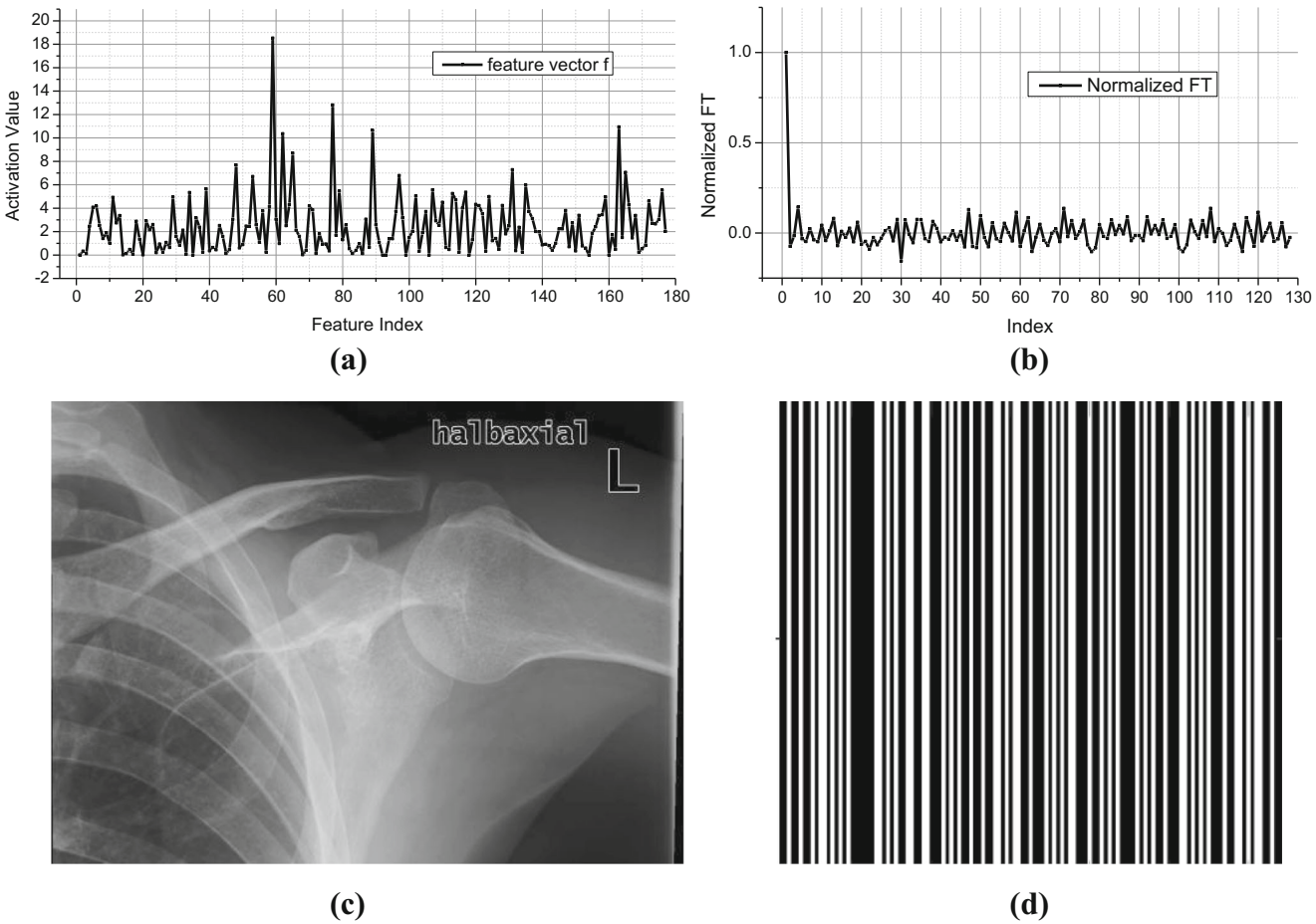

Fig. 4 **a** Sample feature vector (**b**) DC normalized Fourier transform with 128 frequencies, **c** input image, **d** the 128-bit hash code (positive frequencies are represented as white bars and negative frequencies are shown as black bars)

images and the black dots refer to irrelevant images. A high degree of correlation between the Euclidean and hamming distances affirm the locality sensitivity property of these hash codes, which can lead to superior retrieval performance using ANN based search approaches. At 32-bits, the correlation was 0.715, which increased to 0.772 at 64-bit codes. We achieved the highest correlation of 0.883 with 128-bits. Further increase in the number of bits did not yield any significant improvements. The complete algorithm for transforming convolutional features to binary codes is provide in Algorithm 2.

Algorithm 2: Features transformation to binary codes

Input:

Feature vector f

Output

n-bit binary code

Steps:

1. Compute FFT of f as
$$FT_k = \sum_{j=0}^{n-1} f_i e^{-i2\pi kj/n},\ k = 0,....,n-1$$
2. Compute the real component of FT as
$$FT_k^R = \text{Real}(FT_k)$$
3. Normalize the real component of Fourier spectrum by the DC component
$$FN_k^R = \frac{1}{FT_0^R}(FT_k^R)$$
4. Select low n frequencies from the spectrum and transform them to binary codes as
$$H_k = \begin{Bmatrix} 1, & FN_k^R > 0 \\ 0, & Otherwise \end{Bmatrix}$$
5. Return the n-bit code

Experimental Evaluation

Two types of experiments were conducted to assess the effectiveness of both the selected convolutional features and the compact hash codes. During the first set of experiments, we retrieved similar images from the dataset using the original features. Performance is evaluated both qualitatively and quantitatively. In the second set of experiments, retrieval performance and efficiency of the compact binary code is determined. Detailed discussions on the datasets, evaluation metrics, and experimental outcomes are provided in the following sub-sections.

Datasets and Evaluation Metrics

The proposed framework is evaluated using two medical image datasets IRMA-2009 [48] and Kvasir [49]. IRMA-2009 dataset consist of 15,363 radiographs acquired from various body parts. The set of training images contain 13,630 images, whereas the test set consist of 1733 images. Since our feature selection algorithm do not require large number of images, we used the test set for feature selection and the training set was used as the target dataset for retrieval. Kvasir dataset contains 4000 annotated endoscopy images, grouped into 8 different categories. The dataset is mainly used for classification and retrieval. In this dataset, we used randomly selected images as queries and the relevant top rank images were used to compute the retrieval performance. Commonly used evaluation metrics precision and recall have been used to report performance scores.

Retrieval Results on IRMA-2009 Dataset with Optimal Features

The first set of experiments were carried out to assess the retrieval performance of selected features against the full feature set, and other feature pooling

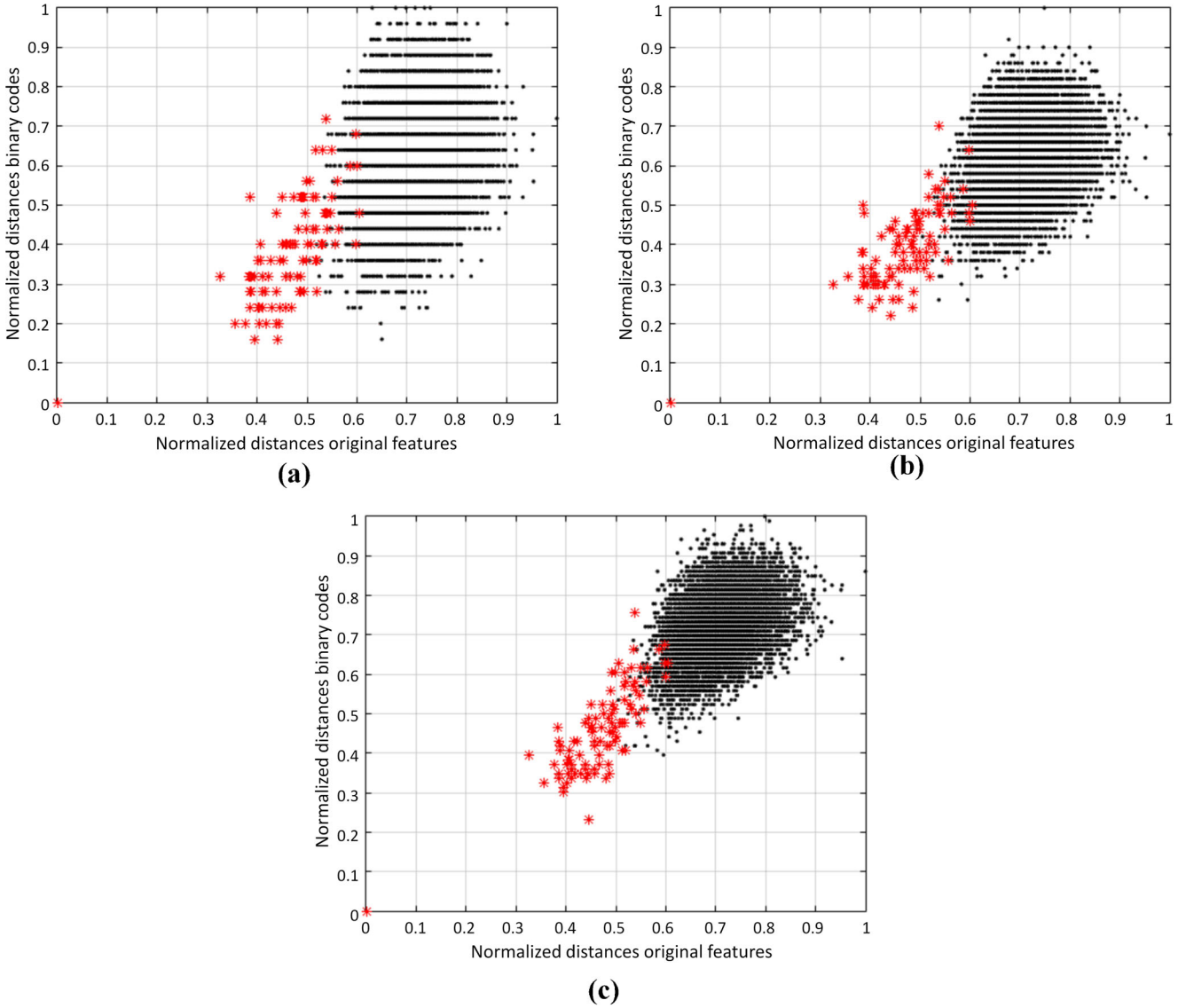

Fig. 5 Locality sensitivity of FFT-based hash codes with (**a**) 32-bits (**b**) 64-bits (**c**) 128-bits

approaches. By varying the value of t in Algorithm 1, we obtained several different subsets of features for image representation. Their retrieval performance was measured in terms of precision and recall. For each experiment, query images were randomly selected from the database, and top-n images were retrieved. Sets of retrieved images for random queries were provided in Fig. 6, where the top-left image is the query image and the remaining ones are the top ranked images according to the distance from the query feature vector. Feature distances computed using Manhattan distance are shown on top of each retrieved image. In all of the results in Fig. 6, the proposed features were able to successfully retrieve correct images at top ranks. Although some images have been incorrectly retrieved in first and last query, these images have been retrieved at lower ranks. Precision-recall curves for various subsets of convolutional features are shown in Fig. 7. It can be seen that removal of irrelevant features not only improves efficiency but also improves retrieval accuracy. With 288 selected features, we obtained the best overall performance as a result of the removal of irrelevant and misleading features. Reducing further degrades performance due to lack of sufficient discriminative capability as can be seen with 177 features. Smaller drop in performance was noticed when the number of features were increased to 355 or 424. These results indicate that the proposed feature

Fig. 6 Top-30 retrieved images for random queries using selected 288 features in IRMA-2009 dataset

selection method can effectively remove irrelevant features and select optimal features from a pre-trained deep CNN when dealing with a particular class of images.

Retrieval Results on Kvasir Dataset with Optimal Features

Similar set of experiments were carried out with this dataset, in order to assess retrieval performance of the proposed features. This dataset is relatively smaller than IRMA-2009 dataset, however it is much more challenging, due to the lack of visible discriminative features as compared to IRMA-2009. Images belonging to different categories have visual similarities, thereby making retrieval more difficult. In such cases, fine-grained features are required for effective representation. In these experiments, we tested the retrieval performance with different subsets of convolutional features, obtained by varying t in Algorithm 1. We obtained five different subsets of features by setting the value of t to 0.1, 0.25, 0.35, and 0.5 to select 242, 342, 396, and 443 features,

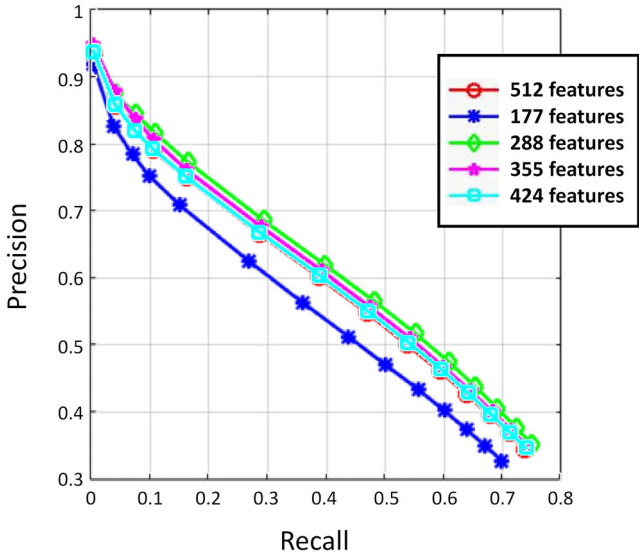

Fig. 7 Precision recall scores computed for retrieval results using different subsets of features in IRMA-2009 dataset

respectively. In the previous experiments, it was observed that the subset having 288 features selected with $t = 0.25$ performed the best. This is because there existed coarse-grained features and a smaller subset was sufficient to effectively represent images. However, this dataset, being more challenging requires a much larger subset for effective representation. Hence, we achieved the best results with 443 features.

In these experiments, we randomly selected query images and retrieved top-n images as shown in Fig. 8. For each query, we retrieved top-30 images from the dataset. Their Euclidean distances from the query image are shown on top of each image. It can be seen in these results that the proposed features were able to retrieve relevant images at top ranks, where high visual similarity can be witnessed at top ranks as compared to the images retrieved at lower ranks. We experimented with various subsets of features in order to determine the optimal

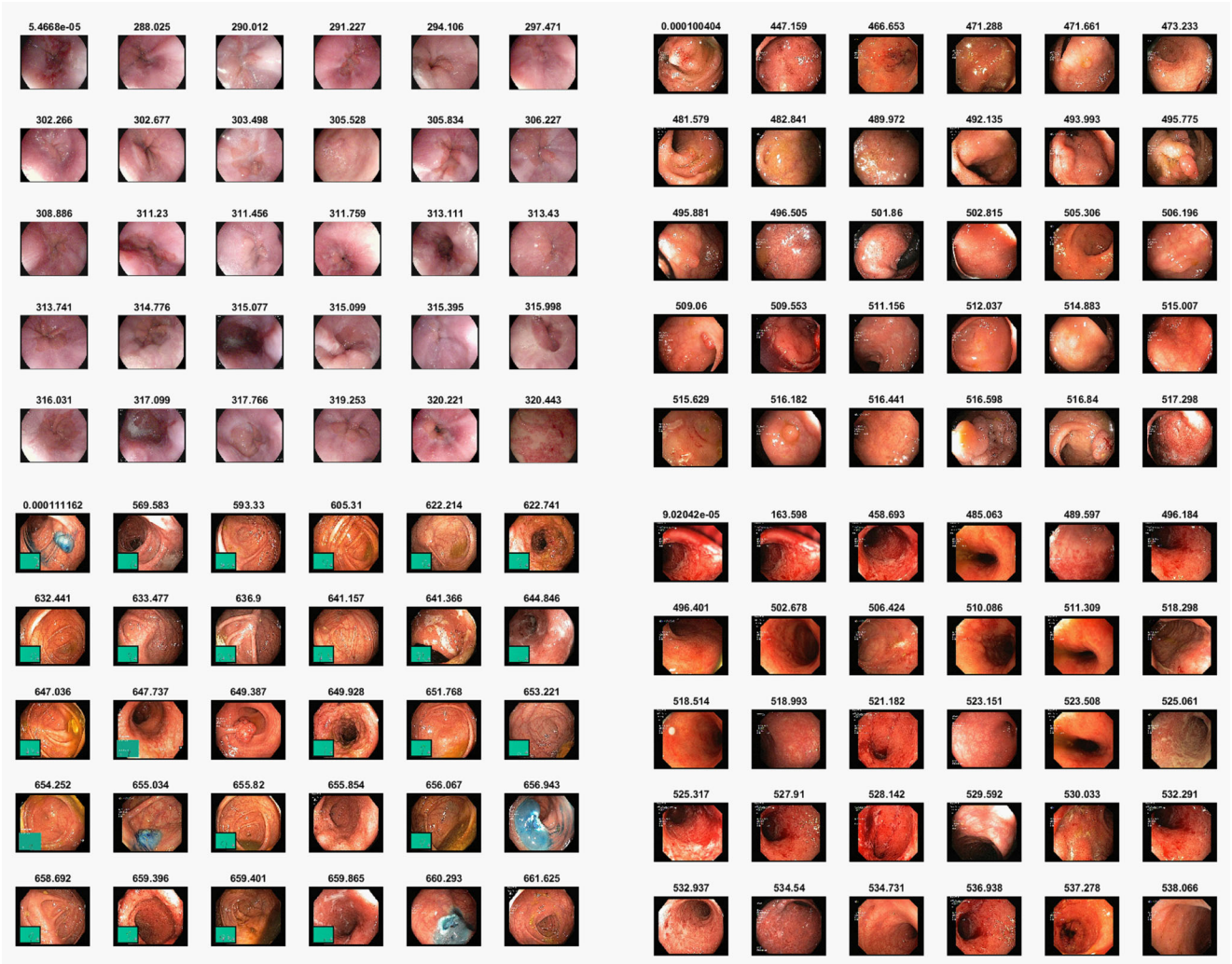

Fig. 8 Top-30 retrieved images for random queries using selected 443 features in Kvasir dataset

subset. We found that the subset with $t = 0.5$ (443 features) is the most optimal one as shown in Fig. 9. However, it is important to note here that the subsets with $t = 0.35$ and $t = 0.25$ performed much better than the full feature set at low recall. However, at higher recall rates, these subsets achieved relatively lower precision scores. These results reveal that the proposed method is capable of selecting optimal set of features for representing endoscopy images.

Retrieval Results on IRMA-2009 Dataset with Compact Binary Codes

When transforming features to binary codes, it is desired that sufficient retrieval precision is achieved with smaller number of bits. We derived codes of varying lengths from the selected convolutional features and retrieved images using the binary codes. Results revealed that relatively smaller performance hit is noticed when images are retrieved with 128-bit codes. Reducing the code length gradually decreases performance. For the smallest code (16 bits), our method was able to retrieve sufficient number of relevant images when a pool of 100 candidate images were retrieved. In such a case, the tiny code can be used as a first phase of image retrieval from large scale datasets, to extract a pool of candidate images, followed by re-ranking using the original feature set. Figure 10 shows retrieval results using 128-bit codes for four random queries. The hamming distances between the query image (top-left) and the retrieved images is shown on top of each image. It can be seen from these results that the proposed method effectively transform features to compact binary codes, having sufficient discriminative ability.

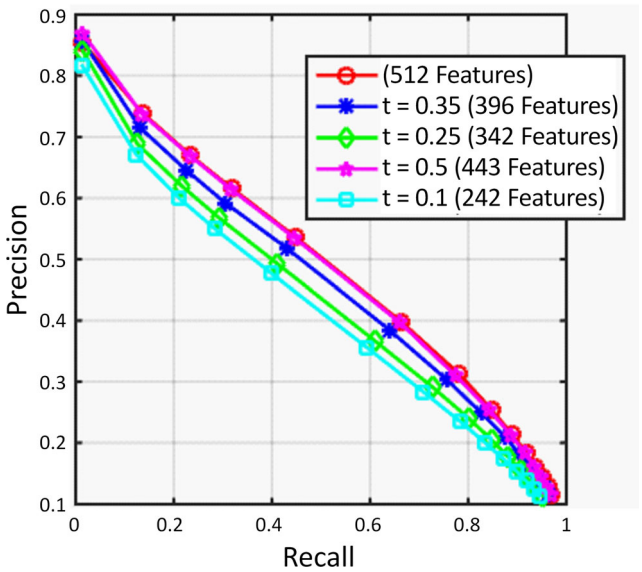

Fig. 9 Precision recall scores computed for retrieval results using different subsets of features in Kvasir dataset

Different experiments were carried out with varying bit lengths ranging from 16-bits to 128-bits. We observed gradual improvements in retrieval performance with the increase in number of bits as shown in Fig. 11. With more bits the signal is approximated in a much better way which results in better performance. With 128-bits hash code, optimal performance was achieved and it was chosen as the optimal bit length due to its efficiency over 256 or higher bit codes. Figure 12 reports retrieval performance in terms of precision-recall curve with 128-bits code using various subsets of features. Four different cases are considered by varying the value of t in Algorithm 1. We experimented with $t = 0.5$, $t = 0.35$, $t = 0.25$, and $t = 0.1$, for which we obtained subsets having 424, 355, 288, and 177 feature maps, respectively. Interestingly, we obtained the best overall results with 288 feature maps, selected by setting $t = 0.25$. For feature sets chosen with $t = 0.35$ and $t = 0.5$, we obtained slightly poor results although the number of features in these cases was higher than $t = 0.25$. By decreasing the value of t to 0.1, performance decreased significantly because of the loss of essential features. With these results, we can say that the proposed feature selection scheme can effectively shortlist features capable of representing medical images in the best possible way. Besides decreasing the space requirements and reducing dimensionality, the optimal feature set results in improved performance, especially at top ranks.

Retrieval Results on Kvasir Dataset with Compact Binary Codes

Image retrieval with hash codes is highly desirable in large scale datasets. In these experiments, we evaluated various aspects of hash codes on retrieval performance, including code length and optimality of feature subset for hash code generation. We already know that this dataset is more challenging than the IRMA-2009 dataset, so greater degradation is expected with precision scores at high recalls. In the first experiment, we evaluated hash codes of varying lengths in order to determine optimal hash code length. Random queries were chosen and top-n images were retrieved with hash codes of varying lengths ranging from 16-bits to 128-bits. Gradual improvement in performance was noticed with increase in hash code length. With 128-bits, we achieved the best scores with the optimal set of features. Figure 13 contains results of four different

Fig. 10 Top-30 retrieved images for random queries in IRMA-2009 dataset using compact (128-bits) binary codes

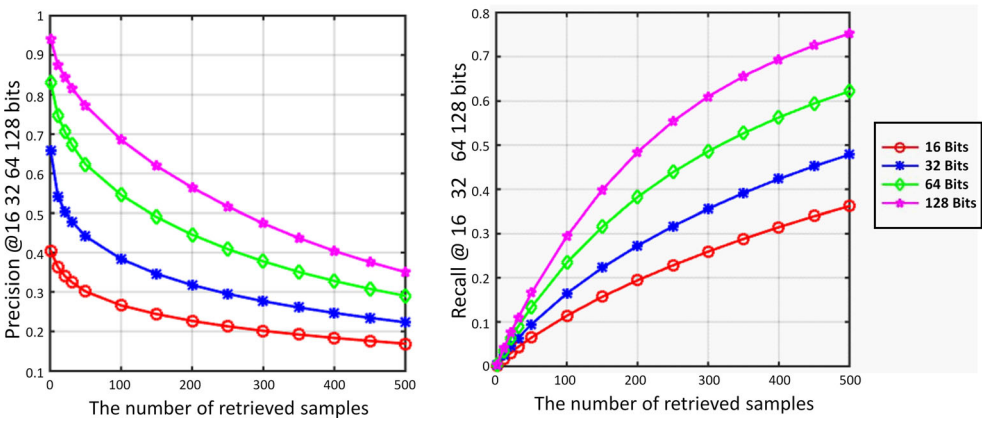

Fig. 11 Precision and recall for binary codes of varying lengths for IRMA-2009 dataset

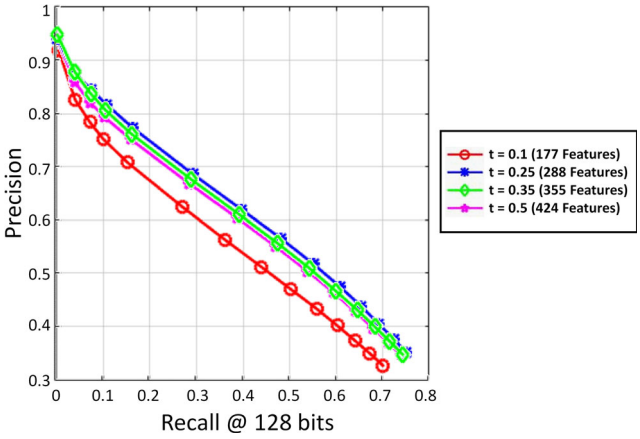

Fig. 12 Precision recall curves for different feature subsets represented with 128-bit binary codes for IRMA-2009 dataset

queries, where top-30 images were retrieved using hash codes of 128-bits. Visual similarity in the retrieved images exhibit the representational strength of proposed hashing method. Precision and recall scores for 16, 32, 64, and 128-bit codes are provided in Fig. 14. Finally, Fig. 15 shows the precision recall curves for 128-bit hash codes obtained from different subsets of features. Interestingly, the subsets with thresholds 0.1, 0.25 and 0.35 yielded better precision scores than the full feature set as well as the subset with 443 features. This outcome indicates that features whose dimensions are closer to the number of bits in the hash codes, will perform better than the other subsets or even the full set of features. It is also important to mention here that this only holds true when sufficient number of features are used in representation. If very small subset is selected, then the corresponding hash codes may not perform as well, which has been seen with the experiments in IRMA-2009 dataset. Hence, for sufficient number of features, hash codes can effectively represent medical images, as is evident from these results.

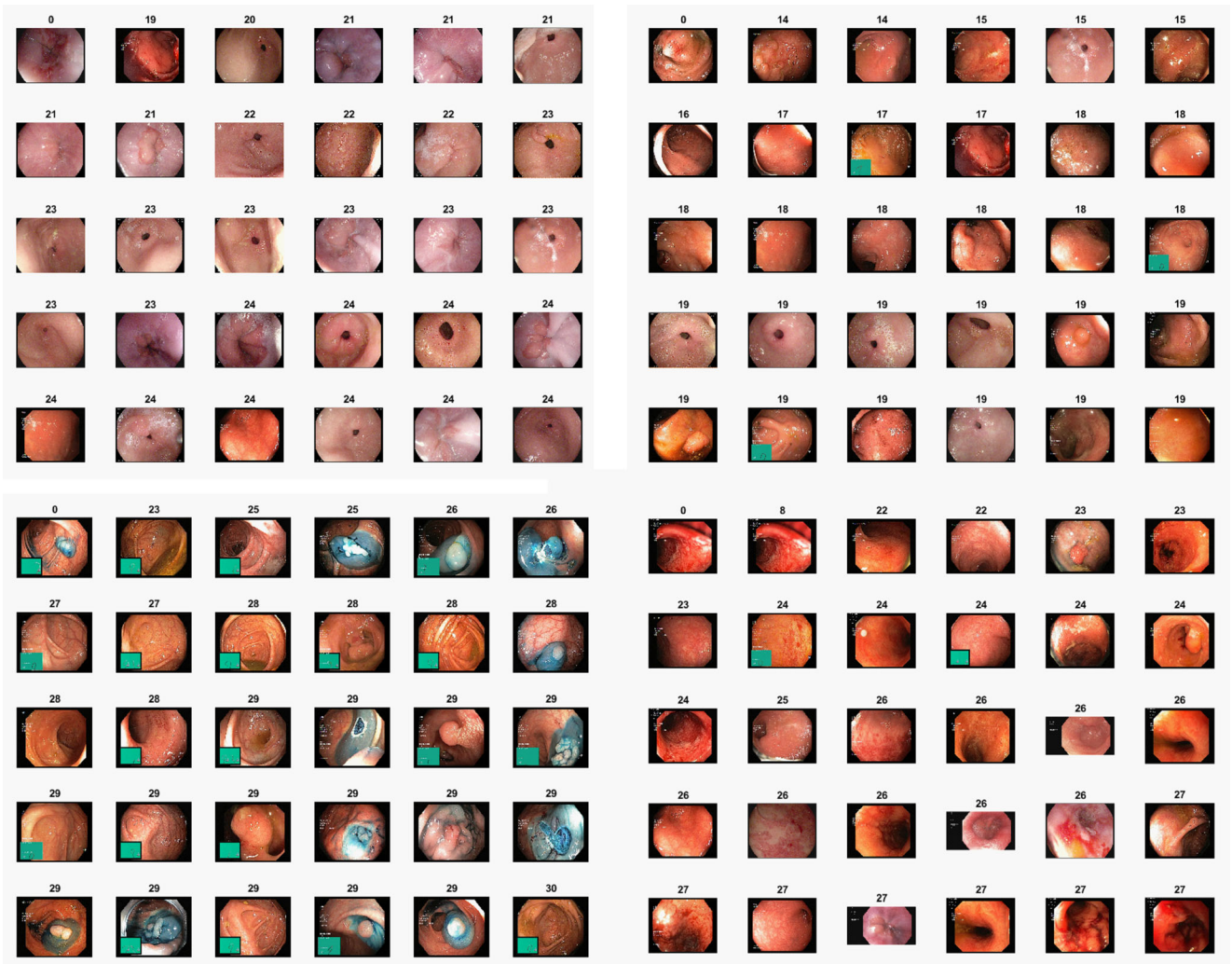

Fig. 13 Top-30 retrieved images for random queries in Kvasir dataset using compact (128-bits) binary codes

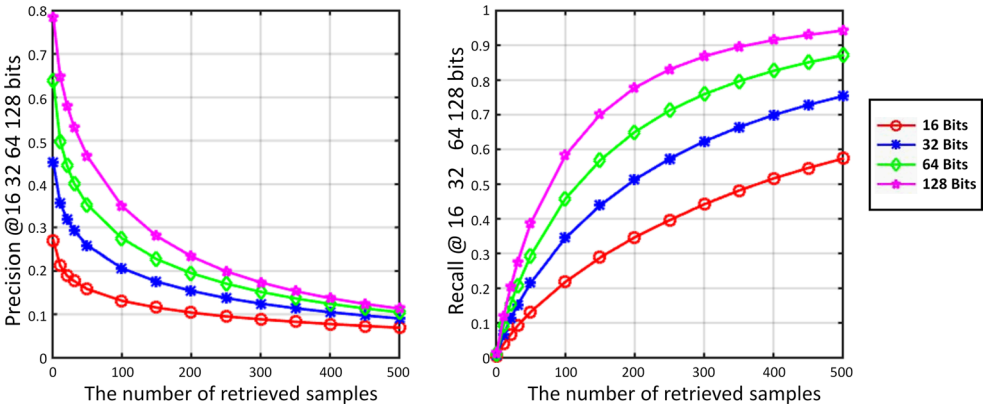

Fig. 14 Precision and recall for binary codes of varying lengths for Kvasir dataset

Comparison with State-of-the-Art

Due to the retrieval efficiency of compact hash codes in huge image repositories, numerous methods have been proposed in the past which yield state-of-the-art performance in a variety of datasets. We compared our method with several of them to determine how it compares with these approaches. Several of these methods like LSH, CBE-opt, SpH, SH, PCA-RR, PCA-H, and DSH require compute intensive training as well as the conversion of feature vector to hash codes. The proposed method do not require any training and the conversion is fast which makes it more suitable for large scale datasets. Further, the conversion speed can be further improved with GPU based computation of FFT. Though the performance of our method is slightly lower than SpH, and PCA-RR at 128-bits, it is much efficient than these methods. Further, our method significantly outperforms PCA-H, CBE-opt, SH, and DSH at 128-bit hash codes. Figure 16 shows precision and recall score comparisons of the proposed method with several other methods. The top row indicates precision rates for different number of retrieved images at 64 and 128-bit representations. The middle row shows recall rates and the last row reports comparison of precision-recall curves. At 64-bits, the proposed method achieves better performance than PCAH, CBE-opt, SH, and LSH. Whereas, the performance is increased when 128-bit encoding is used, surpassing DSH and nearly reaching SpH at high recall rates.

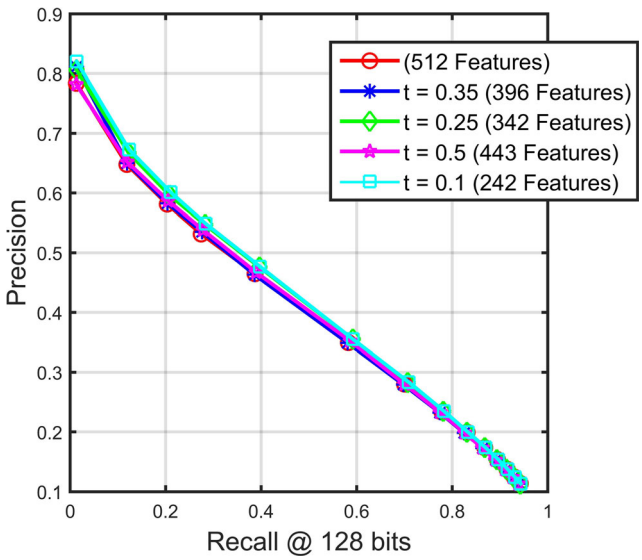

Fig. 15 Precision recall curves for different feature subsets represented with 128-bit binary codes for Kvasir dataset

Similar results were achieved with the Kvasir dataset where FFT-based hash codes outperformed PCAH, SH, CBE-opt, and LSH methods with 64-bit codes as shown in Fig. 17. However, its performance was poor than DSH, SpH, and PCA-RR. When the number of bits were increased to 128, performance with proposed hash codes increased significantly and outperformed PCAH, CBE-opt, SH, LSH, and DSH. Though its performance was still slightly lower than SpH and PCA-RR, the efficient computation of FFT-based codes make it more suitable and convenient for implementation in real applications. Further, due to its ease of implementation, it can be applied to efficiently extract a set of candidate images using very short binary codes and then use relatively longer codes to re-rank them, in order to achieve a high degree of precision in an efficient manner.

Conclusions and Future Work

In this paper, we presented an efficient method to select optimal subset of features from deep convolutional layers of a pre-trained CNN. We showed that the chosen subset of features perform much better than the full set of features in medical image retrieval from large datasets. We also proposed a highly efficient method to represent these features as compact binary codes using Fast Fourier Transform (FFT). The feature vector

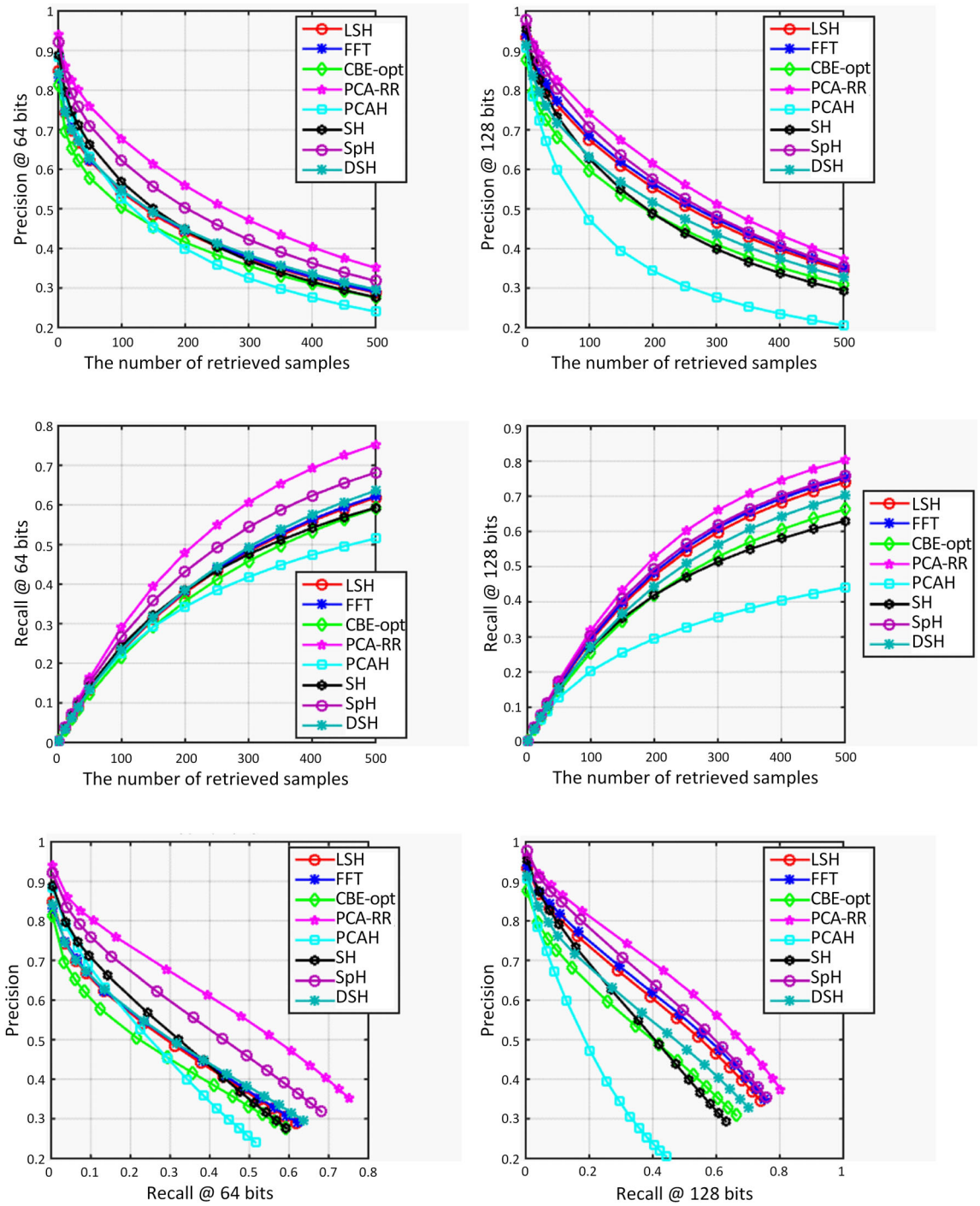

Fig. 16 Precision recall scores computed for retrieval results in IRMA-2009 dataset using compact binary codes

is treated as a one-dimensional signal and is transformed to the frequency domain using FFT. The transformed feature vector is then transformed into bits of our choice by selecting the required number of frequency components and transforming them into bits using simple linear transformation. The acquired binary codes serve as hash codes and allow efficient retrieval in large scale datasets. These proposed hash codes have been extensively compared with several state-of-the-art methods, most of which it significantly outperformed. However, the proposed method could not perform very well with very short codes, primarily due to the use of simple transformation of Fourier spectrum to binary codes.

In future, we plan to extend the framework by developing more efficient means to transform feature vectors

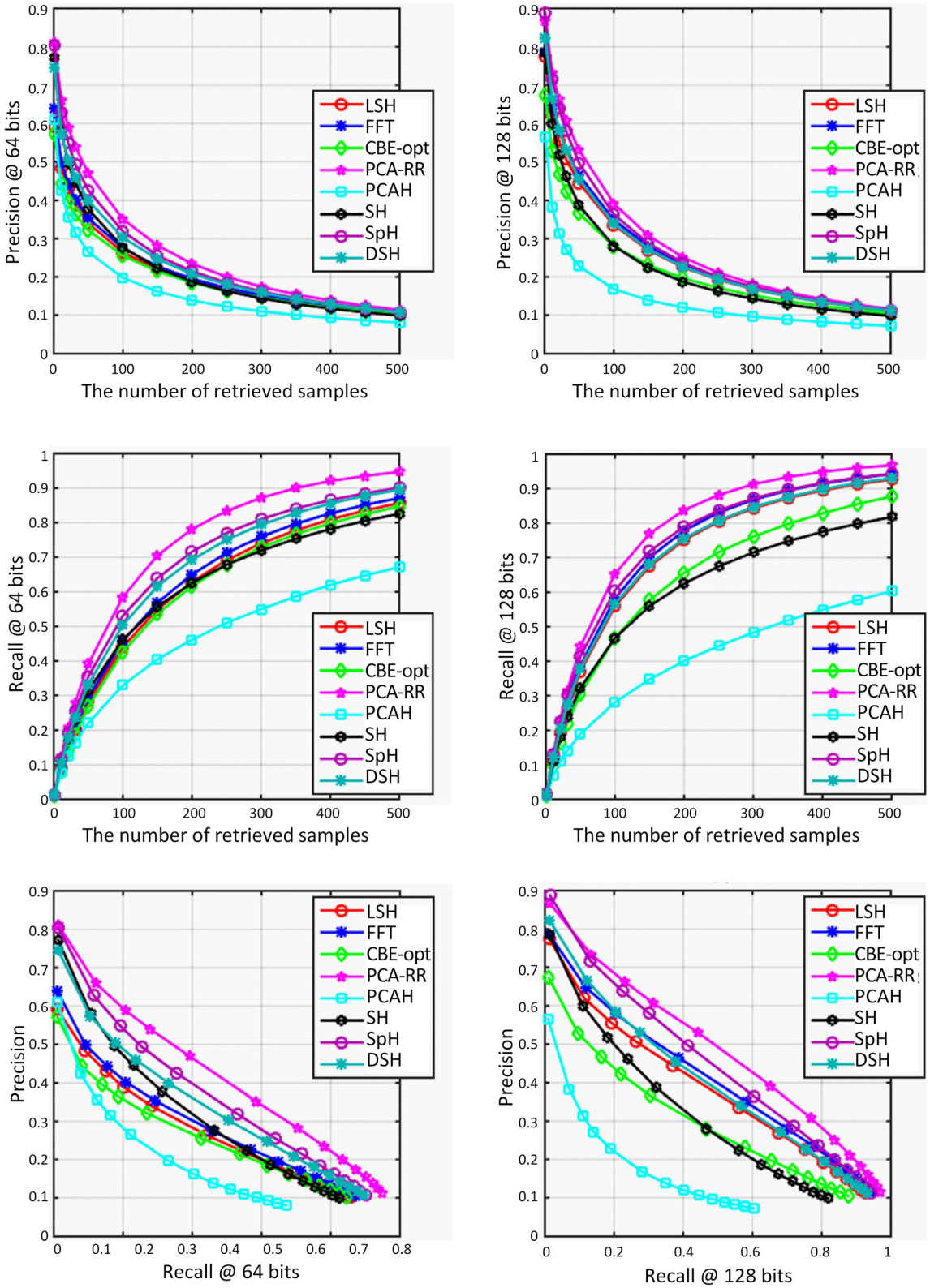

Fig. 17 Precision recall scores computed for retrieval results in Kvasir dataset using compact binary codes

into hash codes. Other frequency domain transformation methods will also be investigated for this purpose. We will also evaluate other techniques to quantize the FFT coefficients for optimal conversion. We strongly hope that the proposed method can be effectively applied to medical image retrieval in large scale datasets.

Funding This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIP) (No.2016R1A2B4011712).

Compliance with Ethical Standards

Conflict of Interest The authors declare that there is no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Aborokbah, M.M., Al-Mutairi, S., Sangaiah, A.K., and Samuel, O.W., Adaptive context aware decision computing paradigm for intensive health care delivery in smart cities—A case analysis. *Sustain. Cities Soc.*, 2017. https://doi.org/10.1016/j.scs.2017.09.004.
2. Samuel, O.W., Asogbon, G.M., Sangaiah, A.K., Fang, P., and Li, G., An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Expert Syst. Appl.* 68:163–172, 2017.
3. Ahmad, J., Sajjad, M., Mehmood, I., and Baik, S.W., SiNC: Saliency-injected neural codes for representation and efficient retrieval of medical radiographs. *PloS One*. 12(8):e0181707, 2017.
4. Ahmad, J., Sajjad, M., Mehmood, I., Rho, S., and Baik, S.W., Saliency-weighted graphs for efficient visual content description and their applications in real-time image retrieval systems. *J. Real-Time Image Proces.* 13(3):431–447, 2017. https://doi.org/10.1007/s11554-015-0536-0.
5. Ahmad, J., Sajjad, M., Rho, S., and Baik, S.W., Multi-scale local structure patterns histogram for describing visual contents in social image retrieval systems. *Multimed. Tools Appl.* 75(20):12669–12692, 2016. https://doi.org/10.1007/s11042-016-3436-9.
6. Jégou, H., Douze, M., and Schmid, C., Improving bag-of-features for large scale image search. *Int. J. Comput. Vis.* 87(3):316–336, 2010.
7. Wang, J., Li, Y., Zhang, Y., Xie, H., and Wang, C. Boosted learning of visual word weighting factors for bag-of-features based medical image retrieval. In: Image and Graphics (ICIG), 2011 Sixth International Conference on, 2011. IEEE, pp 1035–1040
8. Lowe, D.G., Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60(2):91–110, 2004. https://doi.org/10.1023/B:VISI.0000029664.99615.94.
9. Dalal, N., and Triggs, B., Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE computer society conference on, 2005. IEEE, pp 886–893
10. Jégou, H., Douze, M., Schmid, C., and Pérez, P., Aggregating local descriptors into a compact image representation. In: Computer Vision and Pattern Recognition (CVPR), 2010 I.E. conference on, 2010. IEEE, pp 3304–3311
11. Douze, M., Ramisa, A., and Schmid, C., Combining attributes and fisher vectors for efficient image retrieval. In: in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011. IEEE, pp 745–752
12. Liu, L., Shen, C., Wang, L., van den Hengel, A., and Wang, C., Encoding high dimensional local features by sparse coding based fisher vectors. In: Advances in neural information processing systems, 2014. pp 1143–1151
13. Oliva, A., and Torralba, A., Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int. J. Comput. Vis.* 42(3):145–175, 2001.
14. Wu, J., and Rehg, J.M., CENTRIST: A visual descriptor for scene categorization. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(8):1489–1501, 2011.
15. Zhang, R., Shen, J., Wei, F., Li, X., and Sangaiah, A. K., Medical image classification based on multi-scale non-negative sparse coding. *Artif. Intell. Med.* 83:44–51, 2017. https://doi.org/10.1016/j.artmed.2017.05.006.
16. He, K., Zhang, X., Ren, S., and Sun, J., Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. pp 770–778
17. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A., Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp 1–9
18. Girshick, R., Donahue, J., Darrell, T., and Malik, J., Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014. pp 580–587
19. Long, J., Shelhamer, E., and Darrell, T., Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp 3431–3440
20. Babenko, A., Slesarev, A., Chigorin, A., and Lempitsky, V., Neural codes for image retrieval. In: Computer Vision–European Conference on Computer Vision (ECCV). Springer, 2014. pp 584–599. doi:https://doi.org/10.1007/978-3-319-10590-1_38
21. Babenko, A., and Lempitsky, V., Aggregating local deep features for image retrieval. In: Proceedings of the IEEE international conference on computer vision, 2015. pp 1269–1277
22. Ahmad, J., Mehmood, I., and Baik, S.W., Efficient object-based surveillance image search using spatial pooling of convolutional features. *J. Vis. Commun. Image Represent.* 45:62–76, 2017.
23. Ahmad, J., Mehmood, I., Rho, S., Chilamkurti, N., and Baik, S.W., Embedded deep vision in smart cameras for multi-view objects representation and retrieval. *Comput. Electr Eng.* 61C:297–311, 2017. https://doi.org/10.1016/j.compeleceng.2017.05.033.
24. Ahmad, J., Muhammad, K., and Baik, S.W., Data augmentation-assisted deep learning of hand-drawn partially colored sketches for visual search. *PloS One*. 12(8):e0183838, 2017.
25. Krizhevsky, A., Sutskever, I., and Hinton, G. E., Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. Curran associates, Inc., pp 1097–1105, 2012.
26. Gong, Y., Wang, L., Guo, R., and Lazebnik, S., Multi-scale orderless pooling of deep convolutional activation features. In: Computer Vision–ECCV 2014. Springer, pp 392-407, 2014
27. Kalantidis, Y., Mellina, C., and Osindero, S., Cross-dimensional weighting for aggregated deep convolutional features. In: European Conference on Computer Vision, 2016. Springer, pp 685–701
28. Alzu'bi, A., Amira, A., and Ramzan, N., Content-based image retrieval with compact deep convolutional features. Neurocomputing 249:95-105, 2017. https://doi.org/10.1016/j.neucom.2017.03.072.
29. Razavian, A. S., Azizpour, H., Sullivan, J., and Carlsson, S., CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. In: 2014 I.E. Conference on Computer Vision and Pattern Recognition Workshops, 23–28 June 2014 2014. pp 512–519. https://doi.org/10.1109/CVPRW.2014.131
30. Azizpour, H., Razavian, A., Sullivan, J., Maki, A., and Carlsson, S., From generic to specific deep representations for visual recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015. pp 36–45.
31. Mohedano, E., McGuinness, K., O'Connor, N. E., Salvador, A., Marqués, F., and Giró-i-Nieto, X., Bags of local convolutional

features for scalable instance search. In: Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval, 2016. ACM, pp 327–331

32. Srinivas, M., Naidu, R.R., Sastry, C., and Mohan, C.K., Content based medical image retrieval using dictionary learning. *Neurocomputing*. 168:880–895, 2015. https://doi.org/10.1016/j.neucom.2015.05.036.
33. Ahmad, J., Muhammad, K., Lee, M.Y., and Baik, S.W., Endoscopic image classification and retrieval using clustered convolutional features. *J. Med. Syst.* 41(12):196, 2017. https://doi.org/10.1007/s10916-017-0836-y.
34. Liao, X., Yin, J., Guo, S., Li, X., and Sangaiah, A.K., Medical JPEG image steganography based on preserving inter-block dependencies. *Comput. Electr. Eng.*, 2017. https://doi.org/10.1016/j.compeleceng.2017.08.020.
35. Charikar, M. S., Similarity estimation techniques from rounding algorithms. In: Proceedings of the thiry-fourth annual ACM symposium on Theory of computing, 2002. ACM, pp 380–388.
36. Weiss, Y., Torralba, A., and Fergus, R., Spectral hashing. In: Advances in neural information processing systems, 2009. pp 1753–1760.
37. Heo, J-P., Lee, Y., He, J., Chang, S-F., and Yoon, S-E., Spherical hashing. In: Computer Vision and Pattern Recognition (CVPR), 2012 I.E. conference on, 2012. IEEE, pp 2957–2964.
38. Kulis, B., and Grauman, K., Kernelized locality-sensitive hashing. *IEEE Trans. Pattern Anal. Mach. Intell.* 34(6):1092–1104, 2012. https://doi.org/10.1109/TPAMI.2011.219.
39. Jin, Z., Li, C., Lin, Y., and Cai, D., Density sensitive hashing. *IEEE trans.cybern.* 44(8):1362–1371, 2014.
40. Gong, Y., and Lazebnik, S., Iterative quantization: A procrustean approach to learning binary codes. In: Computer Vision and Pattern Recognition (CVPR), 2011 I.E. Conference on, 2011. IEEE, pp 817–824.
41. Yu, F., Kumar, S., Gong, Y., and Chang, S.- F., Circulant binary embedding. In: International conference on machine learning, 2014. pp 946–954.
42. Zhang, T., Du, C., and Wang, J., Composite Quantization for Approximate Nearest Neighbor Search. In: ICML, 2014. vol 2. pp 838–846
43. Erin Liong, V., Lu, J., Wang, G., Moulin, P., and Zhou, J., Deep hashing for compact binary codes learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp 2475–2483
44. Lai, H., Pan, Y., Liu, Y., and Yan, S., Simultaneous feature learning and hash coding with deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp 3270–3278.
45. Zhao, F., Huang, Y., Wang, L., and Tan, T., Deep semantic ranking based hashing for multi-label image retrieval. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. pp 1556–1564.
46. Deng, J., Dong, W., Socher, R., Li, L-J., Li, K., and Fei-Fei, L., Imagenet: A large-scale hierarchical image database. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 2009. IEEE, pp 248–255
47. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 14091556
48. Welter, P., Deserno, T.M., Fischer, B., Günther, R.W., and Spreckelsen, C., Towards case-based medical learning in radiological decision making using content-based image retrieval. *BMC Med. Inform. Decis. Mak.* 11(1):1, 2011. https://doi.org/10.1186/1472-6947-11-68.
49. Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D-T., Lux, M., and Schmidt, P. T., Kvasir: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. In: Proceedings of the 8th ACM on Multimedia Systems Conference, 2017. ACM, pp 164–169.