

Attention-based end-to-end CNN framework for content-based X-ray image retrieval

Şaban ÖZTÜRK^{1,*}, Adi ALHUDHAIF², Kemal POLAT³

¹Department of Electrical and Electronics Engineering, Amasya University, Amasya, Turkey

²Department of Computer Science, College of Computer Engineering and Sciences in Al-kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia

³Department of Electrical and Electronics Engineering, Bolu Abant İzzet Baysal University, Bolu, Turkey

Received: 26.05.2021

Accepted/Published Online: 09.08.2021

Final Version: 04.10.2021

Abstract: The widespread use of medical imaging devices allows deep analysis of diseases. However, the task of examining medical images increases the burden of specialist doctors. Computer-assisted systems provide an effective management tool that enables these images to be analyzed automatically. Although these tools are used for various purposes, today, they are moving towards retrieval systems to access increasing data quickly. In hospitals, the need for content-based image retrieval systems is seriously evident in order to store all images effectively and access them quickly when necessary. In this study, an attention-based end-to-end convolutional neural network (CNN) framework that can provide effective access to similar images from a large X-ray dataset is presented. In the first part of the proposed framework, a fully convolutional network architecture with attention structures is presented. This section contains several layers for determining the saliency points of X-ray images. In the second part of the framework, the modified image with X-ray saliency map is converted to representative codes in Euclidean space by the ResNet-18 architecture. Finally, hash codes are obtained by transforming these codes into hamming spaces. The proposed study is superior in terms of high performance and customized layers compared to current state-of-the-art X-ray image retrieval methods in the literature. Extensive experimental studies reveal that the proposed framework can increase the current precision performance by up to 13

Key words: X-ray, attention, retrieval, hash, CNN

1. Introduction

Today, we know that many diseases that threaten human life can be treated by early diagnosis. Thanks to the increase in technological developments and scientific knowledge, there are positive developments in health systems almost everywhere in the world. Medical imaging systems have become one of the most widely used medical techniques recently [1]. However, the time required to examine the images created by these systems, which are frequently used by specialist doctors for detection of diseases, requires an additional budget. Early diagnosis cannot be achieved in countries where the number of physicians per patient is insufficient due to the time allocated for the examination of each image [2]. Artificial intelligence techniques are actively used for automated medical image analysis tasks. These systems, which can be used as a decision support mechanism or decision-making system, generally perform tasks such as classification, segmentation, detection, and tracking [3]. Medical image analysis based on these tasks has been studied for many years, and many problems have

*Correspondence: saban.ozturk@amasya.edu.tr

found solutions. However, many problems related to content-based image retrieval (CBIR) problems are still open. Especially the storage of images and quick access to these stored images is still an issue that needs improvement. CBIR, which is preferred for providing access to images with similar content, is an area that has been studied for 20 years and is still an active research area. In CBIR systems, each image in a dataset is represented by a low-dimensional vector according to its content. In this way, the space required for the storage of each image is reduced. In the retrieval phase, the desired number of images are requested to be presented to the user automatically in response to a query image [4]. The use of the nearest neighbor (NN) search in the process of identifying similar images provides high performance. However, using exact NN (ENN) creates an unacceptable time and memory load for large datasets [5]. An approximate nearest neighbor (ANN) search is used to alleviate this problem. ANN search exhibits very high probability and low accuracy loss performance in a short time.

Hashing is the most popular ANN method. While many traditional ENN solutions use the Euclidean space, hashing-based solutions use the Hamming space. For this purpose, real-valued data points are encoded into binary codes. Thus, low storage cost and fast retrieval can be provided [6]. CBIR systems aim to solve the 'semantic gap' between linguistic information generated by an expert and machine-generated information. Hashing methods are generally examined in two separate groups, data-dependent, and data-independent. Since data-independent methods produce random variables, they are not sufficient to relieve the semantic gap problem. For this reason, data-dependent methods are preferred today. Data-dependent methods are examined under three groups as unsupervised methods, semisupervised methods, and supervised methods. Detailed information about the development of these methods can be found in [7]. The development of hashing methods in terms of feature extraction can be divided into hand-crafted features and CNN-based features. Detailed literature analysis related to the feature extraction approach can be examined from [8]. It is seen from the studies in the literature that supervised hashing methods produce higher performance than other hashing methods if a sufficient number of labeled data can be reached. On the other hand, using a CNN-based feature extraction method in this supervised technique provides higher precision performance than hand-crafted methods. The latest state-of-the-art studies in the literature generally prefer such types of approaches [9].

CBIR systems are extensively studied for natural image processing problems, but retrieval for medical images remains a challenging task. One of the important reasons for this is that the complexity of medical images makes automatic analysis difficult. In addition, while the differences in the medical images within the same classes are quite high, the differences between the classes are small. Therefore, they are very challenging for computer-assisted systems. On the other hand, today's deep learning architectures are generally designed specifically for natural image processing problems. For this reason, many deep architectures offered as pretrained cannot produce high performance in medical image processing tasks [10]. In addition, a large number of labeled data is needed to train deep architectures. It is challenging to access medical datasets containing a large number of labeled data. For this reason, most of the medical image retrieval systems in the literature benefit from hand-crafted feature extraction methods [11–13]. Due to the higher performance of deep learning-based methods, some shallow CNN methods are used for this task [14, 15]. However, shallow learning does not provide high performance for large-scale datasets and does not provide high performance for real-time problems. For this reason, researchers are actively working on high-performance deep learning methods in high medical image retrieval tasks [16, 17].

Chest X-ray is one of the most commonly used radiological examination methods. Although the automatic X-ray analysis studies have gained importance due to the COVID-19 pandemic today, the retrieval task still has

not received the necessary attention [18]. While X-ray image retrieval studies performed with radon barcodes and traditional methods [19] and Siamese deep learning architectures [20] are quite remarkable, their performance is insufficient. End-to-end style deep learning architectures are better at dealing with today's X-ray image retrieval problems [21]. However, the difficulties in X-ray images still do not allow these methods to be used in real-time. Generally, the very similar background in X-ray images and blurry patterns create problems. In the literature, attention approaches are generally preferred to solve these problems [22]. Attention approaches play an important role in determining the saliency points in the image and enable the determination of important regions. Thus, feature extraction approaches pay more attention to important regions.

This work presents an attention-based end-to-end CNN framework for X-ray image retrieval tasks. Unlike many methods in the literature, attention deep CNN architecture is recommended instead of using hand-crafted feature extraction to solve the 'semantic gap' problem. Attention deep CNN is the first part of the proposed framework. In this section, new layers are proposed by creating an architecture suitable for the characteristics of X-ray images. In the second stage, a pretrained ResNet18 model is used for hashing. The proposed method is trained using the largest X-ray data in the literature. This study is the first X-ray image retrieval study performed using residual attention CNN to the best of the authors' knowledge. The major contributions of this paper can be summarized as follows:

- The proposed method determines the areas of attention using the residual saliency mechanism. It is the first for X-ray image retrieval in the literature.
- Siamese-style training is carried out to relieve the unbalanced problem. Moreover, two different loss functions and a combination strategy are proposed to release this problem.
- The results of the proposed framework produced by the experiments using X-ray images are superior to other state-of-the-art methods.

The rest of this paper is organized as follows. Details of the proposed method and loss functions are described in Section 2. Then, experimental results and dataset properties are presented in Section 3. Finally, the conclusion is discussed in Section 4.

2. Methodology

In this section, some notations and proposed framework used for content-based image retrieval (CBIR) are introduced. Let $X = [x_1, x_2, \dots, x_n]^T$ be the dataset, n represents the number of X-ray images. Let the hash codes of this data matrix be expressed with $Y = [y_1, y_2, \dots, y_n]^T$. Each y is made up of real numbers, and if its length is expressed as d , then a matrix $R_{n \times d}$ is formed. Also, let $F_n = [f_1, f_2, \dots, f_m]$ for learned features for each X-ray images, m represents hash code lengths. CBIR architectures generate descriptive codes (F) of a certain length for each image. The lengths of these codes are usually 32, 64, and 128 in the literature [6, 8]. The direct binary generation of these codes by the CNN architecture is very challenging. In order to provide relaxation, these codes are provided in Euclidean space. These codes are then converted into the binary form using Equation 1, and the problem is pulled into the Hamming space.

$$\text{sign}(f) = \begin{cases} 1 & f \geq 0 \\ -1 & f < 0 \end{cases} \quad (1)$$

The proposed framework for the content-based X-ray image retrieval is shown in Figure 1. The proposed framework works end-to-end, but it is divided into two parts to be examined in more detail. Firstly, the transformation of X-ray images is performed. In this section, which we can call preprocessing, it is aimed to highlight the lung sections. Then attention deep CNN architecture suggests saliency regions in X-ray images. In this part, there is a traditional fully convolutional network (FCN) structure reinforced with attention layers. One or more attention areas can be suggested automatically by the network. In the second part, there is a pretrained ResNet18 architecture as a hashing network. Attention map and original X-ray image are used as input to this section. The codes produced by the hashing network are finally converted into binary form as postprocessing. For the training process, we use a combination of weighted cross-entropy loss and contrastive loss.

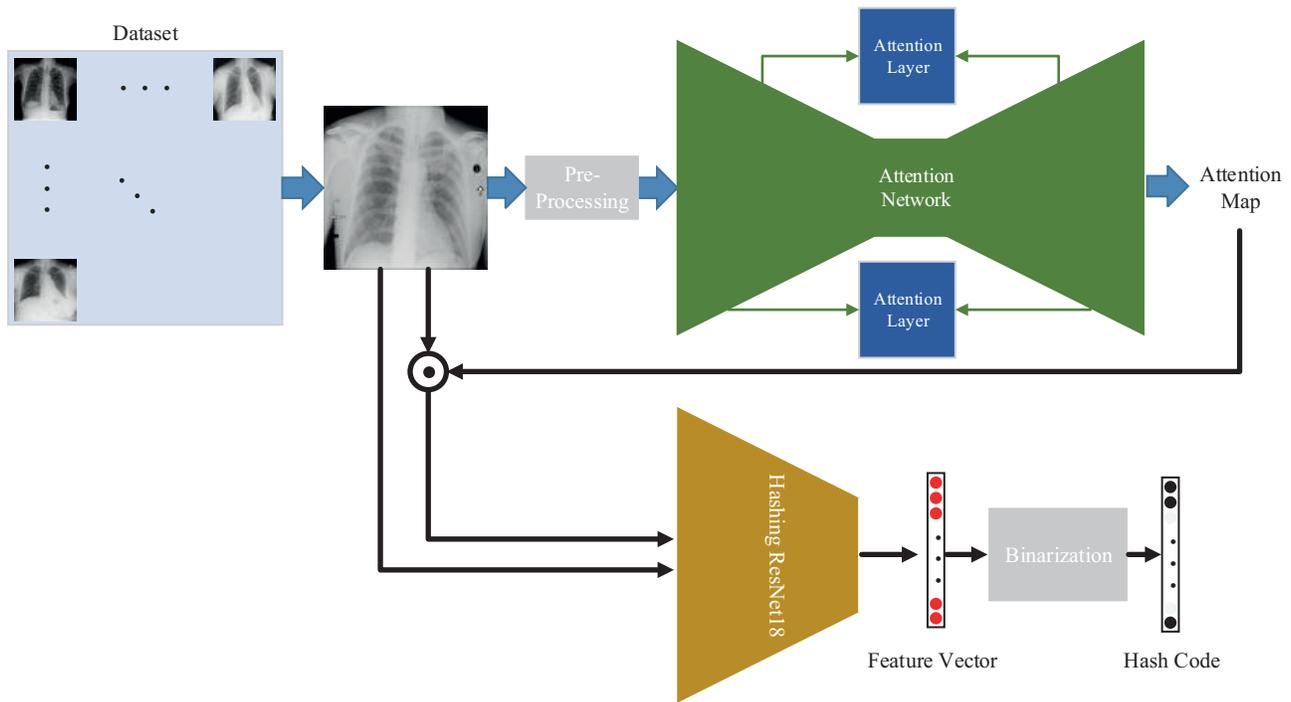


Figure 1. Proposed X-ray retrieval framework.

2.1. Preprocessing of X-ray Images

Making the lung areas more prominent in X-ray images and cleaning the noise is an essential factor for increasing performance. However, cleaning the noise causes the loss of important information in some cases. To get the most efficiency from the automatic learning process in deep learning algorithms and to adapt to real-time image processing problems, a hard cleaning process is not performed in this study. Equation 2 is used to bring the information in the lung regions to the forefront.

$$im_p = uint(255 * (1 - normalize(im_{or}))) \tag{2}$$

where im_p represents the preprocessed image, im_{or} represents the original image, $normalize$ denotes 0-1 standard normalization process.

2.2. Attention network structure

The attention network part of the proposed framework is used to determine the saliency points of X-ray images. No additional training or additional weighting is performed in this study to obtain saliency points [23]. This task is partially fulfilled by the X-ray image preprocessing layer. If this part of the proposed framework is called *Attent*, the saliency map that this framework will generate is as in Equation 3. The next step, the hashing network entry, is calculated as in Equation 4.

$$\text{saliencypoints} = \textit{Attent}(x_{\text{previous}}) \tag{3}$$

$$x_{\text{current}} = \text{saliencypoints} \odot x_{\text{previous}} \tag{4}$$

The attention part of the proposed end-to-end framework is shown in detail in Figure 2. There are five convolutional layers, five batch normalization layers, five elu layers, and two max-pooling layers in the encoder section. All convolutional layers are 5×5 pixels in size, and the depth increases from 32 up to 128. The decoder section includes three deconvolutional layers, two batch normalization layers, and two elu layers. In the attention section, two attention blocks have the same layers. Thanks to its residual form of connection, it can protect low-frequency mainstream information.

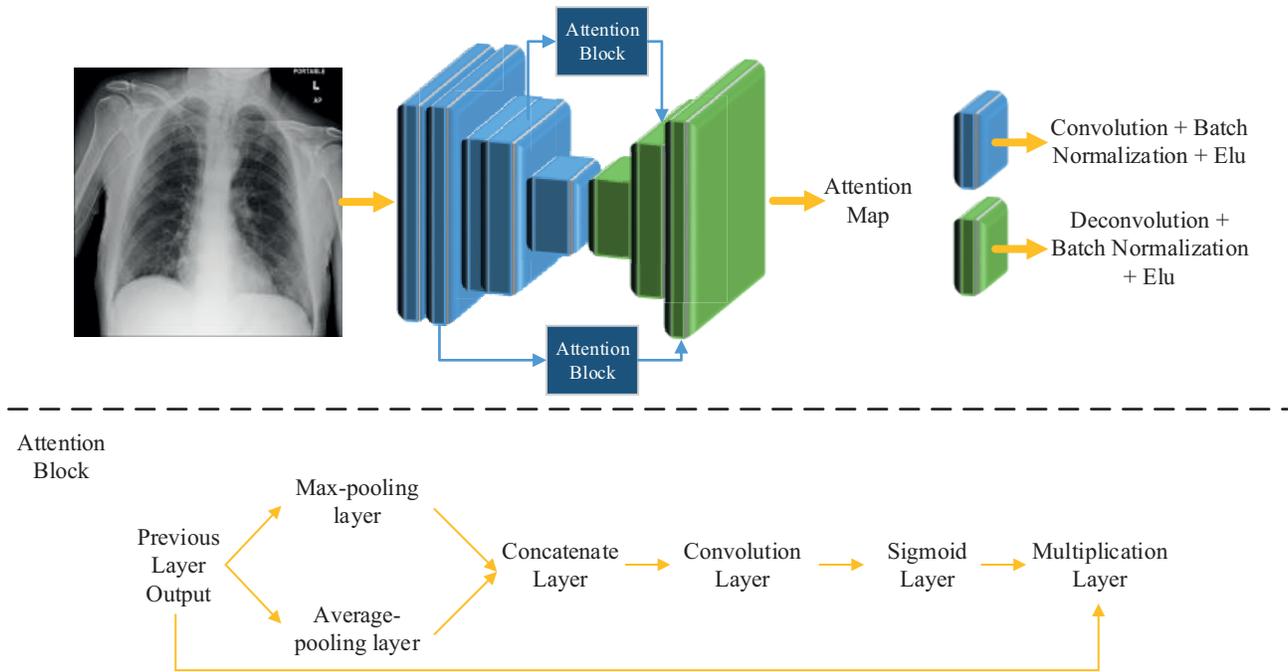


Figure 2. Attention CNN architecture

2.3. ResNet18 network for X-ray hashing

Pretrained ResNet18 architecture is transformed into a hashing network by modifying the input layers and output layers [24]. The original pixel information and the saliency map information are used as the input information for this section. At the network output, the existing fully connected layers are changed. The final layer size is determined according to the desired hash code length. One of the most critical parts of the output

section is selecting the number of fully connected layers (FCL). Choosing as few FCL layers as possible has positive effects on retrieval performance. For this reason, only one FCL layer before the hash layer is used in the proposed framework.

2.4. Loss function

The weighted cross-entropy loss function and the contrastive loss function are used for training the proposed framework. For this purpose, the contrastive loss function is calculated as in Equation 5.

$$L_c = \frac{1}{2}LD^2 + \frac{1}{2}(1-L)\{\max(0, m-D)\}^2 \tag{5}$$

where $D = \|f(I_1) - f(I_2)\|_2$

in which L represents pairwise labels. If two images belong to the same class, $L = 1$, if different, $L = 0$. $f(I_1)$, and $f(I_2)$ denote feature vectors. m is the margin. The other loss function is the weighted maximum likelihood function [25]. Let $H = [h_1, h_2, \dots, h_n]$, where $P(S|H)$ is the weighted likelihood function. Two loss functions are also compatible with pairwise training. In this way, the data imbalance problem can be tackled more effectively. The weighted likelihood problem is expressed as in Equation 6.

$$\min_{\theta} \sum_{s_{ij} \in S} w_{ij} (\log(1 + \exp(\alpha \langle h_i, h_j \rangle)) - \alpha s_{ij} \langle h_i, h_j \rangle) \tag{6}$$

where s_{ij} represents pairwise similarity information. θ represents all parameters in the proposed framework. w_{ij} denotes similar-dissimilar balance function from [25]. The total loss function is determined by combining Equations 5 and 6. Equation 7 is used for this purpose.

$$L_{total} = \alpha L_c + \beta L_w \tag{7}$$

where L_{total} denotes total loss, L_c denotes contrastive loss, L_w denotes weighted cross-entropy loss. α and β represent weights of loss values. Various optimization techniques can be used to determine these values. In this study, a value of 0.5 is chosen for both of them in order to examine the effects more fairly.

3. Experiments and experimental results

In this section, hardware information, experimental parameters, ablation study, and comparisons with current state-of-the-art methods are presented. In addition, the results of the proposed framework are visualized.

The proposed method is trained on a computer with Intel Core i9-10900K CPU (3.7 GHz), 64 GB DDR4 RAM, and NVIDIA GeForce RTX 3090 graphic card.

3.1. Dataset

In this study, the ChestX-ray8 dataset [26], which is the most comprehensive and large X-ray dataset in the literature, is used in order to evaluate the proposed framework appropriately. This dataset consists of a total of 112120 frontal X-ray images. It consists of images taken from 30805 unique patients, including 16630 males and 14175 females. Images were collected between 1992-2015 with the help of text-mined labels. The size of the images is 1024×1024 pixels, and the images are gray level. There are fifteen different classes in total, including the healthy images in the data. Sample images belonging to these classes can be seen in Figure 3.

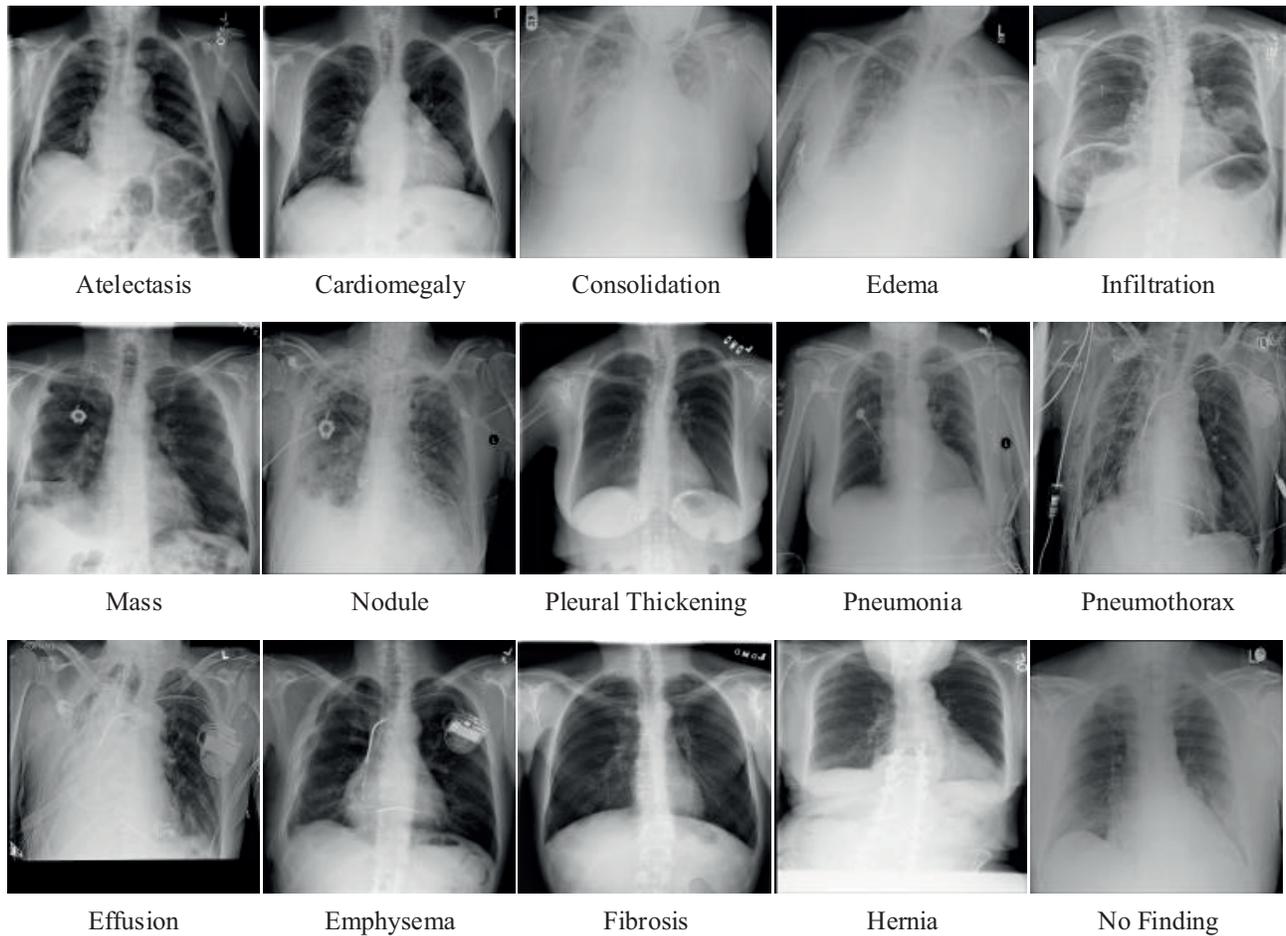


Figure 3. Sample images for each class of the ChestX-ray8 dataset.

ChestX-ray dataset images are relatively challenging. One of the main reasons for this is the difficulties due to the nature of medical images. The second difficulty is that problems belonging to more than one class are found together in some images. In such cases, if only one of the owned classes is estimated, the prediction is determined correctly. Another challenge is the imbalance of data numbers between classes. This imbalance is seen in Figure 4.

3.2. Experimental setup, parameters, and performance indicators

In this section, first of all, details about the parameters used in the experiments are presented. Then, the strategy for determining evaluation parameters is explained. Separation of the dataset for training and testing during experiments is carried out as specified in dataset rules. Training and validation separation is made in the form of 80-20. While the contrastive loss margin value is chosen as 1, the weighted cross entropy weights are determined according to the classes. The number of minibatches is used as 20. The number of training iterations is determined as 15,000. The initial learning rate is determined as 0.0001. This value is multiplied by 0.6 every 4000 iterations. Gradient decay is selected as 0.9. Finally, the Adam optimizer is used to update the parameters. Five-fold cross validation is applied.

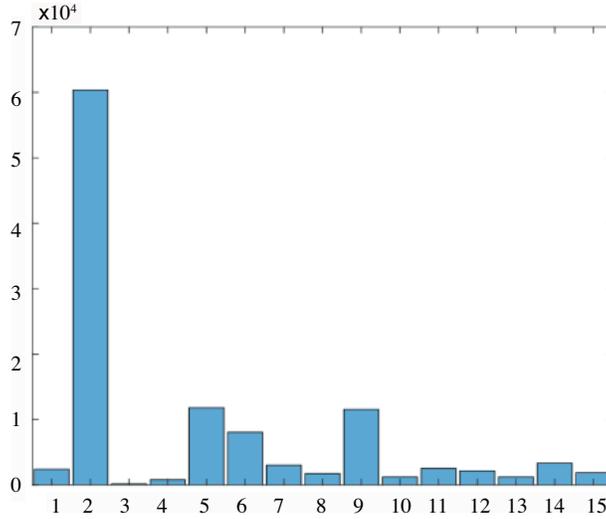


Figure 4. ChestX-ray dataset class imbalance ratio.

The imbalance problem in the data is quite challenging. In order to simply overcome this problem, an approach to batch selection is applied. Accordingly, each minibatch consisting of 20 images can contain up to three samples of each class. In this way, the influence of dominant classes can be brought to the same level as other classes.

In order to compare the performance of the proposed retrieval method with other methods in the literature, the most frequently used performance indicators are preferred in this study. Average precision (ARP), the discounted cumulative gain (DCG), and the normalized discounted cumulative gain (nDCG) are used. Equation 8 is used to calculate the precision value, Equation 9 is used to calculate the AP value, Equation 10 is used to calculate the DCG value, and Equation 11 is used to calculate the nDCG value.

$$P(I_q) = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \tag{8}$$

$$ARP = \frac{1}{n} \sum_{i=1}^n P(I_i) \tag{9}$$

$$DCG = \sum_{n=1}^K \frac{2^{r_n} - 1}{\log(n + 1)} \tag{10}$$

$$nDCG = \frac{DCG}{DCG_{ideal}} \tag{11}$$

3.3. Ablation study

One of the most important factors affecting the success of the proposed framework is the saliency maps suggested by the attention network. Saliency maps that cover the whole image or cover large areas cannot make a positive contribution to performance. Conversely, small saliency points may not usually identify suitable areas. In order

to avoid these situations, this study aims to identify meaningful chest regions and decision-making regions. In this section, the suitability of the proposed framework for the desired purposes is tested. First, the saliency map generation performances of residual attention blocks are compared. Figure 5 shows the saliency maps of architectures that do not contain attention blocks, with one attention block and with two attention blocks. It usually creates a scattered saliency map in experiments where no attention layer is used. It is challenging to understand what these attention points contain visually. Hashing networks generally cannot produce a high performance for this reason. In the architecture that uses a single attention layer, the saliency areas are pretty broad. Generally, important regions are within the saliency areas. However, so much unnecessary information can create complexity. On the other hand, some markings are emphasized too much by the single attention block. Architecture using two attention layers suggests fewer and appropriate saliency zones compared to other methods.

One of the other important factors that can affect attention performance is the activation function. Various activation functions in the literature are used for various tasks. However, the sigmoid function is generally preferred for attention tasks [27]. Therefore, this study examines the effects of sigmoid, hyperbolic tanh, and eLU activation functions on attention. For this purpose, saliency maps are being investigated, as in Figure 6.

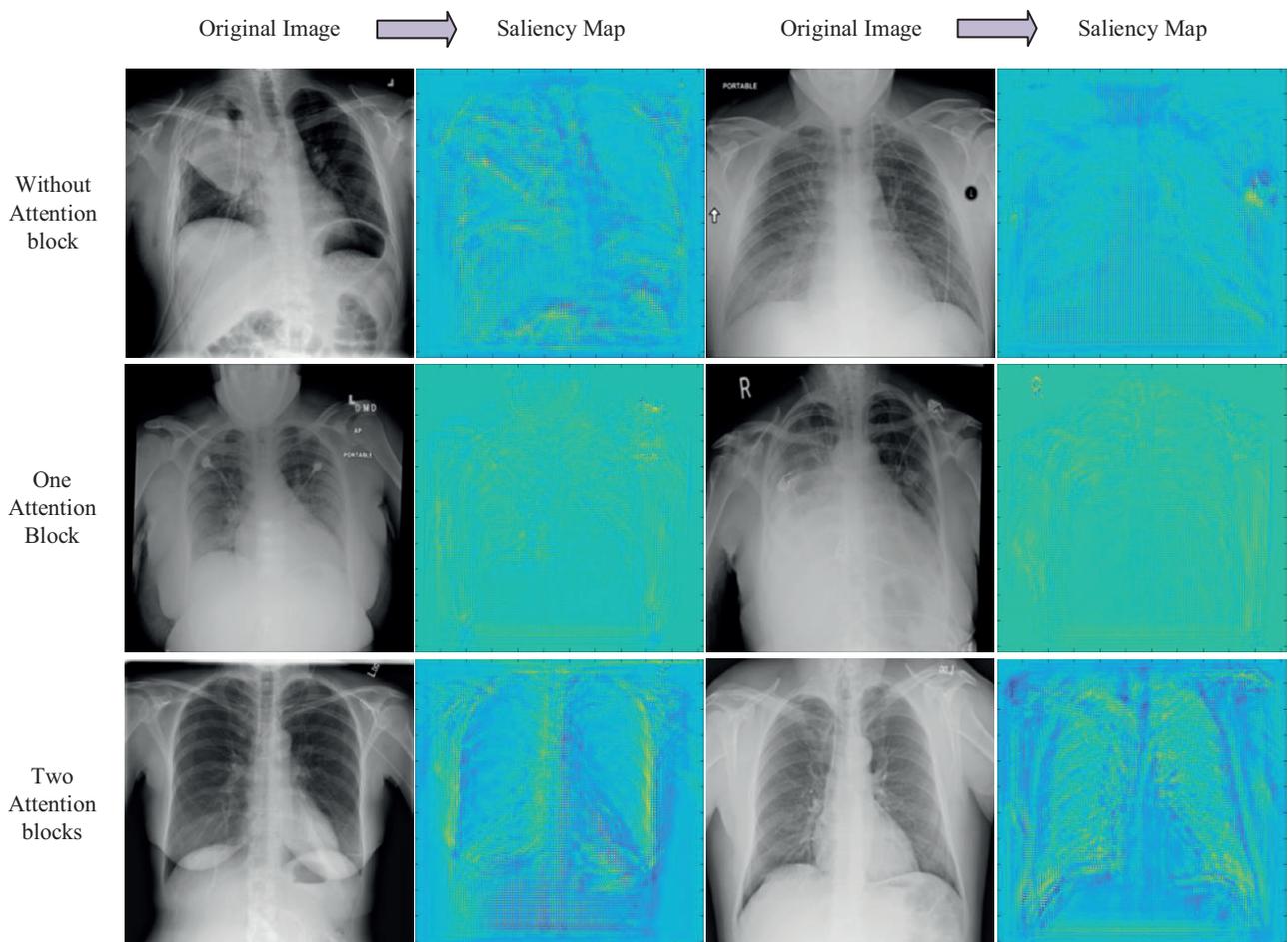


Figure 5. Attention block effect on the saliency maps.

Figure 6 shows some of the attention maps generated by the sigmoid, hyperbolic tanh, and eLU activation functions. It seems that the saliency maps produced by the Sigmoid activation function are more suitable for the main idea of the proposed approach. On the other hand, regions determined by saliency maps produced by tanh and eLU activation functions are also suitable. However, the number of regions to be considered is relatively high, and irrelevant regions are marked. These regions can create a false redirect for analysis to be performed by the hashing network. In spite of that, for a different strategy than the one followed in our framework, the saliency maps suggested by tanh or eLU activation functions may be appropriate. The standard sigmoid activation function, which is generally used in the literature, is preferred for this study.

Parametric analysis of the proposed method is shown in Table 1. The performance of retrieved images in response to a query image is presented in Table 1. For this purpose, each image from the test data is selected in order. Results of retrieved 10, 20, 30, 50, and 100 images are calculated for each image. Equation 9 is used for the realization of the calculations mentioned in the formula. It should not be forgotten that many images have more than one label in the ChestX-ray dataset.

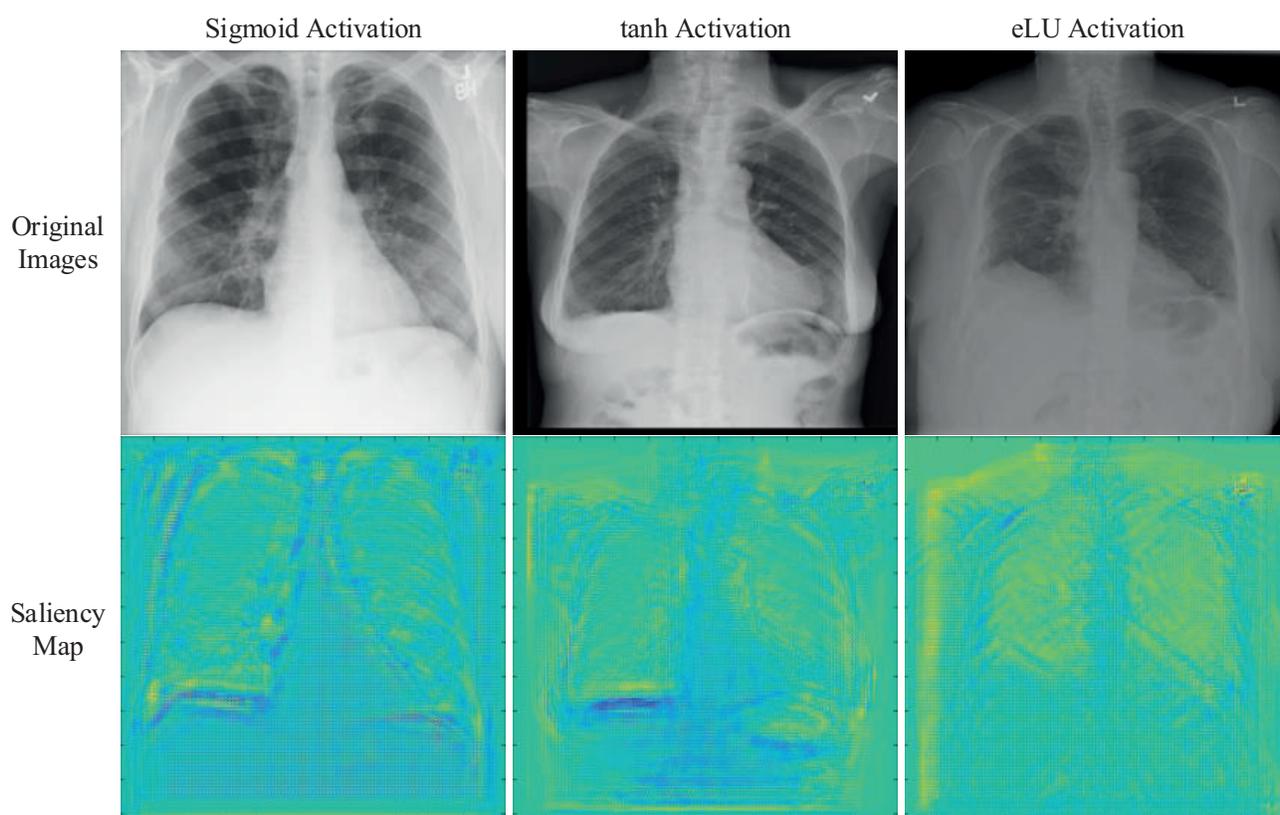


Figure 6. Attention maps from different activation functions.

Table 1 shows that the proposed framework is an effective solution. It increases precision performance by almost 5% compared to traditional architectures. The effect of the proposed attention part is examined piece by piece. The architecture with two attention blocks has almost 7% higher precision performance than the architecture without the attention block. Finally, the proposed architecture offers an almost 13% increase in performance compared to traditional architectures.

Table 1. Precision performance evaluation of the proposed method and variants.

Method	Retrieved images (P@)				
	10 (%)	20 (%)	30 (%)	50 (%)	100 (%)
ResNet18	56.27	55.87	54.74	53.11	51.28
Proposed Framework without Attention Block (PF0)	62.86	60.19	59.51	57.97	56.48
Proposed Framework with One Attention Block (PF1)	68.29	65.57	64.36	63.08	61.49
Proposed Framework with Two Attention Block (PF2)	70.33	67.63	66.43	65.14	63.53

3.4. Comparison with state-of-the-art

This section includes comparisons of the proposed framework with other current X-ray retrieval state-of-the-art methods in the literature. Firstly, P@10 performances, which are frequently used in the literature, are examined. Table 2 presents the results of the comparison of the P@10 performance of the proposed framework on the ChestX-ray data. For a fair comparison, studies that contain only deep learning and produce results only in the ChestX-ray dataset are preferred in Table 2. According to this, the variation with only one attention layer produces higher precision performance than other methods in the literature. The proposed framework, implemented with two attention layers, reveals a performance approximately 7% higher than the highest precision performance in the literature.

Table 2. Comparison of precision performance of the proposed framework with other ChestX-ray retrieval systems (methods from [21]).

	NIH_{LR}	NIH_{HR}	$NIH - U_{LR}$	$NIH - U_{HR}$	ResNet18	PF_0	PF_1	PF_2
P@10 (%)	48	52	58	63	56.27	62.86	68.29	70.33

Table 3 provides a broader comparison, including seven methods in the literature. In this section, nDCG results are evaluated as the comparison performance on a subset of the used dataset. As shown in Table 3, the proposed framework is superior to other methods with a value of 0.36 nDCG.

Table 3. Comparison of retrieval performance of the proposed framework (methods reported in [28]).

Method	nDCG
[29]	0.15
[30]	0.16
[31]	0.19
[32]	0.17
[28]	0.24
[33]	0.15
[21]	0.31
Proposed framework	0.36

4. Conclusion

This study provides a deep learning framework that can effectively generate hash code for X-ray images. The proposed framework is an end-to-end architecture consisting of two separate parts: the attention network and the hashing network. The attention network section, which is designed similar to FCN architectures, includes two attention layers. This is the first study in the literature to use residual attention connections for X-ray image retrieval and hashing tasks. Since the information in the early layers combines the saliency points with the high-level features, the attention network section works quite efficiently. This is one of the most important factors behind the success of the proposed framework. On the other hand, performing a specific normalization to X-ray images to suggest more than one saliency region with moderate probabilities may be another factor. In the second part of the framework, hash codes are obtained with RexNet18 architecture. In order to relieve the imbalance problem that many medical image datasets have, pairwise training is carried out using a special selection criterion. Furthermore, the effects of many layers on the saliency regions are investigated by performing extensive experiments. Finally, the results produced by the proposed framework are compared with the current state-of-the-art methods in the literature. The results obtained reveal that the proposed framework produces superior results in ChestX-ray dataset experiments compared to other methods. In future studies, X-ray hashing techniques that do not require label information will be studied. In addition, intensive studies are planned to increase the precision performance obtained.

Conflict of interest

The authors declare that they have no conflicts of interest.

Human and animal rights

The paper does not contain any studies with human participants or animals.

Acknowledgment

This research is funded by Scientific and Technological Research Council of Turkey (TÜBİTAK) under grant number 120E018.

References

- [1] Domingues I, Pereira G, Martins P, Duarte H, Santos J et al. Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET. *Artificial Intelligence Review* 2019; 53(6): 4093-160. doi: 10.1007/s10462-019-09788-3
- [2] Gates M, Wingert A, Featherstone R, Samuels C, Simon C et al. Impact of fatigue and insufficient sleep on physician and patient outcomes: a systematic review. *BMJ Open* 2018; 8 (9). doi: 10.1136/bmjopen-2018-021967
- [3] Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M et al. Medical Image Analysis using Convolutional Neural Networks: A Review. *Journal of Medical Systems* 2018; 42 (11). doi: 10.1007/s10916-018-1088-1
- [4] Öztürk Ş. Stacked auto-encoder based tagging with deep features for content-based medical image retrieval. *Expert Systems with Applications* 2020; 161. doi: 10.1016/j.eswa.2020.113693
- [5] Pan Z, Wang L, Wang Y, Liu Y. Product quantization with dual codebooks for approximate nearest neighbor search. *Neurocomputing* 2020; 401: 59-68. doi: 10.1016/j.neucom.2020.03.016
- [6] Shen HT, Liu L, Yang Y, Xu X, Huang Z et al. Exploiting Subspace Relation in Semantic Labels for Cross-modal Hashing. *IEEE Transactions on Knowledge and Data Engineering*. 2020. doi: 10.1109/tkde.2020.2970050

- [7] Öztürk Ş. Comparison of Pairwise Similarity Distance Methods for Effective Hashing. IOP Conference Series: Materials Science and Engineering 2021; 1099 (1). doi: 10.1088/1757-899x/1099/1/012072
- [8] Öztürk Ş. Class-driven content-based medical image retrieval using hash codes of deep features. Biomedical Signal Processing and Control 2021; 68. doi: 10.1016/j.bspc.2021.102601
- [9] Nie X, Zhou X, Shi Y, Sun J, Yin Y. Classification-enhancement deep hashing for large-scale video retrieval. Applied Soft Computing 2021. doi: 10.1016/j.asoc.2021.107467
- [10] Bhatt C, Kumar I, Vijayakumar V, Singh KU, Kumar A. The state of the art of deep learning models in medical science and their challenges. Multimedia Systems 2020. doi: 10.1007/s00530-020-00694-1
- [11] Nowaková J, Prílepok M, Snášel V. Medical Image Retrieval Using Vector Quantization and Fuzzy S-tree. Journal of Medical Systems 2016; 41(2). doi: 10.1007/s10916-016-0659-2
- [12] Shamna P, Govindan VK, Abdul Nazeer KA. Content based medical image retrieval using topic and location model. Journal of Biomedical Informatics 2019; 91. doi: 10.1016/j.jbi.2019.103112
- [13] Shamna P, Govindan VK, Abdul Nazeer KA. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University - Computer and Information Sciences 2018. doi: 10.1016/j.jksuci.2018.10.002
- [14] Khatami A, Babaie M, Tizhoosh HR, Khosravi A, Nguyen T et al. A sequential search-space shrinking using CNN transfer learning and a Radon projection pool for medical image retrieval. Expert Systems with Applications 2018; 100: 224-33. doi: 10.1016/j.eswa.2018.01.056
- [15] Ahmad J, Muhammad K, Baik SW. Medical Image Retrieval with Compact Binary Codes Generated in Frequency Domain Using Highly Reactive Convolutional Features. Journal of Medical Systems 2017; 42 (2). doi: 10.1007/s10916-017-0875-4
- [16] Song T, Camalan S, Niazi MKK, Moberly AC, Teknos T et al. OtoMatch: Content-based eardrum image retrieval using deep learning. Plos One 2020; 15 (5). doi: 10.1371/journal.pone.0232776
- [17] Dureja A, Pahwa P. Medical image retrieval for detecting pneumonia using binary classification with deep convolutional neural networks. Journal of Information and Optimization Sciences 2020; 41 (6): 1419-31. doi: 10.1080/02522667.2020.1809096
- [18] Wang S-H, Govindaraj VV, Górriz JM, Zhang X, Zhang Y-D. Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. Information Fusion 2021; 67: 208-29. doi: 10.1016/j.inffus.2020.10.004
- [19] Babaie M, Tizhoosh HR, Zhu S, Shiri MEJae-p. Retrieving similar X-ray images from big image data using radon barcodes with single Projections2017 January 01, 2017: [arXiv: 1701.00449 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2017arXiv170100449B>.
- [20] Chung Y-A, Weng W-HJae-p. Learning deep representations of medical images using siamese CNNs with application to content-based image Retrieval2017 November 01, 2017: [arXiv: 1711.08490 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2017arXiv171108490C>.
- [21] Haq NF, Moradi M, Wang ZJ. A deep community based approach for large scale content based X-ray image retrieval. Medical Image Analysis 2021; 68. doi: 10.1016/j.media.2020.101847
- [22] Guan Q, Huang Y, Zhong Z, Zheng Z, Zheng L et al. Thorax disease classification with attention guided convolutional neural network. Pattern Recognition Letters 2020; 131: 38-45. doi: 10.1016/j.patrec.2019.11.040
- [23] Jin S, Yao H, Sun X, Zhou S, Zhang L et al. Deep saliency hashing for fine-grained retrieval. IEEE Transactions on Image Processing 2020; 29: 5336-51. doi: 10.1109/tip.2020.2971105
- [24] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE conference on computer vision and pattern recognition (CVPR)2016. p. 770-778. doi: 10.1109/cvpr.2016.90

- [25] Cao Z, Long M, Wang J, Yu PS. HashNet: Deep Learning to Hash by Continuation. 2017 IEEE International Conference on Computer Vision (ICCV)2017. p. 5609-5618. doi: 10.1109/iccv.2017.598
- [26] Wang X, Peng Y, Lu L, Lu Z, Bagheri M et al. ChestX-Ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)2017. p. 3462-3471. doi: 10.1109/cvpr.2017.369
- [27] Wang Y, Ji S, Lu M, Zhang Y. Attention boosted bilinear pooling for remote sensing image retrieval. *International Journal of Remote Sensing* 2019; 41(7): 2704-2724. doi: 10.1080/01431161.2019.1697010
- [28] Chen Z, Cai R, Lu J, Feng J, Zhou J. Order-sensitive deep hashing for multimorbidity medical image retrieval. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Lecture Notes in Computer Science*2018. p. 620-628. doi: 10.1007/978-3-030-00928-1_70
- [29] Jun W, Kumar S, Shih-Fu C. Semi-supervised hashing for large-scale search. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2012; 34 (12): 2393-2406. doi: 10.1109/tpami.2012.48
- [30] Gong Y, Lazebnik S. Iterative quantization: A procrustean approach to learning binary codes. *Cvpr 2011*2011. p. 817-824. doi: 10.1109/cvpr.2011.5995432
- [31] Liong VE, Jiwen L, Gang W, Moulin P, Jie Z. Deep hashing for compact binary codes learning. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)2015. p. 2475-2483. doi: 10.1109/cvpr.2015.7298862
- [32] Liu H, Wang R, Shan S, Chen X. Deep Supervised Hashing for Fast Image Retrieval. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)2016. p. 2064-2072. doi: 10.1109/cvpr.2016.227
- [33] Lan R, Zhong S, Liu Z, Shi Z, Luo X. A simple texture feature for retrieval of medical images. *Multimedia Tools and Applications* 2017; 77 (9): 10853-10866. doi: 10.1007/s11042-017-5341-2