Content-based Image Retrieval for Breast Ultrasound Images using Convolutional Autoencoders: A Feasibility Study

Mohammad I. Daoud, Amro Saleh, Ismail Hababeh, and Rami Alazrai
School Electrical Engineering and Information Technology, German Jordanian University
Amman, Jordan
{mohammad.aldaoud, a.saleh1, ismail.hababeh, rami.azrai }@gju.edu.jo

***Abstract*—Ultrasound imaging is one of the most widely used medical imaging modalities for detecting breast cancer. However, the accuracy of diagnosing the tumors in breast ultrasound (BUS) images might vary based on the experience level of the radiologist. Content-based image retrieval (CBIR) systems can be employed to improve the diagnosis accuracy by providing the radiologist with BUS images of previous, clinically relevant cases. In this study, a CBIR system is developed based on deep learning technology to support the diagnosis of BUS images. In particular, each query BUS image submitted to the system is analyzed using a custom-made convolutional autoencoder (CAE) to extract a latent features vector that represents the image. An important advantage of the proposed CAE is its ability to extract the latent features vector without the need to localize or outline the tumor. The latent features vector of the query image is analyzed using a similarity measure to identify and retrieve the most relevant BUS images from a reference BUS image database. Finally, the retrieved BUS images are displayed to the radiologist. The performance of the proposed CBIR system has been evaluated using a set of BUS images that includes benign and malignant breast tumors. The results reported in this study suggest the feasibility of employing the proposed CBIR system to improve the diagnosis of BUS images.**

***Index Terms*—Ultrasound imaging, breast cancer, content-based image retrieval, convolutional autoencoder**

I. Introduction

Breast cancer is the most common cancer in women and it ranks among the leading causes of death in females [1]. According to the World Health Organization, in 2018 around 627,000 female deaths were related to breast cancer [1]. Due to the advantages of low-cost, accuracy, and safety, ultrasound imaging is commonly used for breast cancer diagnosis. However, the interpretation of the breast ultrasound (BUS) images to diagnose the tumor is operator-dependent and can vary based on the experience level of the radiologist [2, 3, 4, 5].

One approach to improve the diagnosis of BUS images is to develop computer-aided diagnosis systems, such as [6, 7, 8, 9, 10], to analyze the BUS images. The interpretation of BUS images can also be improved by developing computational models, such as [11], to related the characteristics of BUS images to tissue anatomy. Another effective approach to enhance the diagnosis accuracy of BUS images is to develop content-based image retrieval (CBIR) systems, such as [12, 13], that provide the radiologist with reference BUS images of previous, clinically relevant cases that have known diagnostic results. Such CBIR systems allow the radiologist to compare the BUS image of interest, denoted by the query image, with relevant BUS images retrieved from a reference BUS image database. The majority of the CBIR systems that are proposed for BUS images involve the extraction of hand-crafted features, such as texture features, from the query image and the use of these features to search for the most relevant BUS images in the reference database. In fact, the process of extracting the hand-crafted features usually requires outlining the tumor in the BUS image or identifying a region of interest (ROI) that includes the tumor. Hence, the capability of retrieving reference BUS images that are relevant to the query image depends on the accuracy of the tumor outline or the ability to identify a ROI that includes the tumor.

In this study, we propose a CBIR system for BUS images based on deep learning technology. The proposed CBIR system can extract features to represent the BUS image without the need to outline the tumor or identify a ROI that includes the tumor. In particular, a custom-made convolutional autoencoder (CAE) is developed to extract a vector of 32 latent features from the BUS image without the requirement of preprocessing or segmenting the image. Hence, when a query BUS image is submitted to the CBIR system, the CAE is used to extract a vector of latent features that represents the image and a similarity measure is employed to analyze the latent features vector to retrieve the most relevant BUS images from the reference database. The performance of the proposed CBIR system is evaluated using a database of BUS images that includes both benign and malignant tumors.

The remainder of the paper is organized as follows. Section II describes the proposed CBIR system. The performance evaluation of the proposed CBIR system and the experimental results are provided in Section III. Finally, the conclusion and future directions are presented in Section IV.

II. The proposed CBIR system for BUS images

Figure 1 illustrates the architecture of the proposed CBIR system. The system has two components: an offline component and an online component. In the offline component, the BUS images in the reference database are analyzed using a convolutional autoencoder (CAE) to extract a vector of 32 latent features from each image. The extracted latent features vectors

978-1-7281-3578-6/19/$31.00 ©2019 IEEE

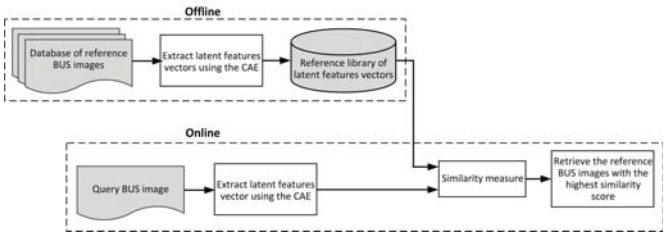

Fig. 1. The architecture of the CBIR system that is proposed for BUS images.

are stored in a reference library. During the online operation of the CBIR system, when a query BUS image is submitted, the CAE is used to extract a vector of 32 latent features from the query image. Furthermore, a similarity measure is employed to evaluate the similarity score between the latent features vector of the query BUS image and the latent features vectors in the reference library. The latent features vectors in the reference library that have the highest similarity score compared with the latent features vector of the query BUS image are identified. Finally, the reference BUS images that correspond to the identified reference latent features vectors are retrieved and displayed. In the following two subsections, we describe the CAE and the similarity measure that are employed in the proposed CBIR system.

A. The convolutional autoencoder (CAE)

In general, the autoencoder aims to represent the input image x using a condensed representation, called the code z. In conventional autoencoders, this process is carried out using two components: the encoder and the decoder. The encoder, denoted by $f(x)$, transforms the input image, x, into the condensed representation, z, such that the dimensionality of z is smaller than x. Furthermore, the decoder, $g(z)$, analyzes the condensed representation, z, to reconstruct a copy of the image, denoted by $\hat{x}$. Hence, the reconstructed copy of the image can be expressed as $\hat{x} = g(f(x))$. The structure and parameters of the autoencoder are usually optimized to minimize the reconstruction error between x and $\hat{x}$ over a set of training images. The convolutional autoencoder (CAE) extends the operation of the autoencoder by employing deep learning technology to find a condensed, latent representation of the input image. A major advantage of CAEs is their capability to preserve the spatial locality of the image [14].

In this study, we propose a CAE that enables the extraction of a condensed vector of latent features for representing the BUS image. The structure of the proposed CAE is shown in Figs. 2. In this architecture, the input BUS image, x, is encoded to generate the latent code z, which corresponds to a vector that includes 32 latent features. In fact, the encoding of x to generate z is performed by processing x using four convolution layers, a flattening layer, and a fully connected layer. The four convolution layers have a filter size of 3×3 and a stride of 2 applied along the horizontal and vertical dimensions. The number of filters in each convolution layer varies between 43 and 6, as shown in Fig. 2. The outputs of each convolution layer are processed using the Rectified Linear Units (ReLU) activation function. After applying the four convolution layers to x, a flatting layer is employed to convert the two-dimensional (2D) feature maps that are generated by the fourth convolution layer to a vector of 384 latent features. Finally, a fully connected layer is applied to generate the condensed, latent representation of the image (z) that includes 32 features. The feature vector, z, is decoded to reconstruct the image, $\hat{x}$, by applying the reverse of the operations that were employed to map x to z. These reverse operations include a fully connected layer that increases the size of the feature vector, a reshape operation that converts the feature vector into 2D feature maps, and four convolutional transpose (deconvolution) layers.

The CAE can be trained in an end-to-end, unsupervised manner using a set of BUS images to minimize the mean squared error (MSE), which can be expressed as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\|x_i - \hat{x_i}\|^2 \quad (1)$$

where N is the total number of BUS images used for training. Moreover, x_i and $\hat{x_i}$ are the i^{th} BUS image used for training and the matching reconstructed image, respectively. The operator $\|.\|$ is the norm operator.

A major advantage of the CAE in Fig. 2 is the constrained dimensionality of the latent code, z, which is smaller than the dimensionality of the input BUS image. In fact, the reduced dimensionality of z allows the CAE to learn salient features that enable effective representation of the BUS image. Another crucial advantage of the proposed CAE is the capability to extract a latent features vector from the BUS image without the need to segment the image.

B. Similarity measure

In this study, the cosine measure [15] is employed to evaluate the similarity score between the latent features vector, z_q, of the query BUS image and a given latent features vector, z_r, in the reference library. In fact, the cosine measure evaluates the angle between the vectors z_q and z_r as follows:

$$cos(z_q, z_r) = \frac{z_q.z_r}{\|z_q\|\|z_r\|} \quad (2)$$

where $z_q.z_r$ is the dot product between the vector z_q and the vector z_r and $\|.\|$ is the norm operator.

III. Performance analysis and results

A. The BUS image database

The BUS image database employed in the current study is composed of 235 BUS images (113 benign and 122 malignant). In fact, the database is formed by combining two datasets. The first dataset includes BUS images that were acquired at the Jordan University Hospital, Amman, Jordan for 110 female patients during the years between 2012 and 2016. For each patient, one BUS image was included in the dataset. Based on biopsy findings, this dataset is composed of 64 images of benign tumors and 46 images of malignant tumors. Each BUS image included exactly one tumor. The mean $\pm$ standard deviation diameter of the tumors was 14.7 $\pm$ 6.0 mm. The ultrasound system used to acquired the images is the Acuson S2000 system (Siemens AG, Munich, Germany)

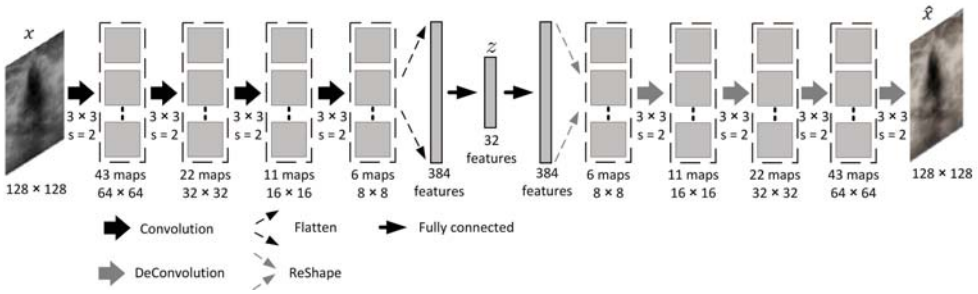

Fig. 2. The structure of the CAE employed by the proposed CBIR system.

equipped with the 14L5 transducer. The resolution of the BUS images was 0.1 mm $\times$ 0.1 mm. The Ethics Committee at the Jordan University Hospital approved the acquisition of the BUS images and the patients provided written consent to the acquisition protocol. The second dataset is composed of 125 BUS images obtained from an online database (http://www.onlinemedicalimages.com) that is commonly used for evaluating the performance of ultrasound image analysis algorithms [16]. The 125 BUS are composed of 49 BUS of benign tumors and 76 BUS images of malignant tumors.

B. Performance evaluation procedure

A five-fold cross validation procedure is used to evaluate the performance of the proposed CBIR system. In this procedure, the BUS image database is divided into five groups, such that each group includes 47 images. In the i^{th} iteration of the cross validation procedure, the image group number i is used as query BUS images and the remaining four groups are combined and employed as the reference BUS image database. The CAE is trained based on the images in the reference database. Furthermore, the trained CAE is used to analyze each query BUS image to extract a latent features vector, which is used to retrieve the most relevant k images from the reference database. This validation procedure is repeated for five iterations to vary the value of i between 1 and 5 with an increment of 1. Using this five-fold cross validation procedure, all 235 images in the BUS image database are employed as query images. The validation procedure is performed using three different values of k, which are 3, 7, and 10. Hence, the performance evaluation is carried out when the CBIR system is configured to retrieve the most relevant 3, 7, and 10 images compared to the query image.

In each iteration of the cross validation procedure, only 188 BUS mages (4 image groups $\times$ 47 image in each group) are included in the reference database. This limited size of the reference database might affect the training of the CAE. To address this potential limitation, we have employed data augmentation to extend the training database. In fact, data augmentation has been used in many previous studies, such as [17], as a solution to increase the number of training images. In the current study, we have employed two augmentation operations: the rotation and transformation. In the rotation-based augmentation, each BUS image in the reference database is rotated using five angles that are randomly selected from the range of $[1°, 180°]$. In the translation-based augmentation, each BUS image in the reference database is translated five times along the horizontal and vertical directions using distances that are randomly generated from the range of $[-20, 20]$ pixels. Hence, the training dataset is composed 2068 images that include the reference dataset in addition to the images generated using the five rotation-based and five translation-based augmentations. It is worth nothing that the images in the reference dataset are not affected by data augmentation.

The performance of the proposed CBIR system is evaluated using the precision metric. The precision is defined as the ratio of the number of retrieved BUS images that are relevant to the query image to the total number of retrieved images. In this study, if the class (benign or malignant) of the retrieved BUS image matches the class of the query image, then the retrieved image is considered to be relevant to the query image.

C. Results

Table I shows the mean $\pm$ standard deviation precision obtained when the proposed CBIR system is employed to retrieve the 3, 7, and 10 most relevant BUS images from the reference database for malignant query images, benign query images, and all (both benign and malignant) query images. In fact, the mean precision values are higher than 80%. Also, the capability of the system to retrieve relevant images is generally higher for benign query images compared to malignant query images. Table I also shows that increasing the number of retrieved images leads to higher precision values.

Figure 3 shows examples of the BUS images retrieved by the proposed CBIR system. In particular, Figs. 3(a) and 3(b) show a malignant query BUS image and the 3 most relevant images retrieved by the system, respectively. Furthermore, Figs. 3(c) and 3(d) show a benign query BUS image and the 3 most relevant images retrieved by the system, respectively. As shown in the figure, the retrieved images reveal large similarity with the corresponding query images.

IV. Conclusion

A CBIR system is proposed for BUS images based on deep learning technology. Specifically, a CAE is developed to extract a vector of 32 latent features for representing the BUS image. A major advantage of the developed CAE is the capability of extracting the latent features vector without the need to segment the BUS image. Moreover, the CAE can be trained in an unsupervised, end-to-end manner. When a query BUS image is submitted to the CBIR system, the CAE is employed to extract a latent features vector to represent the image and the

TABLE I
THE MEAN ± STANDARD DEVIATION PRECISION OBTAINED WHEN THE NUMBER OF RETRIEVED IMAGES, k, IS 3, 7, AND 10.

	k = **3**	k = **7**	k = **10**
Benign tumors	$81.1\% \pm 18.9\%$	$84.3\% \pm 12.5\%$	$84.8\% \pm 10.2\%$
Malignant tumors	$82.0\% \pm 19.2\%$	$82.9\% \pm 12.8\%$	$83.3\% \pm 10.0\%$
All tumors	$81.5\% \pm 19.0\%$	$83.6\% \pm 12.6\%$	$84.0\% \pm 10.1\%$

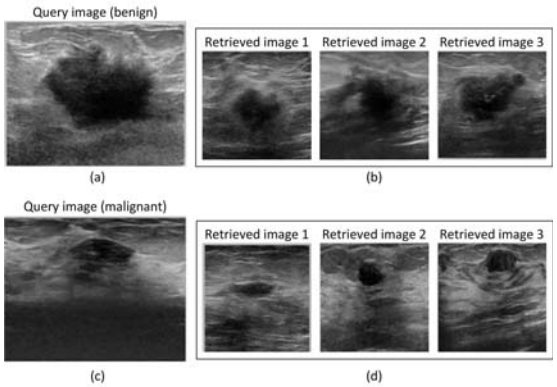

Fig. 3. (a) and (b) A malignant query BUS image and the corresponding 3 most relevant images retrieved by the CBIR system, respectively. (c) and (d) A benign query BUS image and the corresponding 3 most relevant images retrieved by the system, respectively.

similarity measure is used to analyze the latent features vector to retrieve the most relevant BUS images from the reference database. The results reported in the current study demonstrate the potential of employing the proposed CAE to achieve an effective CBIR system for retrieving relevant BUS images. In the future, we are planning to improve the evaluation of the proposed system by comparing the reference BUS images that are retrieved by the system with matching reference BUS images identified by experienced radiologists.

V. ACKNOWLEDGMENT

Dr. Mahasen Al-Najar from the Jordan University Hospital, Amman, Jordan is acknowledged for providing the first dataset of BUS images.

REFERENCES

[1] World Health Organization, Geneva, Switzerland, "Breast cancer," accessed: 2019-02-27. [Online]. Available: https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/

[2] A. Jalalian, S. B. Mashohor *et al.*, "Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review," *Clinical Imaging*, vol. 37, pp. 420–420, 2013.

[3] M. I. Daoud, A. A. Atallah *et al.*, "Automatic superpixel-based segmentation method for breast ultrasound images," *Expert Systems with Applications*, vol. 121, pp. 78–96, 2019.

[4] M. I. Daoud, M. M. Baba *et al.*, "Accurate segmentation of breast tumors in ultrasound images using a custom-made active contour model and signal-to-noise ratio variations," in *Proc. Eighth International Conference on Signal Image Technology and Internet Based Systems*, 2012, pp. 137–141.

[5] M. I. Daoud, A. A. Atallah *et al.*, "Accurate and fully automatic segmentation of breast ultrasound images by combining image boundary and region information," in *Proc. IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 2016, pp. 718–721.

[6] M. I. Daoud, T. M. Bdair *et al.*, "A fusion-based approach for breast ultrasound image classification using multiple-ROI texture and morphological analyses," *Computational and Mathematical Methods in Medicine*, vol. 2016, article ID 6740956, 12 pages, 2016.

[7] J. Shan, S. K. Alam *et al.*, "Computer-aided diagnosis for breast ultrasound using computerized BI-RADS features and machine learning methods," *Ultrasound in Medicine & Biology*, vol. 42, pp. 980–988, 2016.

[8] F. Imani, M. Z. Wu *et al.*, "Monitoring of tissue ablation using time series of ultrasound RF data," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2011, pp. 379–386.

[9] M. I. Daoud, P. Mousavi *et al.*, "Computer-aided tissue characterization using ultrasound-induced thermal effects: analytical formulation and in-vitro animal study," in *Proc. SPIE*, vol. 7968, 2011, pp. 79 680G–1 – 79 680G–6.

[10] F. Imani, M. Daoud *et al.*, "Tissue classification using depth-dependent ultrasound time series analysis: in-vitro animal study," in *Proc. SPIE*, vol. 7968, 2011, pp. 79 680F–1 – 79 680F–7.

[11] M. I. Daoud and J. C. Lacefield, "Stochastic modeling of normal and tumor tissue microstructure for high-frequency ultrasound imaging simulations," *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 2806–2815, 2009.

[12] H. C. Cho, L. Hadjiiski *et al.*, "A similarity study of content-based image retrieval system for breast cancer using decision tree," *Medical Physics*, vol. 40, pp. 012 901/1–012 901/13, 2013.

[13] C. Muramatsu, T. Takahashi *et al.*, "Similar image retrieval of breast masses on ultrasonography using subjective data and multidimensional scaling," in *Proc. International Workshop on Breast Imaging*, 2016, pp. 43–50.

[14] J. Masci, U. Meier *et al.*, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. International Conference on Artificial Neural Networks (ICANN 2011)*, 2011, pp. 52–59.

[15] B. Zheng, A. Lu *et al.*, "A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment," *Medical Physics*, vol. 33, pp. 111–117, 2006.

[16] A. Rodtook and S. S. Makhanov, "Multi-feature gradient vector flow snakes for adaptive segmentation of the ultrasound images of breast cancer," *Journal of Visual Communication and Image Representation*, vol. 24, pp. 1414–1430, 2013.

[17] K. Lekadir, A. Galimzianova *et al.*, "A convolutional neural network for automatic characterization of plaque composition in carotid ultrasound," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, pp. 48 –55, 2017.