

Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network

Michał Byra^{a,b,*}, Piotr Jarosik^c, Aleksandra Szubert^d, Michael Galperin^e, Haydee Ojeda-Fournier^b, Linda Olson^b, Mary O'Boyle^b, Christopher Comstock^f, Michael Andre^b

^a Department of Ultrasound, Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland

^b Department of Radiology, University of California, San Diego, USA

^c Department of Information and Computational Science, Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland

^d Maria Skłodowska-Curie Memorial Cancer Centre and Institute of Oncology, Warsaw, Poland

^e Almen Laboratories, Vista, USA

^f Memorial Sloan-Kettering Cancer Center, New York, USA

ARTICLE INFO

Article history:

Received 2 February 2020

Received in revised form 3 May 2020

Accepted 5 June 2020

Available online 26 June 2020

Keywords:

Attention mechanism

Breast mass segmentation

Convolutional neural networks

Deep learning

Receptive field

Ultrasound imaging

ABSTRACT

In this work, we propose a deep learning method for breast mass segmentation in ultrasound (US). Variations in breast mass size and image characteristics make the automatic segmentation difficult. To address this issue, we developed a selective kernel (SK) U-Net convolutional neural network. The aim of the SKs was to adjust network's receptive fields via an attention mechanism, and fuse feature maps extracted with dilated and conventional convolutions. The proposed method was developed and evaluated using US images collected from 882 breast masses. Moreover, we used three datasets of US images collected at different medical centers for testing (893 US images). On our test set of 150 US images, the SK-U-Net achieved mean Dice score of 0.826, and outperformed regular U-Net, Dice score of 0.778. When evaluated on three separate datasets, the proposed method yielded mean Dice scores ranging from 0.646 to 0.780. Additional fine-tuning of our better-performing model with data collected at different centers improved mean Dice scores by ~6%. SK-U-Net utilized both dilated and regular convolutions to process US images. We found strong correlation, Spearman's rank coefficient of 0.7, between the utilization of dilated convolutions and breast mass size in the case of network's expansion path. Our study shows the usefulness of deep learning methods for breast mass segmentation. SK-U-Net implementation and pre-trained weights can be found at github.com/mbyr/bus_seg.

© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Breast cancer is the most common invasive cancer in women [1]. Ultrasound (US) imaging has been widely used for the breast mass evaluation. In comparison to other medical imaging modalities, for instance magnetic resonance imaging, US is highly accessible and inexpensive, and when applied by expert radiologists it can accurately differentiate malignant and benign breast masses.

Assessment of breast US images requires extensive knowledge of characteristic image features shown to be related to benign or malignant breast masses. Various computer aided diagnosis

systems (CAD) have been developed to aid radiologists with the interpretation of breast US images [2]. Mass segmentation is an important step in CAD systems since accurate segmentation enables better analysis of features related to breast mass shape. However, automatic segmentation in US imaging is considered a challenging task due to relatively low US image contrast, speckle noise and large variations in breast mass sizes and shapes [3,4]. Recently, deep learning algorithms are showing promise for breast mass image analysis. These effective data driven methods process input images to learn high level image representations and calculate, for instance, the segmentation mask or classification decision [5]. Convolutional neural networks (CNNs) have been successfully applied for the detection, segmentation and classification of breast masses in US [6–14].

Several deep learning based approaches have been investigated for the breast mass segmentation. Yap et al. applied

* Corresponding author at: Department of Ultrasound, Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland.
E-mail address: mbyra@ippt.pan.pl (M. Byra).

transfer learning to develop fully convolutional networks for breast mass segmentation, and achieved good automatic segmentation performance [10]. Similarly, Xu et al. investigated the usefulness of fully convolutional networks and U-Net for breast mass segmentation [13]. They presented that utilization of dilated convolutions at deeper layers of fully convolutional networks may help improve segmentation performance. Dilated convolutions increase network's receptive field, resulting in more efficient extraction of spatial details in comparison to conventional convolutions [13]. Han et al. proposed a semi-supervised approach to the development of segmentation networks [14]. They used generative adversarial networks to guide a fully convolutional neural network to generate more accurate segmentation maps [14].

In this work, we propose a novel variation of U-Net model for breast mass segmentation in US. U-Net is perhaps the most popular CNN for semantic object segmentation [15,16]. Standard U-Net architecture consists of contracting and expanding paths. First, in the contraction path (encoder) input image is processed using convolutional and pooling operators to produce a compressed image representation. Second, the representation is upsampled with convolutional operators in the expansion path (decoder) to produce the segmentation mask indicating object location. Additionally, skip connections are used to propagate feature maps from the contraction to expansion path [15,17]. However, standard U-Nets utilize convolutions of fixed receptive field. The segmentation method proposed in this study is based on selective kernels (SKs) that can automatically adjust network's receptive field via an attention mechanism, and mix feature maps extracted with both dilated and conventional convolutions [18]. Li et al. have shown that classification networks utilizing SKs can better recognize objects presented at different scales, and achieve better performance on the ImageNet dataset [18]. Here, we investigate whether a replacement of conventional convolutional blocks in U-Net architecture with SK blocks may help improve segmentation performance. Our network, named SK-U-Net, can automatically adjust receptive field and more efficiently utilize spatial information extracted at different scales, resulting in better segmentation of breast masses than the previous models that utilized fixed receptive fields [10,13,14]. We show that dilated convolutions are mainly utilized in the expansion path of the network to generate segmentations for larger breast masses, efficiently addressing the problem of large variations in breast mass sizes. Moreover, our study presents the robustness of the proposed deep learning segmentation method, which achieved good performance on three datasets of breast mass US images collected at different medical centers.

2. Materials and methods

2.1. Ultrasound data

To develop the deep learning segmentation methods we used a dataset of 882 breast mass US images, consisting of 678 benign and 204 malignant breast masses. The dataset was divided into training, validation and test sets with a 632/100/150 split. The ratio of malignant and benign breast masses was maintained for each set. The Health Insurance Portability and Accountability Act compliant retrospective study was approved by the Human Research Protection Program at the University of California, San Diego, USA. The data were collected at an American College of Radiology accredited center by experienced medical experts with one of three scanners: Siemens Acuson (59%), GE L9 (21%), and ATL HDI (20%). Manually selected regions of interest (ROIs) indicating breast mass areas were outlined by a single medical expert. Malignancy of breast

masses was confirmed by biopsy, while benign breast masses were assessed either by biopsy or a clinical follow up of at least two years [11]. Several US images of malignant and benign breast masses from our dataset are presented in Fig. 1.

To better evaluate the proposed segmentation method we employed three publicly available datasets of breast US images collected at different medical centers. These data were used for testing only. For each dataset, manual ROIs generated by medical experts were provided by the authors. The first dataset, named UDIAT, consisted of 163 US images corresponding to 110 benign and 53 malignant breast masses (one mass per image) [7]. These data were acquired using Siemens ACUSON scanner from the UDIAT Diagnostic Centre of the Parc Tauli Corporation, Sabadell, Spain. The UDIAT dataset was utilized in several papers for breast mass segmentation and classification [7,10]. The second dataset, named OASBUD, consisted of 100 US images corresponding to 48 benign and 52 malignant breast masses (one mass per image) [19]. This dataset was collected from patients of the Maria Skłodowska-Curie Memorial Cancer Centre and Institute of Oncology, Warsaw, Poland. The OASBUD was originally used to investigate methods for breast mass classification [9,20,21]. The third dataset, named BUSI, consisted of 697 US images collected at the Baheya Hospital, Cairo, Egypt using LOGIQ E9 and LOGIQ E9 Agile US scanners [22]. Since all other datasets used in our study included only one breast mass per image, to make the later performance comparisons more straightforward, we removed US images from the BUSI dataset that included several breast masses. This modification resulted in 630 US images corresponding to 421 benign and 209 malignant breast masses. The BUSI dataset was originally used for breast mass classification with deep neural networks [23]. Several US images from the three datasets are presented in Fig. 2.

The same approach was applied to pre-process US images from each dataset. In the case of our dataset, we removed scanner annotations from US images. Next, all US Images were resized using bi-cubic interpolation method to dimensions of 224×224 , and processed with a 3×3 median filter. Manual ROIs were resized to 224×224 using nearest neighbor interpolation technique.

2.2. Segmentation method

General schemes of the SK-U-Net architecture and SK block are presented in Fig. 3. The SK-U-Net architecture was based on the U-Net architecture, with conventional convolution/batch normalization/activation function blocks replaced with SK blocks [18]. The aim of each SK block was to adaptively adjust network's receptive field and mix feature maps determined using different convolutions, effectively addressing the problem of large variations of breast mass sizes. Each SK block included two branches. The first one was based on convolutions with dilation size equal to 2 and 3×3 kernels filters, and the second one utilized 3×3 kernel filters with no dilation. The resulting feature maps were summed, and global average pooling was applied to convert feature maps to a single feature vector. Next, the vector was compressed using fully connected layer, with compression ratio equal to 0.5 (the number of features was halved). Compressed feature vector was decompressed with fully connected layer and processed with sigmoid activation function to determine attention coefficients for each feature map. Next, the obtained attention coefficients were used to weight the feature maps and calculate output of the SK block, using:

$$F_i = a_i F_i^d + (1 - a_i) F_i^c \quad (1)$$

where F_i is the i th feature map, and F_i^d and F_i^c stand for the feature maps calculated using dilated and conventional convolutions, respectively. a_i is the i th feature map attention parameter ranging from 0 to 1. For attention parameter of 1 the SK block fully utilized

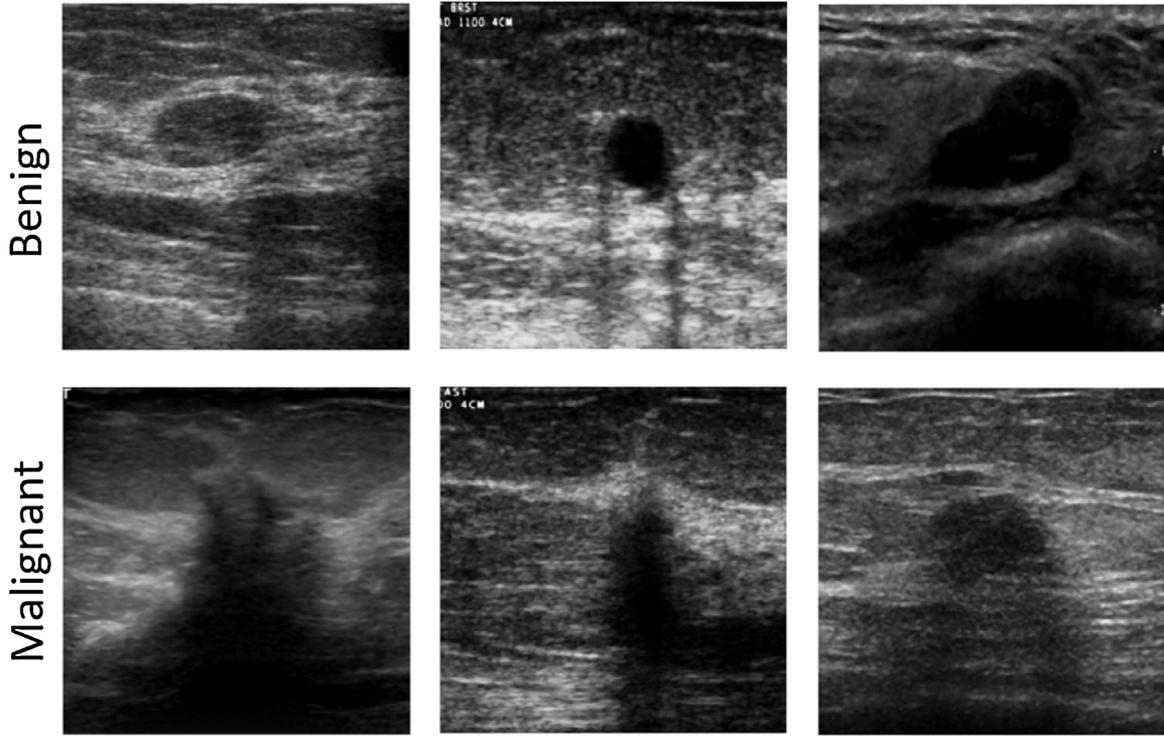


Fig. 1. Several US images presenting benign and malignant breast masses used to develop the segmentation network.

dilated convolutions, while for 0 conventional convolutions were applied.

2.3. Training and evaluation

First, we trained a SK-U-Net segmentation model. Second, for comparison we also developed a regular U-Net, in this case the SK blocks were replaced with conventional 3×3 convolutional filters, batch normalization layer and rectifier linear unit (ReLU) activation function. Both models were trained from scratch using the training set. The cost function J was defined in the following way:

$$J(A, M) = 1 - \text{Dice}(A, M) = 1 - \frac{2|A \cap M|}{|A| + |M|}, \quad (2)$$

where A and M stand automatic (predicted by the network) and manual segmentations, respectively. We used this cost function, because Dice score is commonly employed for the evaluation of segmentation models. Moreover, Dice score-based function is a good choice for segmentation of objects that strongly vary in size [24,25]. For both networks, the first block included 16 convolutional filters, and we doubled number of feature maps after each max pooling layer. In expansion path of the network, we halved the number of feature maps in blocks after each concatenation. The networks were trained using back-propagation algorithm and Adam optimization method [26]. Learning rate and the momentum were set to 0.001 and 0.9, respectively. The learning rate was exponentially decreased every 4 epochs by a factor of 0.9 if no improvement was observed on the validation set. Batch size was set to 12. The training was stopped if no improvement in respect to the Dice Score was observed on the validation set after 11 epochs. After the training, the better-performing model on the validation set was selected for further evaluation. To improve the training, we additionally applied data augmentation, US images were horizontally flipped and blurred with Gaussian noise.

After the training, we evaluated the models on the test set using standard segmentation performance metrics. First, mean Dice score

was calculated for each model based on manual and automatically generated segmentation masks. Following the approach presented in the previous paper on breast mass segmentation, we also calculated mean Dice score for the results with Dice score >0.5 [10]. Second, for each model we calculated accuracy and area under the receiver operating characteristic curve (AUC) to assess how good were the networks at detecting mass pixels. Third, we determined the detection rate of each model as the ratio of correctly detected breast masses. Breast mass was considered correctly detected if the centroid of the automatic ROI was within the manual ROI.

We applied the same evaluation procedure to assess the segmentation performance of the SK-U-Net on the UDIAT, OASBUD and BUSI datasets. First, each dataset was separately used for testing. Second, we also investigated whether additional fine-tuning of the SK-U-Net can improve segmentation performance on these datasets. We applied the same fine-tuning and evaluation procedure for the UDIAT, OASBUD and BUSI datasets. Each dataset was randomly divided into training and testing sets with a 50%/50% split, with the ratio of benign and malignant breast masses maintained. Using the training set, we fine-tuned the SK-U-Net for 10 epochs. The same training approach was applied as before, but in this case the learning rate and batch size were set to 0.0005 and 8, respectively. After the training, the fine-tuned model was evaluated on the test set. Next, we swapped the training and test sets, and repeated the procedure.

The calculations were performed in Matlab (Mathworks, USA) and Python. The networks were implemented in Keras with Tensorflow backend [27]. SK-U-Net implementation and pre-trained weights can be found at github.com/mbyr/bus_seg.

2.4. Attention mechanism

Two experiments were conducted to better understand the performance of SKs in breast mass segmentation. First, we assessed how dilated and conventional convolutions were utilized at different parts of the SK-U-Net. We calculated the mean sample attention

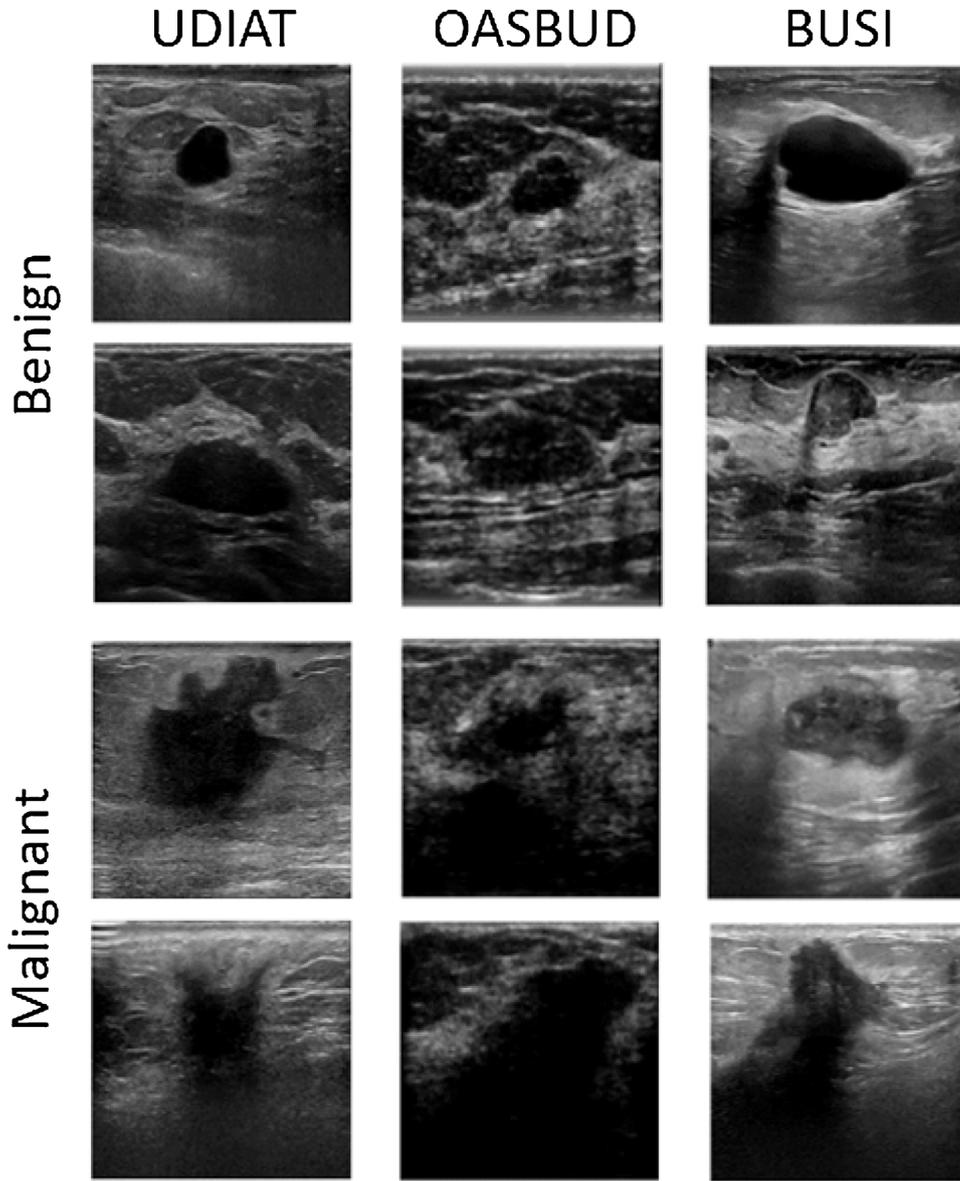


Fig. 2. US images from the UDIAT, OASBUD and BUSI datasets. These US images collected at different medical centers were used to test the proposed segmentation method.

corresponding to each breast mass and each SK block. The mean sample attention was defined in the following way:

$$\alpha_k(X) = \frac{1}{n} \sum_{i=1}^n a_i^k(X), \quad (3)$$

where X indicates sample US image, k stands for the k th SK block and n is the number of feature maps in the k th SK block. Block numbering is depicted in Fig. 3. Next, for each block we calculated the mean attention over all test sets using:

$$\bar{\alpha}_k = \frac{1}{m} \sum_{j=1}^m \alpha_k(X_j), \quad (4)$$

where m is the number of US images in the combined test set, equal to 1043. Second, we investigated whether the dilated convolutions were more utilized for the processing of US images presenting large breast masses. For each SK block we determined the Spearman's rank correlation coefficient between the mean sample attention and breast mass size expressed as the total number of ROI pixels (after image resizing).

3. Results

Table 1 presents segmentation performance scores obtained for the regular U-Net and the proposed SK-U-Net on our test set of 150 US images. Overall, our method achieved better performance than the U-Net. Mean Dice, accuracy, AUC and detection rate were equal to 0.826, 0.979, 0.958 and 0.900 for the SK-U-Net, and 0.778, 0.976, 0.909 and 0.817 for the U-Net. In the case of the benign and malignant breast mass segmentation, mean Dice score achieved by the SK-U-Net for benign masses, 0.820, was worse than for malignant masses, 0.842, due to lower detection rate of benign breast masses. Nevertheless, median Dice score presented that the network generally performed better at benign breast mass segmentation. The SK-U-Net achieved median Dice score of 0.914 for benign masses and 0.898 for malignant masses. Several segmentation results for benign and malignant cases with Dice scores around median are presented in Fig. 4.

The SK-U-Net trained on our dataset achieved good segmentation performance on US images from different medical centers. However, the performance was lower than on our dataset. The sum-

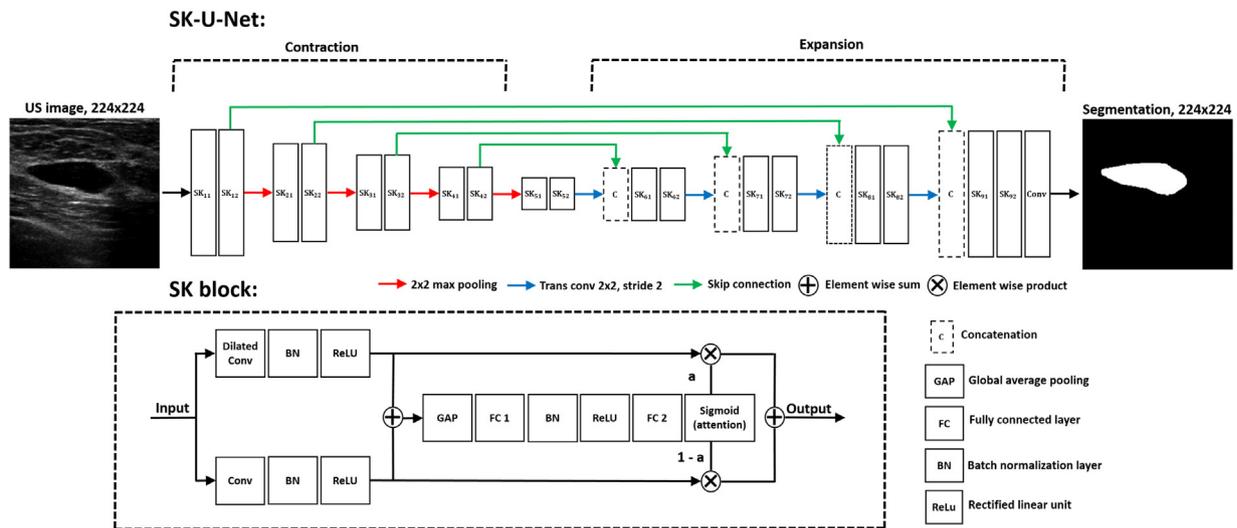


Fig. 3. The architecture of the proposed SK-U-Net, a modification of the U-Net. The SK block consisted of two branches. The first one utilized 3x3 dilated convolutions, while the second one used conventional 3x3 convolutions. Both feature maps were summed and used to calculate attention coefficients determining the usefulness of feature maps corresponding to different receptive fields.

Table 1
Breast mass segmentation performance scores (plus median and standard deviation) achieved by the U-Net and SK-U-Net calculated using test set of 150 breast masses, 39 malignant and 111 benign.

Method	Mass type	Dice	Dice > 0.5	Accuracy	AUC	Detection rate
U-Net	Benign	0.768 (0.896 ± 0.291)	0.881 (0.909 ± 0.092)	0.979 (0.991 ± 0.031)	0.904 (0.975 ± 0.167)	0.793
	Malignant	0.813 (0.856 ± 0.142)	0.836 (0.871 ± 0.106)	0.968 (0.982 ± 0.035)	0.924 (0.947 ± 0.076)	0.885
	All	0.778 (0.891 ± 0.261)	0.868 (0.900 ± 0.098)	0.976 (0.989 ± 0.032)	0.909 (0.968 ± 0.149)	0.817
SK-U-Net	Benign	0.820 (0.914 ± 0.227)	0.886 (0.916 ± 0.091)	0.981 (0.991 ± 0.028)	0.956 (0.994 ± 0.126)	0.892
	Malignant	0.842 (0.898 ± 0.165)	0.883 (0.902 ± 0.081)	0.973 (0.984 ± 0.033)	0.965 (0.988 ± 0.059)	0.923
	All	0.826 (0.907 ± 0.212)	0.885 (0.914 ± 0.088)	0.979 (0.990 ± 0.030)	0.958 (0.992 ± 0.113)	0.900

Table 2
Breast mass segmentation performance scores (plus median and standard deviation) achieved by the SK-U-Net on test images collected at different centers. First, the SK-U-Net pre-trained on our dataset was used to calculate the scores. Second, for each test set the SK-U-NET was additionally fine-tuned.

Dataset	Fine-tuning	Mass type	Dice	Dice > 0.5	Accuracy	AUC	Detection rate
UDIAT	No	Benign	0.800 (0.894 ± 0.242)	0.873 (0.908 ± 0.095)	0.989 (0.995 ± 0.016)	0.948 (0.997 ± 0.154)	0.908
		Malignant	0.738 (0.851 ± 0.251)	0.833 (0.876 ± 0.112)	0.973 (0.980 ± 0.043)	0.910 (0.957 ± 0.129)	0.889
		All	0.780 (0.877 ± 0.246)	0.860 (0.894 ± 0.102)	0.984 (0.993 ± 0.029)	0.935 (0.994 ± 0.147)	0.902
	Yes	Benign	0.819 (0.906 ± 0.230)	0.877 (0.916 ± 0.093)	0.989 (0.995 ± 0.017)	0.941 (0.996 ± 0.160)	0.917
		Malignant	0.739 (0.855 ± 0.265)	0.824 (0.867 ± 0.121)	0.975 (0.982 ± 0.039)	0.904 (0.953 ± 0.131)	0.889
		All	0.791 (0.888 ± 0.245)	0.860 (0.898 ± 0.105)	0.985 (0.993 ± 0.027)	0.929 (0.993 ± 0.152)	0.908
OASBUD	No	Benign	0.710 (0.827 ± 0.263)	0.819 (0.852 ± 0.099)	0.973 (0.982 ± 0.037)	0.938 (0.996 ± 0.153)	0.813
		Malignant	0.645 (0.762 ± 0.273)	0.784 (0.791 ± 0.084)	0.959 (0.967 ± 0.031)	0.897 (0.984 ± 0.178)	0.750
		All	0.676 (0.783 ± 0.269)	0.802 (0.824 ± 0.093)	0.966 (0.977 ± 0.035)	0.916 (0.993 ± 0.167)	0.780
	Yes	Benign	0.790 (0.881 ± 0.221)	0.845 (0.889 ± 0.113)	0.980 (0.989 ± 0.036)	0.938 (0.991 ± 0.143)	0.917
		Malignant	0.667 (0.801 ± 0.291)	0.806 (0.837 ± 0.107)	0.967 (0.972 ± 0.030)	0.874 (0.951 ± 0.159)	0.808
		All	0.726 (0.837 ± 0.266)	0.826 (0.863 ± 0.112)	0.973 (0.984 ± 0.033)	0.905 (0.977 ± 0.154)	0.860
BUSI	No	Benign	0.650 (0.796 ± 0.330)	0.843 (0.880 ± 0.113)	0.956 (0.984 ± 0.060)	0.888 (0.974 ± 0.171)	0.741
		Malignant	0.637 (0.713 ± 0.257)	0.756 (0.768 ± 0.117)	0.919 (0.935 ± 0.064)	0.820 (0.847 ± 0.137)	0.842
		All	0.646 (0.761 ± 0.308)	0.812 (0.845 ± 0.121)	0.944 (0.970 ± 0.064)	0.865 (0.937 ± 0.164)	0.775
	Yes	Benign	0.720 (0.881 ± 0.327)	0.869 (0.910 ± 0.103)	0.969 (0.990 ± 0.054)	0.905 (0.994 ± 0.177)	0.798
		Malignant	0.689 (0.814 ± 0.288)	0.810 (0.842 ± 0.115)	0.930 (0.954 ± 0.069)	0.856 (0.925 ± 0.165)	0.828
		All	0.709 (0.857 ± 0.315)	0.849 (0.889 ± 0.111)	0.956 (0.982 ± 0.062)	0.889 (0.977 ± 0.175)	0.808

many of the obtained results is presented in Table 2. For instance, our method achieved mean Dice scores of 0.780, 0.646 and 0.676 on the UDIAT, BUSI and OASBUD datasets, respectively. The corresponding detection rates were equal to 0.902, 0.775 and 0.780, respectively. Therefore, the network could efficiently detect breast masses on US images from another medical center. Additional fine-tuning improved the performance of the SK-U-Net on all test sets, especially in the case of the BUSI and OASBUD datasets. Due to

the fine-tuning the mean Dice scores increased to 0.791, 0.709 and 0.726 for the UDIAT, BUSI and OASBUD datasets, respectively. Moreover, for all three test sets we obtained better performance in the case of the benign breast mass segmentation. For instance, mean Dice scores obtained on the UDIAT dataset for the fine-tuned model were equal to 0.819 and 0.739 for benign and malignant breast masses, respectively. Representative automatic segmentations for all test sets from another centers are presented in Fig. 5.

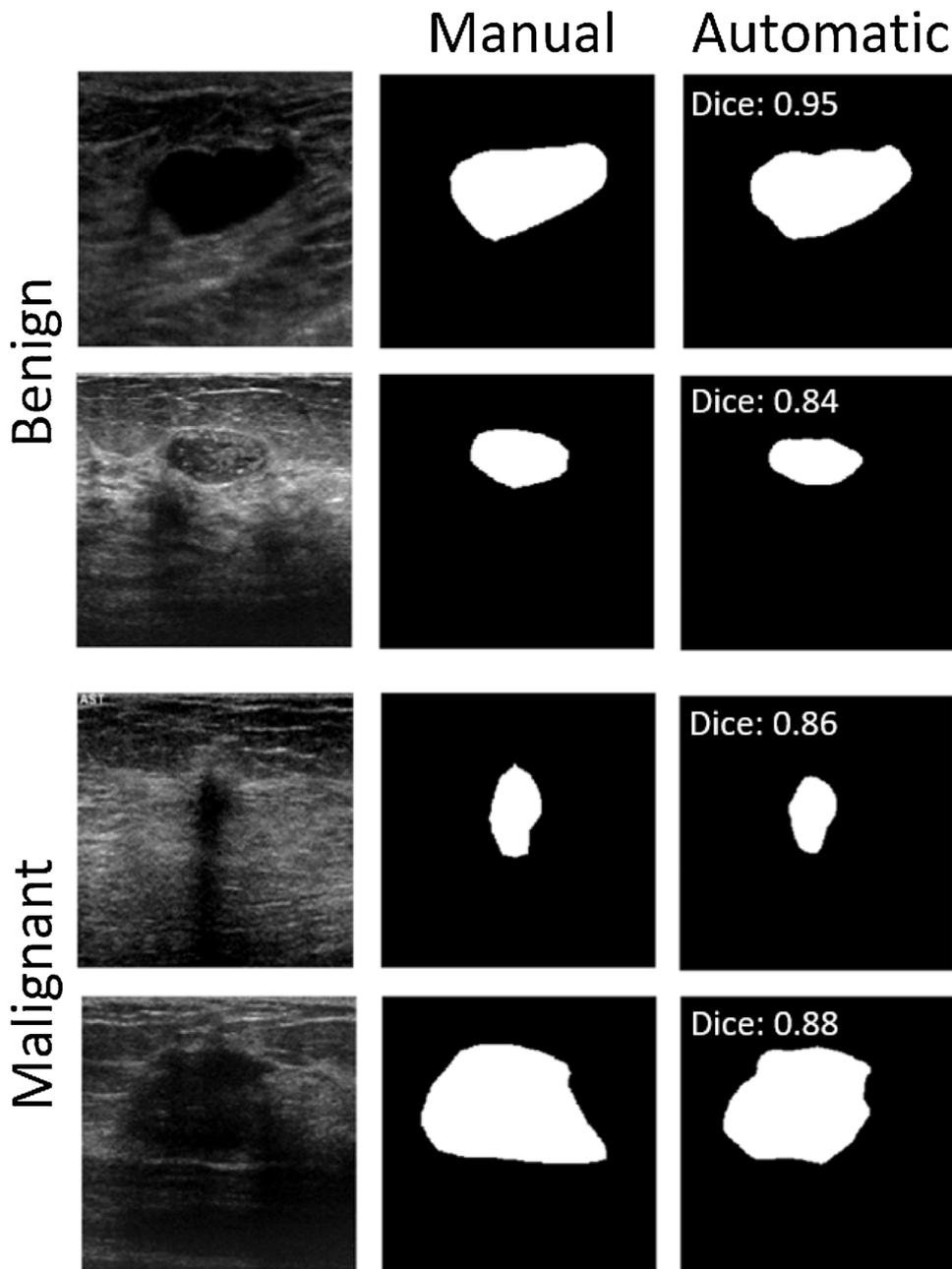


Fig. 4. Representative segmentation results (Dice score around test set median) obtained with the SK-U-Net for the test set of US images collected at our center.

Moreover, Fig. 6 shows malignant cases from each dataset for which our network achieved segmentation performance below median Dice scores. These breast mass images include examples of indistinct margins and posterior acoustic shadowing.

Utilization of the SKs for each network block is depicted in Fig. 7a). Mean attention coefficients calculated using all test US images were equal to around 0.5 for the SK blocks corresponding to the contraction and expansion paths, illustrating that the SK-U-Net used both dilated and conventional convolutions to process US images. In contrary, the SKs in the middle of the network utilized more dilated convolutions, the mean attention in this case was equal to around 70%. Fig. 7b) shows the Spearman's correlation coefficients between the mean sample attention coefficient and breast mass size obtained for each SK block. We found high correlation of around 0.7 in the case of the first SK blocks of the expansion path. The network utilized dilated convolutions to reconstruct ROIs corresponding to larger breast masses.

4. Discussion

Our study shows the usefulness of deep learning methods for breast mass segmentation in US. The proposed deep learning method based on U-Net equipped with SKs achieved good segmentation scores, and outperformed regular U-Net, illustrating the usefulness of automatic receptive field adjustment for efficient semantic object segmentation. Moreover, the proposed method was robust. We evaluated our model on three datasets of US images collected at different medical centers, and obtained good segmentation performance. However, the results obtained for other datasets were slightly lower than for our data. The mean Dice scores were equal to 0.780, 0.676 and 0.646 for the UDIAT, OAS-BUD and BUSI datasets, respectively. The better result obtained for UDIAT dataset might be due to the fact, that the UDIAT US images were collected with the ACUSON US scanner, the same US scanner used to collect our data. Appearance of US images is

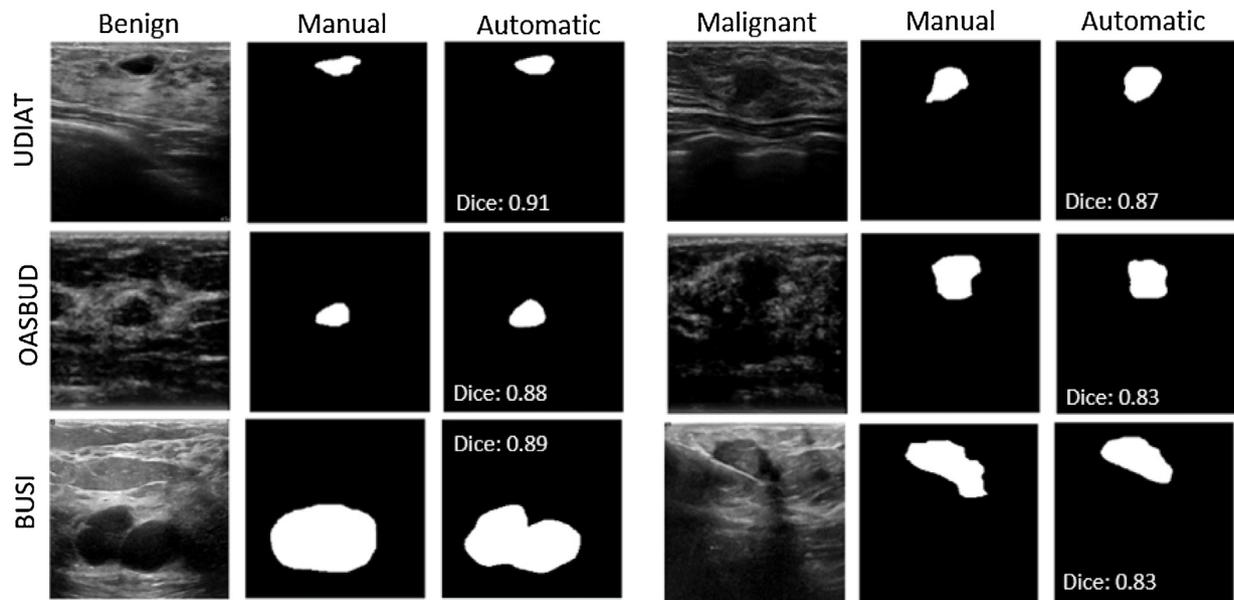


Fig. 5. Representative segmentation results (Dice score around test set median score) obtained with the SK-U-Net for the test sets collected at different medical centers. In this case, the SK-U-Net was additionally fine-tuned using US images collected at particular center.

related to the US image reconstruction method implemented by scanner manufacturer. The obtained results suggest that the US image reconstruction method might have impact on breast mass segmentation.

We presented that the network developed using data collected at one center can be fine-tuned with the data from another center to improve segmentation performance. After the fine-tuning of our SK-U-Net, the mean Dice scores increased by around 6%. The improvement was especially higher in the case of the segmentation of US images from OASBUD and BUSI datasets.

The segmentation methods proposed in this work achieved better or comparable segmentation performance in comparison to other studies [10,13]. However, it is difficult to directly compare our results with those reported by other authors due to different methods and evaluation schemes. Yap et al. used two datasets, including the UDIAT dataset, to develop deep learning segmentation methods based on fully convolutional networks [10]. They applied 5-fold cross-validation to evaluate the methods. Their better performing model achieved mean Dice scores of 0.763 and 0.548 for segmentation of benign and malignant breast masses, respectively. In comparison, our SK-U-Net (without fine-tuning), evaluated using entire UDIAT dataset, achieved mean Dice scores of 0.800 and 0.738 for segmentation of benign and malignant breast masses, respectively. While they did not report the segmentation scores for the UDIAT dataset separately, our results suggest that the SK-U-Net outperformed their segmentation method [10]. Moreover, Hu et al. developed deep learning segmentation methods based on a set of 570 US images collected from 89 patients [13]. Their better performing method based on fully convolutional network achieved high mean Dice score of 0.890. They presented that introduction of dilated convolutions at deeper layers of segmentation network may improve performance. In our study, we obtained similar results, but in an automatic fashion. The expansion path of our SK-U-Net utilized dilated convolutions (larger receptive fields) to generate ROIs for larger breast masses. Our results also showed that the SK-U-Net used both conventional and dilated convolutions to process US images in the contraction path. Moreover, the dilated convolutions were more utilized in the middle part of the segmentation network, which suggests that utilization of larger receptive fields might be especially important before the first feature map

up-sampling. The better-performing segmentation network proposed by Han et al. achieved mean Dice score of 0.871 on a test set of 800 US images. Moreover, they evaluated their method on the UDIAT dataset, and achieved mean Dice score of 0.798, comparable to our result. To the best of our knowledge, the OASBUD and BUSI datasets have not been used to evaluate segmentation methods [19,23]. Both datasets were originally used to develop methods for differentiation of malignant and benign breast masses. Therefore in this paper for the first time we reported the segmentation performance scores for these datasets.

The approach proposed in this paper has several advantages. First, by replacing conventional convolution/batch normalization/activation function blocks with SK blocks we improved the performance of the segmentation network. In a regular U-Net, the receptive field is fixed, but in the proposed SK-U-Net the receptive field can be automatically adjusted to better address the segmentation problem. Our results presented that the SKs can be useful for the segmentation of objects that, like breast masses, vary in size. While we used the SKs to improve the performance of U-Net, the approach is general. In future, it would be interesting to investigate the usefulness of the SKs for other segmentation networks, for instance fully convolutional networks.

There are several issues related to our study. First, we did not examine the usefulness of different deep learning based segmentation networks, such as fully convolutional networks or residual networks. Usefulness of SKs was investigated only in the case of the standard U-Net architecture. However, we could also assess other versions of the U-Net model, such as residual U-Nets, dense U-Nets or attention gated U-Nets [28–30]. We would like to explore the usefulness of other deep learning architectures for breast mass segmentation in future. Second, segmentation performance of machine learning methods is related to the quality of manual segmentations [31]. In our study, ROIs for each of four datasets were prepared by a different medical expert. Ideally, ROI annotator interobserver agreement should be considered to better evaluate automatic segmentation methods. In practice, a segmentation method that achieved a similar level of agreement as between several medical experts can be considered satisfactory. Even so, the accuracy of manual segmentation is not fully knowable, especially for malignant masses, where US artifacts may obscure portions of

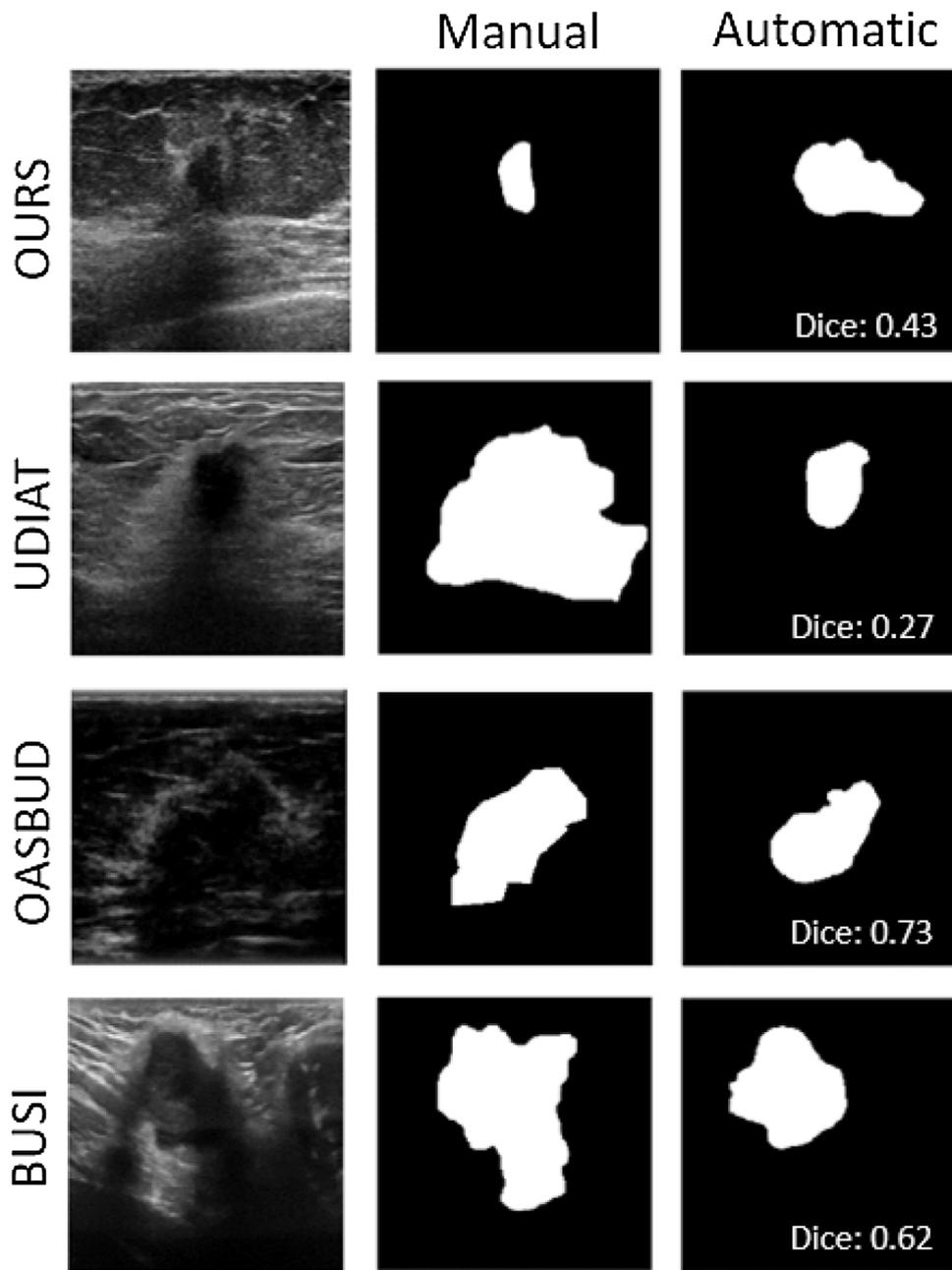


Fig. 6. Test segmentation results obtained for malignant breast mass images presenting indistinct margins and posterior acoustic shadows. In these cases, our SK-U-Net achieved segmentation performance below median Dice scores calculated for each dataset.

the mass boundary as seen in Fig. 6. Segmentation of US images presenting such masses, with indistinct margins and posterior acoustic shadows, is considered problematic and challenging. Third, we did not utilize post-processing methods to further improve the automatic segmentations. Various methods, such as conditional random fields, could be applied to improve automatic segmentations generated by deep neural networks [32]. Forth, in this work we developed the U-Net and SK-U-Net models from scratch based on a relatively small dataset of 882 US images. Hypothetically, segmentation performance of the proposed method could be further improved by incorporating more data for training if required for the intended clinical applications. Segmentation performance of the models could be additionally improved by using transfer learning, but this requires further studies.

5. Conclusions

In this work, we presented a deep learning method for efficient segmentation of breast masses in ultrasound. The proposed method was based on U-Net model equipped with selective kernels. Due to the selective kernels, the network could automatically adjust receptive fields to provide better segmentation performance in comparison to standard U-Net model. We presented the usefulness of the proposed network on three datasets of ultrasound images collected at different medical centers. We believe that the results and techniques presented in our study may serve as an important step to the development of deep learning methods for breast mass recognition. In future, we plan to investigate the usefulness of other deep learning methods for breast mass segmentation. We also plan

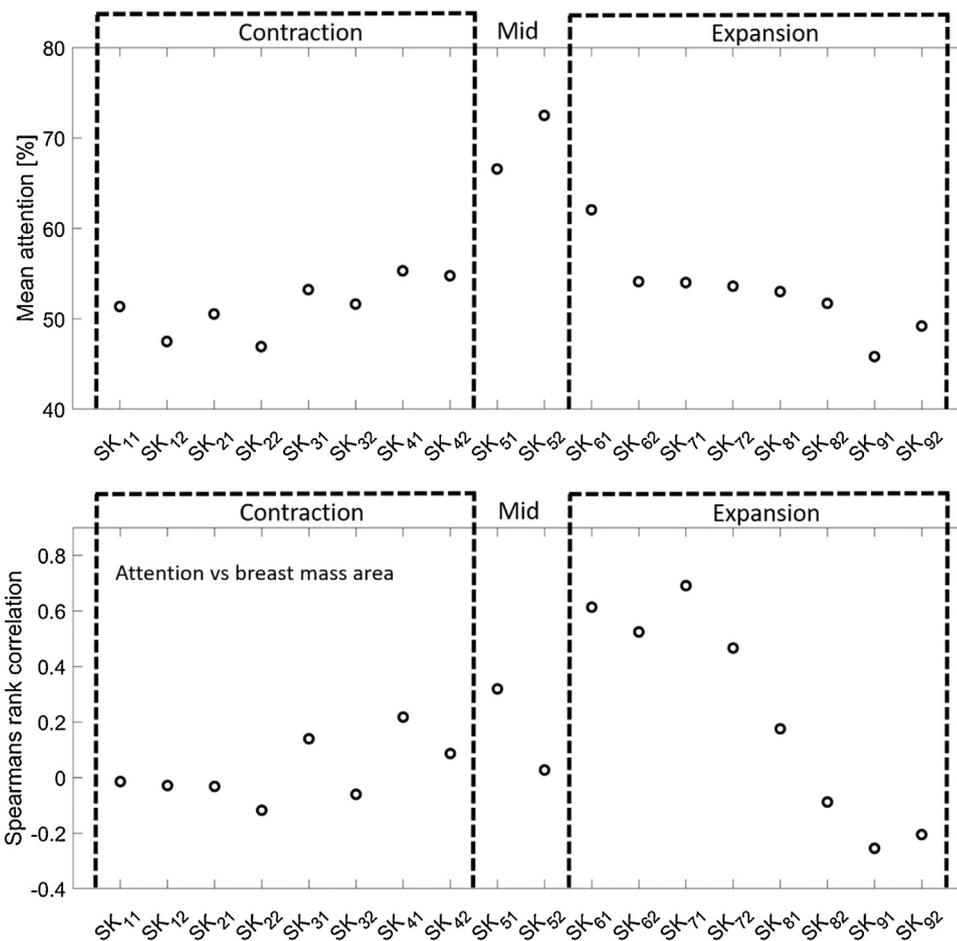


Fig. 7. (a) Mean attention calculated using all test sets for each SK block of the SK-U-Net. The middle SK blocks utilized more dilated convolutions (mean attention > 50%). (b) Spearman's rank correlation coefficients between mean attention and breast mass size calculated for each SK block. The network utilized dilated convolutions in the expansion path to reconstruct ROIs corresponding to larger breast masses.

to take into account the agreement between annotators to better evaluate automatic segmentation methods.

Authors' contribution

Michal Byra: conceptualization, methodology, software, formal analysis, investigation, resources, data curation, writing – original draft, writing – review & editing, visualization. Piotr Jarosik: conceptualization, writing – review & editing. Aleksandra Szubert: data curation. Michael Galperin: investigation, resources, data curation, funding acquisition. Haydee Ojeda-Fournier: investigation, resources, data curation. Mary O'Boyle: investigation, resources, data curation. Christopher Comstock: investigation, resources, data curation. Michael Andre: conceptualization, investigation, resources, writing – original draft, writing – review & editing, supervision, project administration funding acquisition.

Acknowledgement

This work was supported in part by Grant 2R44CA112858 from the National Institutes of Health, National Cancer Institute, USA and by the Gustavus and Louise Pfeiffer Research Foundation, NJ, USA.

Conflict of interest

The authors do not have any conflicts of interest.

References

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: Cancer J. Clin.* 68 (2018) 394–424.
- [2] G.-C. Wu, L.-Q. Zhou, J.-W. Xu, J.-Y. Wang, Q. Wei, Y.-B. Deng, X.-W. Cui, C.F. Dietrich, Artificial intelligence in breast ultrasound, *World J. Radiol.* 11 (2019) 19.
- [3] J.A. Noble, D. Boukerroui, Ultrasound image segmentation: a survey, *IEEE Trans. Med. Imaging* 25 (2006) 987–1010.
- [4] M. Xian, Y. Zhang, H.-D. Cheng, F. Xu, B. Zhang, J. Ding, Automatic breast ultrasound image segmentation: a survey, *Pattern Recogn.* 79 (2018) 340–355.
- [5] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, *Med. Image Anal.* 42 (2017) 60–88.
- [6] N. Antropova, B.Q. Huynh, M.L. Giger, A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets, *Med. Phys.* 44 (2017) 5162–5171.
- [7] M.H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwigelaar, A.K. Davison, R. Martí, Automated breast ultrasound lesions detection using convolutional neural networks, *IEEE J. Biomed. Health Informatics* 22 (2017) 1218–1226.
- [8] S. Han, H.-K. Kang, J.-Y. Jeong, M.-H. Park, W. Kim, W.-C. Bang, Y.-K. Seong, A deep learning framework for supporting the classification of breast lesions in ultrasound images, *Phys. Med. Biol.* 62 (2017) 7714.
- [9] M. Byra, Discriminant analysis of neural style representations for breast lesion classification in ultrasound, *Biocybern. Biomed. Eng.* 38 (2018) 684–690.
- [10] M.H. Yap, M. Goyal, F.M. Osman, R. Martí, E. Denton, A. Juetter, R. Zwigelaar, Breast ultrasound lesions recognition: end-to-end deep learning approaches, *J. Med. Imaging* 6 (2018) 1–8, <http://dx.doi.org/10.1117/1.JMI.6.1.011007>.
- [11] M. Byra, M. Galperin, H. Ojeda-Fournier, L. Olson, M. O'Boyle, C. Comstock, M. Andre, Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion, *Med. Phys.* 46 (2019) 746–755.

- [12] X. Qi, L. Zhang, Y. Chen, Y. Pi, Y. Chen, Q. Lv, Z. Yi, Automated diagnosis of breast ultrasonography images using deep neural networks, *Med. Image Anal.* 52 (2019) 185–198.
- [13] Y. Hu, Y. Guo, Y. Wang, J. Yu, J. Li, S. Zhou, C. Chang, Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model, *Med. Phys.* 46 (2019) 215–228, <http://dx.doi.org/10.1002/mp.13268>.
- [14] L. Han, Y. Huang, H. Dou, S. Wang, S. Ahamad, H. Luo, Q. Liu, J. Fan, J. Zhang, Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network, *Comput. Methods Programs Biomed.* (2019) 105275.
- [15] O. Ronneberger, P. Fischer, T. Brox, U-net. Convolutional networks for biomedical image segmentation, *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2015) 234–241.
- [16] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald, et al., U-net: deep learning for cell counting, detection, and morphometry, *Nat. Methods* 16 (2019) 67–70.
- [17] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015) 3431–3440.
- [18] X. Li, W. Wang, X. Hu, J. Yang, Selective kernel networks, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2019).
- [19] H. Piotrkowska-Wróblewska, K. Dobruch-Sobczak, M. Byra, A. Nowicki, Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions, *Med. Phys.* 44 (2017) 6105–6109.
- [20] M. Byra, A. Nowicki, H. Wróblewska-Piotrkowska, K. Dobruch-Sobczak, Classification of breast lesions using segmented quantitative ultrasound maps of homodyned k distribution parameters, *Med. Phys.* 43 (2016) 5561–5569.
- [21] Y. Ouyang, P.-H. Tsui, S. Wu, W. Wu, Z. Zhou, Classification of benign and malignant breast tumors using h-scan ultrasound imaging, *Diagnostics* 9 (2019) 182.
- [22] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images, *Data Brief* 28 (2020) 104863, <http://dx.doi.org/10.1016/j.dib.2019.104863>.
- [23] W. Al-Dhabyani, M. Gomaa, H. Khaled, F. Aly, Deep learning approaches for data augmentation and classification of breast masses using ultrasound images, *Int. J. Adv. Comput. Sci. Appl.* 10 (2019).
- [24] F. Milletari, N. Navab, S.-A. Ahmadi, V-net. Fully convolutional neural networks for volumetric medical image segmentation, *2016 Fourth International Conference on 3D Vision (3DV)* (2016) 565–571.
- [25] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M.J. Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (2017) 240–248.
- [26] D.P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization, 2014 arXiv:1412.6980.
- [27] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al., Tensorflow: a system for large-scale machine learning, *OSDI*, vol. 16 (2016) 265–283.
- [28] Z. Zhang, Q. Liu, Y. Wang, Road extraction by deep residual u-net, *IEEE Geosci. Rem. Sens. Lett.* 15 (2018) 749–753.
- [29] S. Li, M. Dong, G. Du, X. Mu, Attention dense-u-net for automatic breast mass segmentation in digital mammogram, *IEEE Access* 7 (2019) 59037–59047.
- [30] M. Byra, M. Wu, X. Zhang, H. Jang, Y.-J. Ma, E.Y. Chang, S. Shah, J. Du, Knee menisci segmentation and relaxometry of 3d ultrashort echo time cones MR imaging using attention u-net with transfer learning, *Magn. Reson. Med.* 83 (2020) 1109–1122.
- [31] L. Maier-Hein, M. Eisenmann, A. Reinke, S. Onogur, M. Stankovic, P. Scholz, T. Arbel, H. Bogunovic, A.P. Bradley, A. Carass, et al., Why rankings of biomedical image analysis competitions should be interpreted with care, *Nat. Commun.* 9 (2018) 5217.
- [32] Z. Zhou, G. Zhao, R. Kijowski, F. Liu, Deep convolutional neural network for segmentation of knee joint anatomy, *Magn. Reson. Med.* 80 (2018) 2759–2770.