Computer Methods and Programs in Biomedicine 162 (2018) 221–231

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks

Mohammed A. Al-masni[a,1], Mugahed A. Al-antari[a,1], Mun-Taek Choi[b], Seung-Moo Han[a], Tae-Seong Kim[a,*]

[a] Department of Biomedical Engineering, College of Electronics and Information, Kyung Hee University, Yongin, Republic of Korea
[b] School of Mechanical Engineering, Sungkyunkwan University, Republic of Korea

ARTICLE INFO

Article history:
Received 20 March 2018
Revised 30 April 2018
Accepted 17 May 2018

Keywords:
Deep learning
Dermoscopy
Full resolution convolutional network (FrCN)
Melanoma
Skin lesion segmentation

ABSTRACT

Background and objective: Automatic segmentation of skin lesions in dermoscopy images is still a challenging task due to the large shape variations and indistinct boundaries of the lesions. Accurate segmentation of skin lesions is a key prerequisite step for any computer-aided diagnostic system to recognize skin melanoma.

Methods: In this paper, we propose a novel segmentation methodology via full resolution convolutional networks (FrCN). The proposed FrCN method directly learns the full resolution features of each individual pixel of the input data without the need for pre- or post-processing operations such as artifact removal, low contrast adjustment, or further enhancement of the segmented skin lesion boundaries. We evaluated the proposed method using two publicly available databases, the IEEE International Symposium on Biomedical Imaging (ISBI) 2017 Challenge and PH2 datasets. To evaluate the proposed method, we compared the segmentation performance with the latest deep learning segmentation approaches such as the fully convolutional network (FCN), U-Net, and SegNet.

Results: Our results showed that the proposed FrCN method segmented the skin lesions with an average Jaccard index of 77.11% and an overall segmentation accuracy of 94.03% for the ISBI 2017 test dataset and 84.79% and 95.08%, respectively, for the PH2 dataset. In comparison to FCN, U-Net, and SegNet, the proposed FrCN outperformed them by 4.94%, 15.47%, and 7.48% for the Jaccard index and 1.31%, 3.89%, and 2.27% for the segmentation accuracy, respectively. Furthermore, the proposed FrCN achieved a segmentation accuracy of 95.62% for some representative clinical benign cases, 90.78% for the melanoma cases, and 91.29% for the seborrheic keratosis cases in the ISBI 2017 test dataset, exhibiting better performance than those of FCN, U-Net, and SegNet.

Conclusions: We conclude that using the full spatial resolutions of the input image could enable to learn better specific and prominent features, leading to an improvement in the segmentation performance.

1. Introduction

Melanoma skin cancer is one of the most common cancers in the world. It originates when a certain type of skin cell, called a melanocyte, begins to grow out of control. This kind of skin cancer is considered a malignant tumor and named as melanoma [1]. According to the annual report of the American Cancer Society in the United States [2], about 87,110 cases were diagnosed as new cases of melanoma, and the estimated deaths from this disease included up to 9730 cases in 2017. It is reported that melanoma is the most deadly skin cancer, with a mortality rate of 1.62% among all other cancers [3]. It is extremely important to diagnose melanoma in its early stage, since this allows treatments and increases the survival rate [4]. Visual inspection during the medical examination of skin lesions suffers from the similarity among the skin lesions and normal tissues, which may result in inaccurate diagnoses [5]. In the past decade, dermoscopy has been used to aid dermatologists in improving the screening for melanoma skin lesions. Dermoscopy is a non-invasive imaging tool that acquires a magnified image of the skin lesion utilizing polarized light [6]. It visualizes deeper details of the skin structure by abolishing the reflection of the skin surface. Although dermoscopy improves the diagnostic accuracy compared to that obtained via visual inspection, screening through dermoscopy images by dermatologists is still complex, time consum-

* Corresponding author.
E-mail addresses: m.almasani@khu.ac.kr (M.A. Al-masni), en.mualshz@khu.ac.kr (M.A. Al-antari), mtchoi@skku.edu (M.-T. Choi), smhan@khu.ac.kr (S.-M. Han), tskim@khu.ac.kr (T.-S. Kim).
[1] Authors equally contributed to this work.

https://doi.org/10.1016/j.cmpb.2018.05.027
0169-2607/

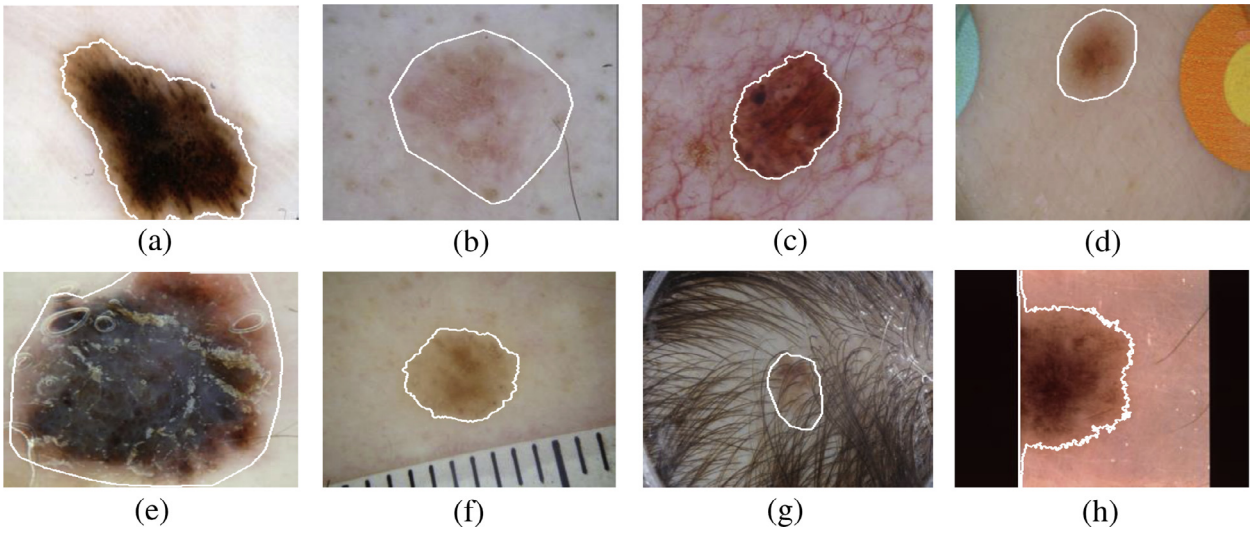

Fig. 1. Examples of some challenging cases of skin lesions such as (a) irregular fuzzy boundaries, (b) low contrast, (c) blood vessels, (d) color illumination, (e) bubbles, (f) ruler mark artifact, (g) hair artifact, and (h) frame artifact. White contours indicate the lesions segmented by expert dermatologists. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ing, subjective, and fault prone [5]. Hence, automated computerized diagnostic systems for skin lesions are highly necessary to help and support dermatologists in decision-making.

An essential preliminary step in any computerized diagnostic system of melanoma is to segment the skin lesions automatically [7,8,9]. Accurate segmentation of the skin lesions from surrounding tissues plays a crucial role in acquiring more specific and prominent features, which are utilized to classify the melanoma. In fact, it is essential to exclude normal tissues via segmentation and extract more representative features from the lesions for accurate diagnosis [9,10]. However, such segmentation is a challenging task due to the large variations among skin lesions in shapes, size, color, texture, and location in the dermoscopy images. Also, their low contrast from neighboring tissues poses additional difficulties. Furthermore, extra factors such as hair, blood vessels, ruler marks, air bubbles, ebony frames, and color illumination contribute extra obstacles to the segmentation task. Fig. 1 illustrates some examples of these challenges.

In general, segmentation techniques for skin cancer lesions can be categorized into the following five groups. *Histogram thresholding methods* determine one or more proper threshold values to differentiate the skin lesions from the surrounding tissues [11,12,13,14,15]. *Unsupervised clustering methods* utilize the color space features of RGB dermoscopy images to acquire homogenous regions [16,17,18,19]. *Edge-based* and *region-based methods* utilize the edge operator and region splitting or merging algorithms, respectively [20,21,11,22]. Active contour methods (e.g., *snakes active contours*) adapt the evolution curve algorithm to segment skin lesions [11,23,24,25]. Lastly, *supervised segmentation methods* segment the skin lesions by training the recognizers such as decision trees (DTs), support vector machines (SVMs), and artificial neural networks (ANN) [16,26]. For further details of these methods, the latest comprehensive reviews of the segmentation methods for skin lesions provide more details [7,8,27,28]. In fact, all of these techniques used the low level attributes that rely only on pixel level features. Thus, these conventional segmentation methods still do not provide satisfactory performance and cannot overcome the challenges of low contrast and hair artifacts.

Recently, segmentation techniques employing deep learning approaches have been getting major attention in the fields of object localization, semantic segmentation, and image classification [29,30,31,32,33,34,35]. Deep learning based on a convolutional neural network (CNN) is a powerful technique since it has the capability to extract more prominent features from the entire image rather than hand-crafted features [36,37,38]. In 2013, the initial CNN approach to the segmentation task was introduced by [39,40]. They applied a pixel segmentation method by dividing the input image into smaller patches (i.e., sliding windows) and passing them into a CNN classifier. The prediction of each patch represented the segmentation of the center pixel in the corresponding patch. They performed this segmentation method only for some rib regions in the chest X-ray images. In 2015, Melinščak et al. proposed a similar CNN segmentation approach for retinal vessels as a binary classification task [41]. The prediction (i.e., vessel or non-vessel) for each patch determined the segmentation of its central pixel. This approach still needs further improvement since the deep attributes extracted from each patch do not exactly reflect the center pixel features. Also, its processing requires a lot of time since the network should be executed separately for each patch [42].

Lately, state-of-the-art approaches have been developed using deep CNN methods to enhance the performance of challenging segmentation tasks. In 2015, Long et al. developed the first end-to-end pixel-wise semantic segmentation technique called a fully convolutional network (FCN) [43,44]. They adapted the well-known recognition models of AlexNet, GoogLeNet, and Visual Geometry Group (VGG)Net into FCNs by substituting the fully connected layers with the convolutional layers. They upsampled the last convolutional layer with the deconvolution and bilinear interpolation processes to produce a map that involves dense predictions with the same size as the input image. FCN achieved a significant improvement in segmentation accuracy on the Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes (PASCAL VOC) dataset against the conventional methods [45]. In 2015, Noh et al. developed a deep learning segmentation method called a deconvolution network (DeconvNet) as an extension of FCN [46]. DeconvNet consisted of a convolution network to extract the features from input images and a deconvolution network to generate object segmentation. DeconvNet utilized the VGGNet of 16 layers in the convolution network, and the same layers were mirrored in the deconvolution network. This method produced a dense map that had a higher spatial resolution compared to that generated by FCN. In 2015, Badrinarayanan et al. proposed a deep convolutional encoder-decoder segmentation method

called SegNet [47]. SegNet is similar to DeconvNet; the encoder produced low resolution feature maps, while the decoder involved a combination of convolution and upsampling layers that transformed the features into a full resolution of the input. A softmax classifier was used to produce the output map, which represented the prediction of the input pixels. In 2015, Chen et al. introduced an image segmentation method via deep convolutional nets called DeepLab [48,49]. They proposed a solution to some of the challenges faced in the semantic segmentation methods. First, they replaced the downsampling process in the last few layers by atrous convolution, which is a dilated convolution that expands the filter size and fills the hole positions with zeros to preserve the feature resolution. Second, they utilized the atrous spatial pyramid pooling for multiscale image representations. To recover an accurate boundary of the segmented objects, they appended their network with the fully connected conditional random field (CRF). In 2015, Ronneberger et al. built a new convolutional segmentation method called U-Net, which was an amendment and extension of the FCN architecture [42]. U-Net consisted of the contracting path, which is similar to the FCN structure, and the expansive path, which applied the upsampling process followed by a convolution operator. The features map in the expansive path was concatenated with the corresponding features in the contracting path. Arguably, FCN is the most successful semantic segmentation method, but other methodologies are also remarkable [45]. More details of these semantic segmentation methods are given in a review paper [45].

Recently, these deep learning approaches have been applied in the field of skin lesion segmentation. In 2017, Yu et al. developed a two-stage approach for the segmentation and classification of melanoma skin lesions utilizing deep residual networks [50]. Their fully convolutional residual network (FCRN) was ranked second in the segmentation challenge with an overall accuracy of 94.9% when utilizing the IEEE International Symposium on Biomedical Imaging (ISBI) 2016: Skin Lesion Analysis Towards Melanoma Detection Challenge Dataset [51]. Furthermore, in the recognition challenge, their deep residual network (DRN) method was ranked first, achieving a recognition accuracy of 85.5%. In 2017, Bi et al. proposed a cascaded multi-stage FCN followed by the parallel integration (PI) method to segment lesions from dermoscopy images [52]. The PI method was utilized for further refinement of the boundaries of the segmented lesions. They evaluated their method using two public datasets, the ISBI 2016 Skin Lesion Challenge and PH2 datasets [53]. They achieved overall segmentation accuracies of 95.51% and 94.24% and dice coefficient indices of 91.18% and 90.66%, respectively, in the two datasets. In 2017, Yuan et al. presented a skin lesion segmentation method via deep FCN [54]. They amended the well-known FCN method by employing the Jaccard distance as a loss function in order to address the imbalance among tissue-lesion pixels. Their method achieved overall accuracies of 95.5% and 93.8% with the ISBI 2016 and PH2 datasets, respectively. In 2017, Lin et al. presented a comparison between two skin lesion segmentation approaches, U-Net-based histogram equalization and C-means clustering [55]. This work was evaluated using the ISBI 2017 skin lesion analysis challenge dataset [56]. Their U-Net method achieved a dice coefficient index of 77%, which significantly outperformed the clustering technique's result of only 61%. In 2017, Goyal et al. developed a multi-class semantic segmentation via FCN that was able to segment three classes of skin lesions (i.e., benign nevi, seborrhoeic keratoses, and melanoma) from the ISBI 2017 dataset [57]. Their approach achieved dice indices of 78.5%, 65.3%, and 55.7% for benign, melanoma, and seborrhoeic keratosis lesions, respectively. In 2017, Yuan et al. presented a skin lesion segmentation using deep convolutional-deconvolutional neural networks (CDNN) [58]. They trained their model with different color spaces of dermoscopy images using the ISBI 2017 dataset. Their approach was ranked first in the ISBI 2017 Challenge, in which it achieved a Jaccard index of 76.5%.

In this paper, we propose a novel deep learning method for skin lesion segmentation called full resolution convolutional networks (FrCN). We amend and extend the deep FCN segmentation method. Unlike the previous deep learning segmentation methods that compensate the subsampled pixel-wise features using an uppooling process, our proposed FrCN segmentation method does not require any pre-processing techniques, such as artifact removal or low contrast enhancement, since it directly learns the full resolution features of each individual pixel of the input data. Furthermore, it directly predicts the segmented skin lesions maps with fine boundaries without the need for the CRF algorithm as a post-processing step. Our results show the improved performance of the proposed FrCN over the latest deep learning segmentation methodologies.

The main contributions of this paper are outlined as follows. We propose a new FrCN segmentation method that learns the full resolution features of each pixel of the input data to achieve more accurate pixel-wise segmentation of the skin lesions. This is achieved by eliminating the subsampling layers in the networks and enabling the convolutional layers to extract and learn the full spatial features of the input image. Due to this, the proposed FrCN produces finely segmented contours of the skin lesions. We compare the performance of our FrCN segmentation method against the well-known deep learning approaches FCN, U-Net, and SegNet under the same conditions and with the same datasets.

The reminder of this paper is organized as follows. First, the details and distributions of the utilized datasets are introduced. Second, we describe in detail the proposed FrCN segmentation method. Then, we present and analyze the results of the proposed method compared to the recent state-of-the-art approaches. Finally, the conclusions of this study are presented.

2. Materials and methods

2.1. Databases

In this study, two well-known and open dermoscopy databases were utilized to evaluate the proposed segmentation method. The first database is the ISBI 2017 challenge dataset called "Skin Lesion Analysis Towards Melanoma Detection." This database is provided by the International Skin Imaging Collaboration (ISIC) archive [56] and is available at [59]. This challenge dataset contains 8-bit RGB dermoscopy images with different image sizes from 540×722 to 4499×6748 pixels. It provides 2000 training images and separate datasets of 150 and 600 images for validation and testing, respectively. All of these dermoscopy images contain skin lesions, which are labeled as benign nevi, melanoma, or seborrheic keratosis. The second public dataset is called PH2 and was collected through a collaboration between the dermatology service of Hospital Pedro Hispano and the research group of the Universidade do Porto, Técnico Lisboa in Matosinhos, Portugal [53]. The PH2 dataset contains 200 dermoscopy images that were acquired under the same conditions utilizing a magnification of $20\times$. They are 8-bit RGB images, and the image size is fixed as 768×560 pixels. The clinical diagnoses of the PH2 dataset include 160 benign nevi (i.e., common nevi and atypical nevi) cases and 40 melanoma cases. Table 1 summarizes the distributions of both datasets. Furthermore, both of these datasets provided the original images paired with the lesion segmentation boundaries, which were annotated by expert dermatologists. Fig. 2 shows instances of these dermoscopy images from both the ISBI 2017 challenge and the PH2 datasets with their ground truth segmentation masks. To avoid overfitting and ensure generalization of deep learning architectures with the unseen data, deep learning networks need to be trained, evaluated,

Table 1
Distribution of the ISBI 2017 challenge and PH2 datasets.

Dataset	Training data				Validation data				Test data				Total
	B*	M*	SK*	Total	B	M	SK	Total	B	M	SK	Total	
ISBI 2017	1372	374	254	2000	78	30	42	150	393	117	90	600	2750
PH2	–	–	–	–	–	–	–	–	160	40	–	200	200

*B, M, and SK indicate benign nevi, melanoma, and seborrheic keratosis classes, respectively.

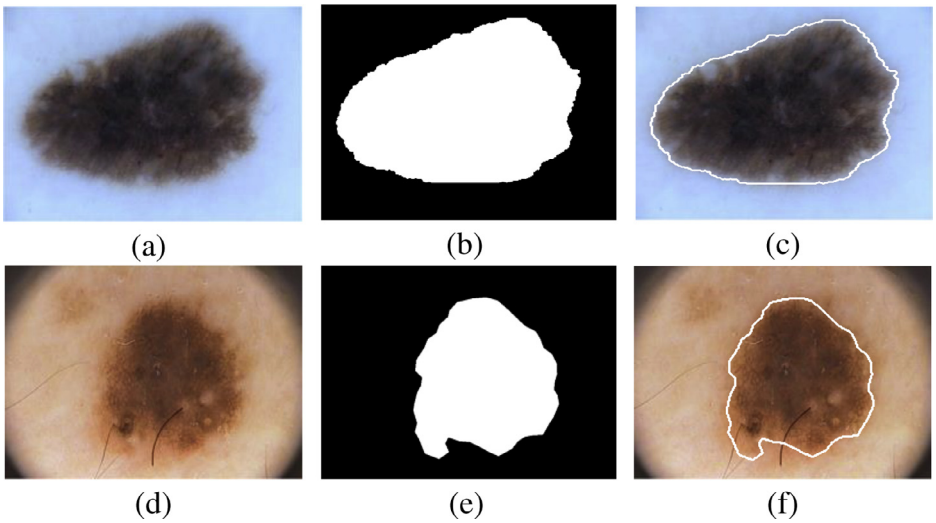

Fig. 2. Exemplary pairs of the original dermoscopy images and their segmentation masks, which were annotated by expert dermatologists for (a) and (b) from the ISBI 2017 challenge, and (d) and (e) from the PH2 datasets. (c) and (f) illustrate the segmented lesion boundaries of (a)–(b) and (d)–(e), respectively.

and tested using independent datasets. A training dataset is utilized to learn the deep learning network with different test parameters, which are evaluated using the validation dataset. The optimal network is selected according to its high performance on the validation set. Then, further assessment is performed on the test dataset to obtain the overall performance of the deep learning model [60,61].

2.2. Data preprocessing and augmentation

In deep learning approaches, proper training requires a large amount of data. Thus, we applied an image augmentation process to enlarge the training dataset. A simple pre-processing procedure is provided to increase and facilitate the learning of our proposed segmentation approach from the training dataset. By taking advantage of the fact that the dermoscopy images existed in the form of RGB, we included the three channels of the Hue-Saturation-Value (HSV) in addition to the RGB images in order to learn different color space features. Therefore, we collected 4000 dermoscopy images for training (i.e., 2000 RGB and 2000 HSV images). In addition, we augmented the training data by rotating all of the 4000 dermoscopy images four times with angles of 0°, 90°, 180°, and 270°. Thus, a total of 16,000 dermoscopy images were utilized to train our proposed FrCN segmentation method. This process reduced the overfitting of the system and improved the robustness of the deep learning. To sum up, the augmentation processes were only applied to the training data to learn and build an optimal deep network. Furthermore, due to the variations in the image sizes of the dataset, in which the majority have height to width ratios similar to 3:4, we resized all of the images into 192×256 pixels using bilinear interpolation, as in [54]. It is possible that resizing images in their aspect ratio may cause geometric distortion. In our case, we found only 4.2% images out of the aspect ratio of $3{:}4 \pm 10\%$ in the training data (e.g., 84 out of 2000 images). However, our proposed deep learning FrCN method learns the global representation of the lesions from all the images. This may be able to overcome the effect from the small amount of distortion. Based on the conclusion in [54], reporting the optimal segmentation performance with the image size of 192×256 among other different image sizes, we adopted the same image size of 192×256 in this study.

2.3. Proposed FrCN architecture

In general, CNN consists of two essential parts to achieve pixel-wise classification for segmentation. The first part is the convolutional and subsampling layers. Convolutional layers are responsible for extracting the deep features of the input image using various filter types of different sizes, while the subsampling layers are used to reduce the sizes of the extracted feature maps [62,63]. In the image-wise classification task (i.e., traditional image classification using CNN), the subsampling process does improve the robustness of the classifier, since it eliminates the redundancy of features, reduces overfitting, and minimizes computation time [43,52]. This is due to the fact that all the reduced features represent a label of the input image. However, in the pixel-wise segmentation task, subsampling causes reduction of the spatial features resolution of the input image. The second part of CNN is the upsampling layers followed by the softmax classifier, which use the extracted features of each pixel to classify it as a lesion or a tissue pixel. Therefore, recent deep learning segmentation methods have utilized complex procedures to compensate for the missing features of certain pixels due to the reduced image size, using upsampling, bilinear interpolation, deconvolution, atrous convolution, or decoding [42,44,46,47,49]. Moreover, these processes require further CRF or PI operations to refine the segmented boundaries of objects [49,52].

In this paper, we propose a new deep learning segmentation method, FrCN, in which the full resolution features of the individual input pixels are reserved by eliminating all the subsampling

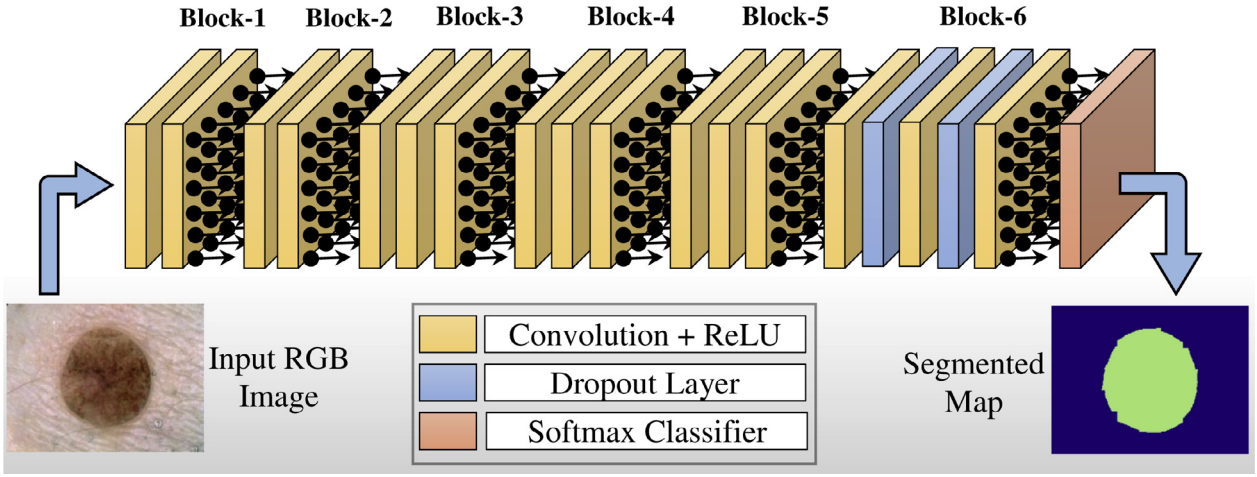

Fig. 3. An illustration of the proposed FrCN segmentation method. There are no subsampling nor fully connected layers.

Table 2
Details of architecture layers of the proposed FrCN method.

Block	Layer	Filter size, maps	Block	Layer	Filter size, maps
Input data	192 × 256 × 3	–		Conv. 11	3 × 3, 512
Block-1	Conv. 1	3 × 3, 64	Block-5	Conv. 12	3 × 3, 512
	Conv. 2	3 × 3, 64		Conv. 13	3 × 3, 512
Block-2	Conv. 3	3 × 3, 128		Conv. 14	7 × 7, 4096
	Conv. 4	3 × 3, 128		Dropout 1	$p = 0.5$
Block-3	Conv. 5	3 × 3, 256	Block-6	Conv. 15	1 × 1, 4096
	Conv. 6	3 × 3, 256		Dropout 2	$p = 0.5$
	Conv. 7	3 × 3, 256		Conv. 16	1 × 1, 2
Block-4	Conv. 8	3 × 3, 512		Softmax	–
	Conv. 9	3 × 3, 512	OutputMap	192 × 256 × 3	–
	Conv. 10	3 × 3, 512			

layers from the architecture. In this case, each pixel can be represented as a training sample. Fig. 3 illustrates the architecture of the proposed FrCN segmentation method. Obviously, the FrCN method allows each pixel in the input image to extract its own features using the convolutional layers. Our FrCN architecture consists of 16 layers, which are inspired by the VGG-16 network layers [64]. The last three fully connected neural network layers in Block-6 were replaced by the convolutional layers. Details of the convolutional filter sizes and the number of maps in each layer are provided in Table 2.

A convolutional network is originally a translation invariant that produces features maps $\boldsymbol{F}$ using convolution operations [65,66]. The basic components of this network are the convolution operator '$*$,' filter kernels $\boldsymbol{W}$, and activation function $\phi(\bullet)$. Therefore, the k^{th} feature map of layer L is computed as follows:

$$\boldsymbol{F}_L^k = \phi\left(\boldsymbol{W}_L^k * \boldsymbol{F}_{L-1}^k + b_L^k\right), \tag{1}$$

where b_L^k is the bias applied to each feature map of each layer. We applied a nonlinear activation operation via rectified linear units (ReLUs) immediately after each convolutional layer to provide nonlinearity to our network. A ReLU activation function was widely utilized rather than tanh and sigmoid functions due to its capability to train the network with higher computational efficiency and to alleviate the vanishing gradient problem [67,68]. The ReLU activation function is defined as follows:

$$\phi(x) = \max(0,\ x) = \begin{cases} x, & if\ x \geq 0 \\ 0, & if\ x < 0\ . \end{cases} \tag{2}$$

In addition, we utilized a dropout of $p = 0.5$ after the convolutional layer numbers 14 and 15, as shown in the proposed FrCN architecture in Fig. 3 and Table 2. Dropout is a method used in deep learning to address the overfitting problem of deep layers [69]. During training, it randomly eliminates some units with their connections to prevent overfitting to the training data. Thus, every unit is retrained with a certain probability, leading the outgoing weights of that unit to be multiplied by p at the test time. Finally, the last layer of the proposed FrCN architecture is fed by a multinomial logistic regression known as the softmax classifier. This produces a map of predictions for each pixel, which contains the segmented map. As a result, the resolutions of the spatial output maps are that same as that of the input data. Throughout the training stage, the overall loss H of each pixel is minimized by calculating the cross entropy loss function between the ground truth annotation y and the predicted segmented map $\hat{y}$ as follows:

$$\mathrm{H} = -\left(y \cdot \log\left(\hat{y}\right) + (1-y) \cdot \log\left(1-\hat{y}\right)\right). \tag{3}$$

The cross entropy loss function is usually used when a deep convolutional network is applied for a pixel-wise classification [44,54]. The objective of training deep learning methods is to optimize the weight parameters in each layer. The optimization process of a single cycle proceeds as follows [36,70]. First, the forward propagation path sequentially computes the output in each layer using the training dataset. In the last output layer, the error between the predicted and ground truth labels is measured utilizing the loss function. To minimize the training error, back-propagation is performed through the network layers. Consequently, the training weights of FrCN can be updated utilizing the training data.

2.4. Training and testing

Medical applications of deep networks still suffer from the scarcity of medical data [71]. This is due to the complexity and cost of the labeling process during data acquisition [72]. To overcome the challenge of insufficient training data in the field of

dermoscopy skin lesion images, we adopted two strategies, data augmentation and transfer learning [52,57,71,72,73,74,75]. For data augmentation, we used the HSV and image rotations as described above. For transfer learning, we used the pre-training model weights trained by the VGG-16 net layers [64] using the large public ImageNet dataset [76]. Subsequently, we re-trained or fine-tuned the proposed deep FrCN segmentation method using the augmented training data of the dermoscopy images. Regarding the optimization of the network of the proposed segmentation method, we used a separate validation dataset containing 150 dermoscopy images. Hence, the evaluations of the network optimization and the final network performance were performed independently using the validation and test datasets, respectively. This unbiased procedure is called a double cross validation scheme and is more reliable and more commonly used than the trial and error process in the most recent deep learning approaches [77,78]. In addition, we evaluated our proposed segmentation methods against the recent deep learning FCN, U-Net, and SegNet methods using two different test subsets that included 600 test images from ISBI 2017 and 200 images from PH2.

The computation time of the training stage was about 17.5 h over 200 epochs with a batch size of 20. However, the decoding (i.e., segmentation) for a single dermoscopy image took about 9.7 s. This should make the proposed deep learning segmentation method applicable for clinical practice. All experiments were conducted on a personal computer with the following specifications: Intel® Core(TM) i7-6850 K with 16 GB RAM, frequency of CPU @ 3.360 GHz, and GPU of NVIDIA GeForce GTX 1080. This work was implemented with Python 2.7.14 on Ubuntu 16.04 OS using the Theano and Keras deep learning libraries [79].

2.5. Evaluation metrics

To evaluate the capability of the proposed deep FrCN segmentation method quantitatively, we used the following skin lesion segmentation evaluation measures. *Sensitivity (SEN)* indicates the proportion of the skin lesion pixels that are correctly segmented, while the *specificity (SPE)* measures the ratio of the skin non-lesion pixels that are not correctly segmented [57,80,81]. The *dice coefficient (DIC)* was also used to quantify the efficiency of the proposed FrCN segmentation method. This measures how similar the segmented lesions are to the annotated ground truths [82,83]. The *Jaccard index (JAC)* is an intersection over union of the segmented lesions with the ground truth masks [84]. *Accuracy (ACC)* is also provided to show the overall pixel-wise segmentation performance. The *Matthew correlation coefficient (MCC)* measures the correlation between the annotated and segmented skin lesion pixels and returns values in a range from −1 to + 1 [84]. A higher MCC value implies a better segmentation performance. All of these metrics were computed in relation with the elements of the confusion matrix as follows.

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{4}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}, \tag{5}$$

$$\mathrm{DIC} = \frac{2\cdot\mathrm{TP}}{(2\cdot\mathrm{TP})+\mathrm{FP}+\mathrm{FN}}, \tag{6}$$

$$\mathrm{JAC} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}}, \tag{7}$$

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FN}+\mathrm{TN}+\mathrm{FP}}, \tag{8}$$

$$\mathrm{MCC} = \frac{\mathrm{TP}\cdot\mathrm{TN}-\mathrm{FP}\cdot\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}, \tag{9}$$

where TP, FP, TN, and FN denote the true and false positives and the true and false negatives, respectively. If the lesion pixels were segmented correctly, they were considered as TPs; otherwise, they were FNs. In contrast, the non-lesion pixels were considered as TNs if their segmentation was classified correctly as non-lesion; otherwise, they were FPs. In addition, the curve of the receiver operator characteristic (ROC) with its area under the curve (AUC) were utilized for segmentation assessment.

3. Results

3.1. Segmentation performance on the ISBI 2017 test dataset

This section shows the segmentation performance of the proposed FrCN against the latest deep learning approaches using 600 dermoscopy images of the ISBI 2017 test dataset. Quantitatively, Tables 3 and 4 summarize the segmentation performance results of the proposed FrCN method compared to FCN, U-Net, and SegNet. In this study, the segmentation of FCN was performed using a stride of 32. Table 3 shows the performances in terms of sensitivity, specificity, and overall pixel segmentation accuracy, while the measurements in Table 4 are presented in terms of the dice, Jaccard, and Matthew correlation indices. Our FrCN outperformed the others with overall accuracy, dice, and Jaccard indices of 94.03%, 87.08%, and 77.11%, respectively. In contrast, U-Net achieved the highest specificity for overall skin lesion segmentation at 97.24%. Moreover, we analyzed the segmentation performance on the diagnosis categories of the skin lesions. Measurements of each diagnostic class including benign, melanoma, and seborrheic keratosis (SK) cases are presented in Tables 3 and 4. The segmentation accuracies of the benign, melanoma, and SK cases in the ISBI 2017 test dataset for our proposed FrCN were 95.62%, 90.78%, and 91.29%, respectively. Finally, the ROC curves with their AUCs of our FrCN versus FCN, U-Net, and SegNet are illustrated, separately for each diagnostic class, in Fig. 4. The overall AUCs of FCN, U-Net, SegNet, and FrCN were 88.32%, 82.19%, 87.71%, and 91.04%, respectively. For qualitative evaluation, Fig. 5(a), (b), and (c) present some exemplar segmentation results of the proposed FrCN compared to the ground truth contours for benign, melanoma, and SK cases from the ISBI 2017 test dataset.

3.2. Segmentation performance on the PH2 dataset

We also evaluated the segmentation performance with the PH2 test dataset, which consists of 200 dermoscopy images. Our segmentation method outperformed the other techniques, as demonstrated in Tables 5 and 6. The results show the robustness of our FrCN. The overall segmentation accuracies of FCN, U-Net, SegNet, and FrCN were 92.82%, 92.55%, 93.36%, and 95.08%, respectively. Fig. 6 illustrates the ROC curves of our proposed FrCN compared to the conventional segmentation methods. It is clear that our method performed slightly better on the benign cases, while it demonstrated a significant improvement in the melanoma cases. The AUCs of the benign cases for FCN, U-Net, SegNet, and FrCN were 94.72%, 92.16%, 94.07%, and 94.97%, while those for the melanoma cases were 86.44%, 84.53%, 86.16%, and 94.06%, respectively. Fig. 5(d), (e), and (f) show the qualitative performances of some segmentation results of the proposed FrCN for common benign, atypical benign, and melanoma cases from the PH2 dataset, respectively.

Table 3
Segmentation performance (%) of the proposed FrCN compared to FCN, U-Net, and SegNet for the ISBI 2017 test dataset.

Method	Benign cases			Melanoma cases			SK cases			Overall		
	SEN	*SPE*	*ACC*	*SEN*	*SPE*	*ACC*	*SEN*	*SPE*	*ACC*	*SEN*	*SPE*	*ACC*
FCN	85.25	96.95	94.45	70.67	96.18	88.25	75.10	95.91	90.96	79.98	96.66	92.72
U-Net	76.76	97.26	92.89	58.71	96.81	84.98	43.81	97.64	84.83	67.15	97.24	90.14
SegNet	85.19	96.30	93.93	73.78	94.26	87.90	70.58	92.50	87.29	80.05	95.37	91.76
FrCN	88.95	97.44	95.62	78.91	96.04	90.78	82.37	94.08	91.29	85.40	96.69	94.03

Table 4
Segmentation performance (%) of the proposed FrCN compared to FCN, U-Net, and SegNet for the ISBI 2017 test dataset.

Method	Benign cases			Melanoma cases			SK cases			Overall		
	DIC	*JAC*	*MCC*	*DIC*	*JAC*	*MCC*	*DIC*	*JAC*	*MCC*	*DIC*	*JAC*	*MCC*
FCN	86.77	76.63	83.28	78.89	65.14	71.84	79.81	66.40	74.26	83.83	72.17	79.30
U-Net	82.16	69.72	78.05	70.82	54.83	63.71	57.88	40.73	53.89	76.27	61.64	71.23
SegNet	85.69	74.97	81.84	79.11	65.45	71.03	72.54	56.91	64.32	82.09	69.63	76.79
FrCN	89.68	81.28	86.90	84.02	72.44	77.90	81.83	69.25	76.11	87.08	77.11	83.22

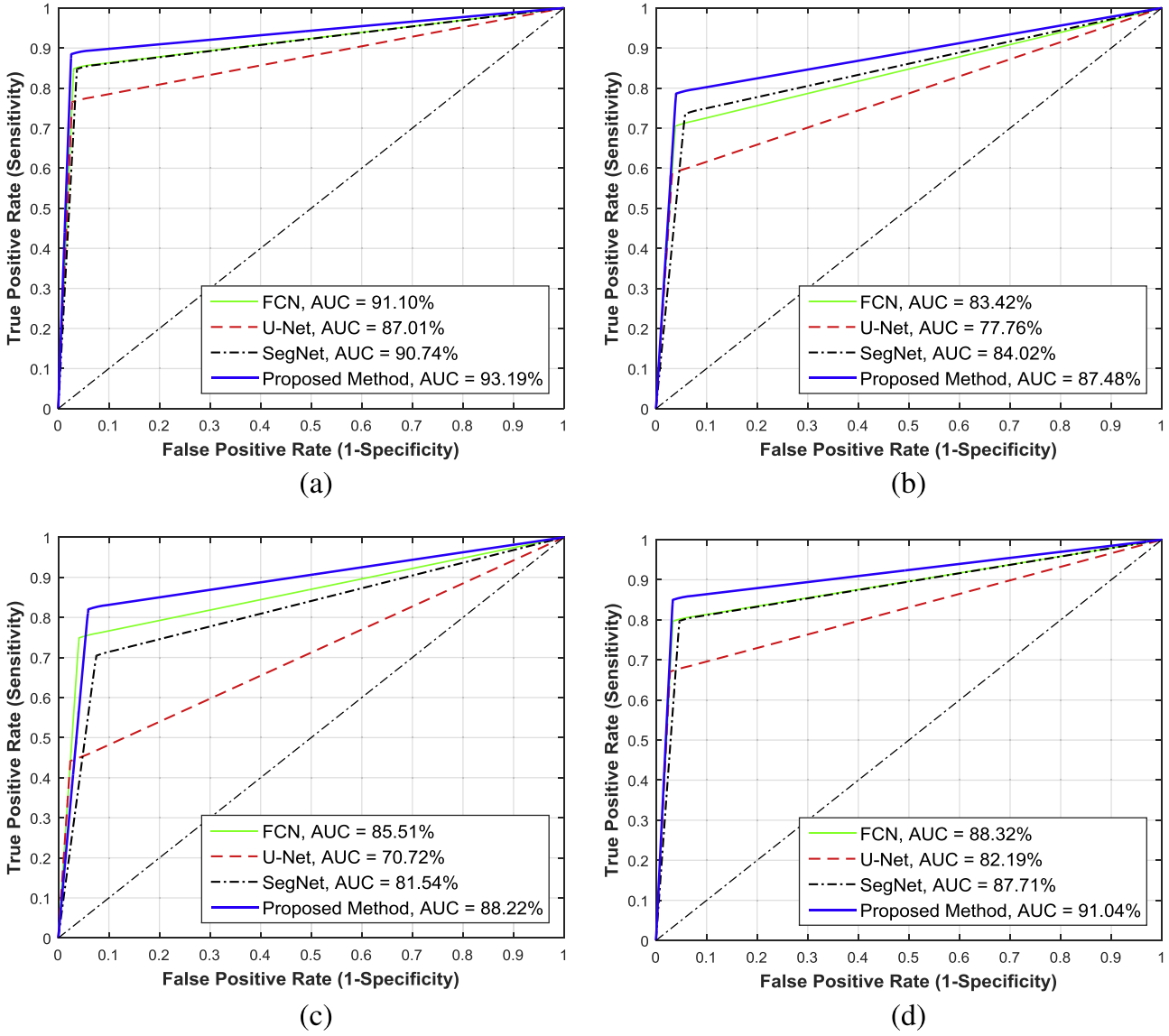

Fig. 4. ROC curves of different deep learning segmentation methods on the ISBI 2017 skin lesion dataset for (a) benign, (b) melanoma, (c) SK, and (d) overall clinical cases.

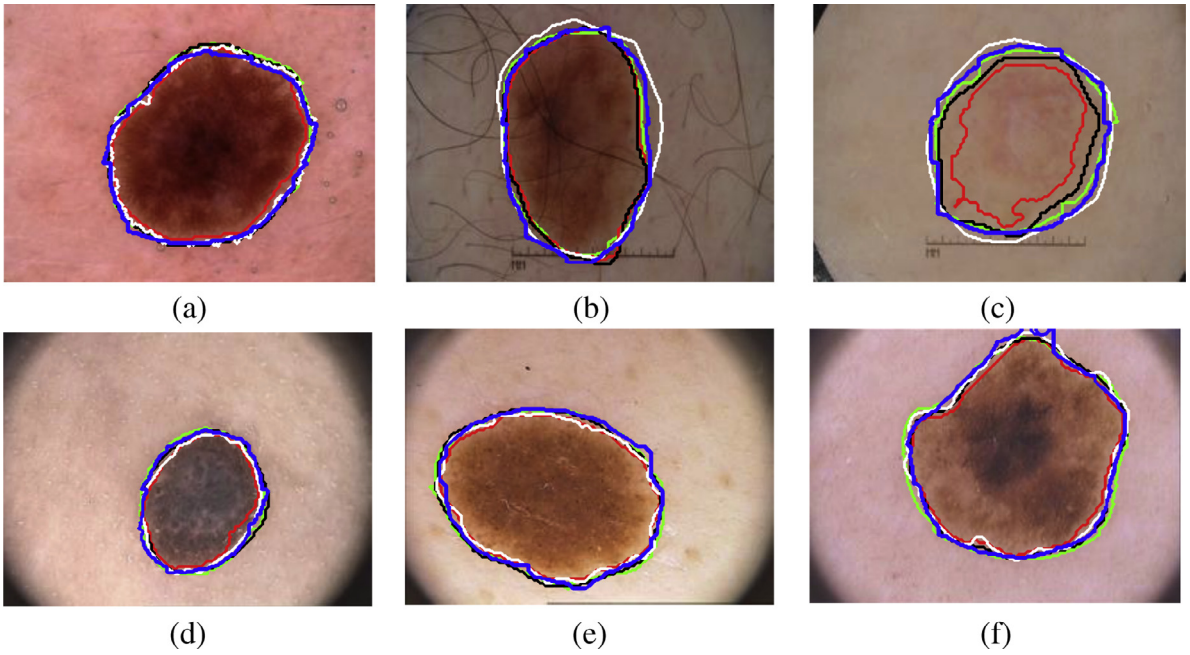

Fig. 5. The segmentation results of the proposed FrCN against FCN, U-Net, and SegNet for (a) benign, (b) melanoma, and (c) seborrheic keratosis cases from the ISBI 2017 dataset, while (d) common nevi, (e) atypical nevi, and (f) melanoma indicate the cases from the PH2 dataset. The segmentation results of the ground truth (white) and FrCN (blue), while FCN (green), U-Net (red), and SegNet (black).(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Segmentation performance (%) of the proposed FrCN compared to FCN, U-Net, and SegNet for the PH2 dataset.

Method	Benign cases			Melanoma cases			Overall		
	SEN	*SPE*	*ACC*	*SEN*	*SPE*	*ACC*	*SEN*	*SPE*	*ACC*
FCN	95.35	94.09	94.44	90.30	94.02	92.82	90.30	94.02	92.82
U-Net	86.68	97.63	94.60	70.58	98.47	84.36	81.63	97.76	92.55
SegNet	91.57	96.57	95.19	75.50	96.83	86.04	86.53	96.61	93.36
FrCN	94.48	95.46	95.20	91.57	96.55	94.64	93.72	95.65	95.08

Table 6
Segmentation performance (%) of the proposed FrCN compared to FCN, U-Net, and SegNet for the PH2 dataset.

Method	Benign cases			Melanoma cases			Overall		
	DIC	*JAC*	*MCC*	*DIC*	*JAC*	*MCC*	*DIC*	*JAC*	*MCC*
FCN	90.46	82.59	86.78	89.03	80.22	83.71	89.03	80.22	83.71
U-Net	89.88	81.63	86.32	82.04	69.55	71.73	87.61	77.95	82.78
SegNet	91.32	84.03	87.99	84.55	73.23	73.89	89.36	80.77	84.64
FrCN	91.38	84.13	88.15	92.92	86.77	88.62	91.77	84.79	88.30

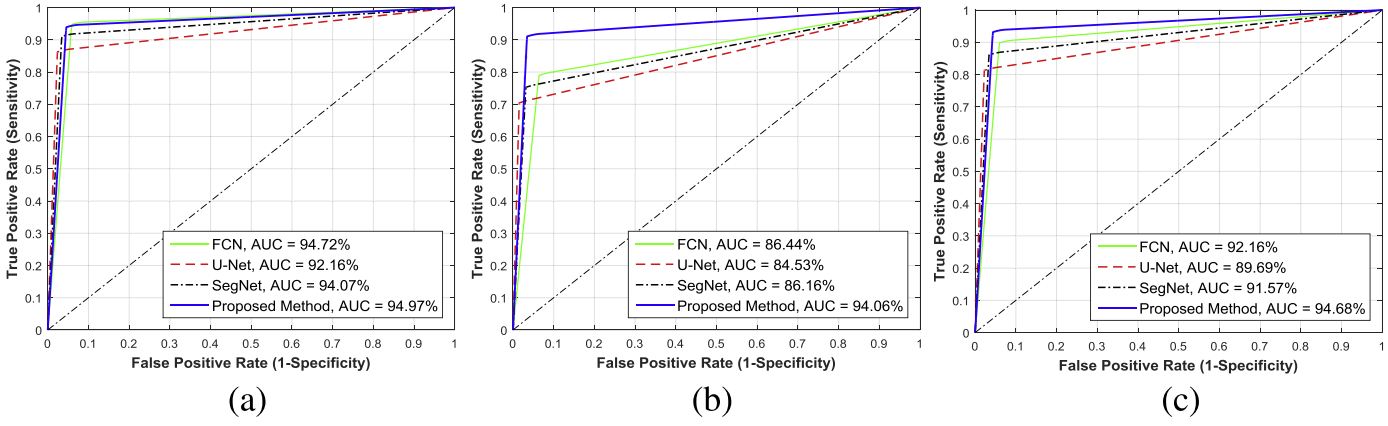

Fig. 6. ROC curves of the tested deep learning segmentation methods for the PH2 skin lesion dataset for (a) benign, (b) melanoma, and (c) overall clinical cases.

4. Discussion

Automatic delineation of skin lesions is highly demanded as a prerequisite step for computer-aided diagnostic systems for melanoma classification. Generally, the overall segmentation performance of deep learning methods improves as the size of the training dataset increases. The effect of different training dataset sizes was shown in [31,71,81], in which the results showed the importance of utilizing larger training sets. In this study, we have augmented the original training data using two approaches. First, we generated HSV images in addition to the RGB. Second, we augmented all the training dermoscopy images by rotating with four rotation angles. The augmented training data assisted the network to learn better representations of the skin lesion features. Our results show the capability of the proposed FrCN method in segmenting skin lesions with higher accuracies than those attained with the conventional deep learning approaches. This is due to the fact that our FrCN method is trained with the full spatial resolution of the input images. In our experiments, we compared the segmentation performance of the proposed FrCN method against three well-known approaches (FCN, U-Net, and SegNet) under the same conditions. In this work, further evaluation of the proposed segmentation method was performed using different dermoscopy images (i.e., the PH2 dataset) without altering any training parameters in the FrCN architecture. Quantitative evaluations with different measures of the recent deep learning techniques are reported in Tables 3 and 4 for the ISBI 2017 test dataset and in Tables 5 and 6 for the PH2 dataset. Significantly, our proposed FrCN method outperformed other skin lesion segmentation techniques. The MCC index demonstrated how the segmented lesions were correlated with the known annotated ground truths. In the case of the ISBI 2017 test dataset, our FrCN method achieved overall increments of 3.92%, 11.99%, and 6.43% in MCC as compared to FCN, U-Net, and SegNet, respectively. However, in the case of the PH2 dataset, our segmentation method achieved an overall MCC of 88.30% which was 4.59%, 5.52%, and 3.66% higher when compared with those of FCN, U-Net, and SegNet, respectively. These promising segmentation results indicate the consistency and capability of the proposed method.

In this study, we performed further analyses to show the segmentation performance for clinical diagnosis cases. In the ISBI 2017 test dataset, benign cases achieved higher segmentation performance than the melanoma and seborrheic keratosis cases, with Jaccard indices of 81.28%, 72.44%, and 69.25%, respectively, as shown in Table 4. This significant improvement for the benign cases is due to the larger existing percentage of these cases in the training dataset (i.e., 1372/2000 cases) and in the test dataset (i.e., 393/600), as clearly demonstrated in the database distribution in Table 1. Note that, due to the inequality of the numbers of each diagnostic type, the segmentation results of the overall indices shown in Tables 3, 4, 5, and 6 were not calculated as direct averages of all cases, but as percentages of their existence.

Fig. 5 illustrates some examples of the segmentation results of the proposed FrCN versus the FCN, U-Net, and SegNet methods compared to the ground truth contours. This figure shows the segmentation results of each clinical diagnostic case in both the ISBI 2017 and PH2 datasets. All methods showed reasonable segmentation performance on the clear skin lesion cases such as those presented in Fig. 5(a) and d–f. In contrast, our FrCN method achieved better segmentation when the dermoscopy image contained hair obstacles, as shown in Fig. 5(b). Also, Fig. 5(c) illustrates how the proposed method segments one of the challenging skin lesion tasks that involved low contrast.

Additionally, the training computation time per epoch and the test time per a single dermoscopy image for each of the segmentation methods are shown in Table 7. Clearly, the proposed FrCN segmentation method seems feasible for future medical practices, since less than 10 s of processing time were required to segment the potential lesion in the dermoscopy image. The computational speed during the training phase of our proposed FrCN method was faster than other segmentation methods. Table 7 shows how fast our FrCN method in reducing the training time of the FCN method to half (i.e., from 651 to 315 s per epoch). This is due to that our FrCN method was executed without the need for the subsampling processes, but keeping the same convolution operations. In summary, our proposed deep FrCN segmentation method outperformed the most popular deep learning approaches of FCN, U-Net, and SegNet on two different dermoscopy datasets.

Table 7
Measurements in seconds of the training time per epoch and test time per dermoscopy image.

Method	Training time/epoch	Test time/single image
FCN	651	10.6
U-Net	577	10.9
SegNet	543	11.13
Proposed FrCN	315	9.7

Table 8
Performance (%) of the proposed FrCN method compared to the latest studies in the literature on skin lesion segmentation.

References	*SEN*	*SPE*	*DIC*	*JAC*	*ACC*
Yuan et al. (CDNN) [58]	82.50	97.50	84.90	76.50	93.40
Li et al. [85]	82.00	97.80	84.70	76.20	93.20
Bi et al. (ResNets) [86]	80.20	98.50	84.40	76.00	93.40
Lin et al. (U-Net) [55]	–	–	77.00	62.00	–
Proposed FrCN	85.40	96.69	87.08	77.11	94.03

Finally, to show the robustness of the proposed FrCN segmentation method, we present a comparison with the latest studies in the literature, as shown in Table 8. All of the segmentation methods in this table were evaluated using the overall diagnostic cases in the ISBI 2017 challenge dataset. In this table, we provide a comparison with the top three methods in the ISBI 2017 challenge, which are described in [58, 85], and [86], respectively. These segmentation methods were ranked according to the performance of the Jaccard index. It is clear that our FrCN method outperformed the top CDNN method [58] with a marginal increment of 0.61% in terms of the Jaccard index. In contrast, the Deep Residual Networks (ResNets) method [86] exhibited a higher specificity of 98.50% for the skin lesion segmentation. Although the skin lesion U-Net-based histogram equalization method in [55] outperformed the clustering algorithm, its segmentation performance was poorer compared to the recent deep learning approach, with dice and Jaccard indices of 77.00% and 62.00%, respectively. With regard to the overall pixel-wise segmentation results of the skin lesion task, our FrCN method proved its feasibility and effectiveness compared to the others with a segmentation accuracy of 94.03%. The higher segmentation achievement of our proposed FrCN method against others is due to that FrCN directly learns the full resolution features of each pixel through the convolutional layers of the network. It should be noted that although we used the fixed image size in this study which was a requirement for the compared conventional segmentation techniques, one can utilize any whole image size in FrCN without resizing. We conclude that the convolutional process in the segmentation task could extract and learn better specific and prominent features using the full spatial resolution without the need for the subsampling operation.

The limitations of this work are as follows. Although the proposed FrCN segmentation method outperformed the other deep learning approaches, it still needs enhancement, in particular with regard to its sensitivity. Moreover, it needs a larger training dataset

in order to achieve higher confidence of the deep learning architecture. In addition, some failure cases such as over- or under-segmented skin lesions are required for further segmentation improvement. In fact, our FrCN network is able to take the input image of any arbitrary size since there are no subsampling layers. However, there is a requirement on the computational resources (i.e., enough memory), since the original dermoscopy images come with very large sizes (i.e., from 540×722 to 4499×6748). In this study, image resizing is applied as a preprocessing step for two reasons. One is to compare the conventional segmentation methods which require the fixed image size. The other is considering our computation resources.

5. Conclusion

In this work, we presented a new full resolution convolutional network (FrCN) method for skin lesion segmentation. Unlike the previous well-known deep learning segmentation approaches, the proposed FrCN method is able to generate full spatial resolution features for each pixel of the input dermoscopy images, leading to an improvement in the performance of pixel-wise segmentation. We evaluated our method using two publicly available databases, the ISBI 2017 challenge and PH2 datasets. The results showed that the proposed FrCN method outperformed the recent deep learning FCN, U-Net, and SegNet approaches. In future work, a larger number of training dermoscopy images is needed to improve the segmentation performance of each class. Developing a computer-aided diagnostic system using segmented lesions maps is also needed to distinguish between the melanoma, benign, and normal skin lesions.

Acknowledgments

This work was supported by International Collaborative Research and Development Programme (funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) (N0002252).

Conflict of interest statement

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] "SEER Cancer Stat Facts: Melanoma of the Skin,". National Cancer Institute, 2017. [Online]. Available: http://seer.cancer.gov/statfacts/html/melan.html. [Accessed 20 01 2017].
[2] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2017, Cancer J. Clin. 67 (1) (2017) 7–30.
[3] American Cancer Society, Cancer Facts and Figures 2016, American Cancer Society, Atlanta, 2016.
[4] C.M. Balch, J.E. Gershenwald, S-J. Soong, J.F. Thompson, M.B. Atkins, D.R. Byrd, A.C. Buzaid, et al., Final version of 2009 AJCC melanoma staging and classification, J. Clin. Oncol. 27 (36) (2009) 6199–6206.
[5] M.E. Vestergaard, P.H.P.M. Macaskill, P.E. Holt, S.W. Menzies, Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting, Br. J. Dermatol. 159 (3) (2008) 669–676.
[6] M. Binder, M. Schwarz, A. Winkler, A. Steiner, A. Kaider, K. Wolff, H. Pehamberger, Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists, Arch. Dermatol. 131 (3) (1995) 286–291.
[7] M.E. Celebi, H. Iyatomi, G. Schaefer, W.V. Stoecker, Lesion border detection in dermoscopy images, Comput. Med. Imaging Graph. 33 (2) (2009) 148–153.
[8] M.E. Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, G. Schaefer, A state-of-the-art survey on lesion border detection in dermoscopy images, in: Dermoscopy Image Analysis, 2015, pp. 97–129.
[9] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, H. Kittler, Automated melanoma recognition, IEEE Trans. Med. Imaging 20 (3) (2001) 233–239.
[10] G. Schaefer, B. Krawczyk, M.E. Celebi, H. Iyatomi, An ensemble classification approach for melanoma diagnosis, Memet. Comput. 6 (4) (2014) 233–240.
[11] M. Silveira, A.C. Nascimento, J.S. Marques, A.R. Marçal, T. Mendonça, S. Yamauchi, J. Maeda, J. Rozeira, Comparison of segmentation methods for melanoma diagnosis in dermoscopy images, IEEE J. Sel. Top. Signal Process. 3 (1) (2009) 35–45.
[12] M.E. Celebi, Q. Wen, S. Hwang, H. Iyatomi, G. Schaefer, Lesion border detection in dermoscopy images using ensembles of thresholding methods, Skin Res. Technol. 19 (1) (2013) e252–e258.
[13] K. Møllersen, H.M. Kirchesch, T.G. Schopf, F. Godtliebsen, Unsupervised segmentation for digital dermoscopic images, Skin Res. Technol. 16 (4) (2010) 401–407.
[14] F. Peruch, F. Bogo, M. Bonazza, V.-M. Cappelleri, E. Peserico, Simpler, faster, more accurate melanocytic lesion segmentation through meds, IEEE Trans. Biomed. Eng. 61 (2) (2014) 557–565.
[15] M.E. Yüksel, M. Borlu, Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic, IEEE Trans. Fuzzy Syst. 17 (4) (2009) 976–982.
[16] F. Xie, A.C. Bovik, Automatic segmentation of dermoscopy images using self–generating neural networks seeded by genetic algorithm, Pattern Recognit. 46 (3) (2013) 1012–1019.
[17] H. Zhou, G. Schaefe, A.H. Sadka, M.E. Celebi, Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images, IEEE J. Sel. Top. Signal Process. 3 (1) (2009) 26–34.
[18] S. Kockara, M. Mete, V. Yip, B. Lee, K. Aydin, A soft kinetic data structure for lesion border detection, Bioinformatics 26 (12) (2010) i21–i28.
[19] S. Suer, S. Kockara, M. Mete, An improved border detection in dermoscopy images for density based clustering, BMC Bioinformatics 12 (10) (2011) S12.
[20] Q. Abbas, M.E. Celebi, I.F. García, Skin tumor area extraction using an improved dynamic programming approach, Skin Res. Technol. 18 (2) (2012) 133–142.
[21] Q. Abbas, M.E. Celebi, I.F. García, M. Rashid, Lesion border detection in dermoscopy images using dynamic programming, Skin Res. Technol. 17 (1) (2011) 91–100.
[22] M.E. Celebi, H.A. Kingravi, H. Iyatomi, Y.A. Aslandogan, R.H. Mos, R.H. Moss, J.M. Malters, Border detection in dermoscopy images using statistical region merging, Skin Res. Technol. 14 (3) (2008) 347–353.
[23] H. Zhou, G. Schaefer, M.E. Celebi, F. Lin, T. Liu, Gradient vector flow with mean shift for skin lesion segmentation, Comput. Med. Imaging Graph. 35 (2) (2011) 121–127.
[24] M. Mete, N.M. Sirakov, Lesion detection in demoscopy images with novel density-based and active contour approaches, BMC Bioinformatics 11 (6) (2010) S23.
[25] B. Erkol, R.H. Moss, R.J. Stanley, W.V. Stoecker, E. Hvatum, Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes, Skin Res. Technol. 11 (1) (2005) 17–26.
[26] H. Wang, R.H. Moss, X. Chen, R.J. Stanley, W.V. Stoecker, M.E. Celebi, J.M. e. a. Malters, Modified watershed technique and post-processing for segmentation of skin lesions in dermoscopy images, Comput. Med. Imaging Graph. 35 (2) (2011) 116–120.
[27] S. Pathan, K.G. Prabhu, P.C. Siddalingaswamy, Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review, Biomed. Signal Process. Control 39 (2018) 237–262.
[28] K. Korotkov, R. Garcia, Computerized analysis of pigmented skin lesions: a review, Artif. Intell. Med. 56 (2) (2012) 69–90.
[29] N. Dhungel, G. Carneiro, A.P. Bradley, A deep learning approach for the analysis of masses in mammograms with minimal user intervention, Med. Image Anal. 37 (2017) 114–128.
[30] G. Carneiro, J. Nascimento, A.P. Bradley, Automated analysis of unregistered multi-view mammograms with deep learning, IEEE Trans. Med. Imaging 36 (11) (2017) 2355–2365.
[31] M.A. Al-masni, M.A. Al-antari, J-M. Park, G. Gi, T.-Y. Kim, P. Rivera, E. Valarezo, M-T. Choi, S-M. Han, T-S. Kim, Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system, Comput. Methods Programs Biomed. 157C (2018) 85–94.
[32] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 779–788.
[33] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.v.d. Laak, B.v. Ginneken, C. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[34] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, Y. Fan, A deep learning model integrating FCNNs and CRFs for brain tumor segmentation, Med. Image Anal. 43 (2018) 98–111.
[35] E. Gibson, W. Li, C. Sudre, L. Fidon, D. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y H., T W., NiftyNet: a deep-learning platform for medical imaging, Comput. Methods Programs Biomed. 158 (2018) 113–122.
[36] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 51 (7553) (2015) 436–444.
[37] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of Advances in Neural Information Processing Systems 25, 2012, pp. 1097–1105.
[38] M.A. Al-Masni, M.A. Al-Antari, J.M. Park, G. Gi, T.Y. Kim, P. Rivera, E. Valarezo, S.-M. Han, T-S. Kim, Detection and classification of the breast abnormalities in digital mammograms via regional Convolutional Neural Network, in: Proceedings of 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2017), 2017, pp. 1230–1233. Jeju Island, Republic of Korea.

[39] C. Cernazanu-Glavan, S. Holban, Segmentation of bone structure in X-ray images using convolutional neural network, Adv. Electr. Comput. Eng. 13 (1) (2013) 87–94.

[40] D. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, in: Proceedings of Advances in Neural Information Processing Systems, 2012, pp. 2843–2851.

[41] M. Melinščak, P. Prentašić, S. Lončarić, Retinal vessel segmentation using deep neural networks, VISAPP 2015 10th International Conference on Computer Vision Theory and Applications, 2015.

[42] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.

[43] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[44] E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (4) (2017) 640–651.

[45] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," arXiv: 1704.06857, 2017.

[46] H. Noh, S. Hong, B. Han, Learning deconvolution network for semantic segmentation, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1520–1528.

[47] V. Badrinarayanan, A. Kendall and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," arXiv: 1505.07293, 2015.

[48] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected CRFs, International Conference on Learning Representations (ICLR), 2015 arXiv:1412.7062v4.

[49] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 834–848.

[50] L. Yu, H. Chen, Q. Dou, J. Qin, P.-A. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE Trans. Med. Imaging 36 (4) (2017) 994–1004.

[51] D. Gutman, N.C. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," arXiv: 1605.01397, 2016.

[52] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, D. Feng, Dermoscopic image segmentation via multi-stage fully convolutional networks, IEEE Trans. Biomed. Eng. 64 (9) (2017) 2065–2074.

[53] T. Mendonça, P.M. Ferreira, J.S. Marques, A.R. Marcal, J. Rozeira, PH 2-A dermoscopic image database for research and benchmarking, in: Proceedings of 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 5437–5440.

[54] Y. Yuan, M. Chao, Y.-C. Lo, Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance, IEEE Trans. Med. Imaging 36 (9) (2017) 1876–1886.

[55] B.S. Lin, K. Michael, S. Kalra and H.R. Tizhoosh, "Skin lesion segmentation: U-Nets versus clustering," arXiv: 1710.01248, 2017.

[56] N.C. Codella, D. Gutman, M.E. Celebi, B. Helba, M.A. Marchetti, S.W. Dusza and A. e. a. Kalloo, "Skin lesion analysis toward melanoma detection: a challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the International Skin Imaging Collaboration (ISIC)," arXiv: 1710.05006, 2017.

[57] M. Goyal and M.H. Yap, "Multi-class semantic segmentation of skin lesions via fully convolutional networks," arXiv: 1711.10449, 2017.

[58] Y. Yuan, M. Chao and Y.-C. Lo, "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks," arXiv: 1703.05165, 2017.

[59] "ISIC, 2017. Skin lesion analysis towards melanoma detection," 2017. [Online]. Available: [Accessed 08 02 2018] https://challenge.kitware.com/#challenge/n/ISIC_2017%3A_Skin_Lesion_Analysis_Towards_Melanoma_Detection.

[60] C. Angermueller, T. Pärnamaa, L. Parts, O. Stegle, Deep learning for computational biology, Mol. Syst. Biol. 12 (7) (2016) 878.

[61] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining,Inference,and Prediction, Second Edn., Springer, 2009.

[62] M. Lin, Q. Chen and S. Yan, "Network in network," in rXiv preprint arXiv:1312.4400, 2013.

[63] D. Scherer, A. Müller, S. Behnke, Evaluation of pooling operations in convolutional architectures for object recognition, in: International Conference On Artificial Neural Networks (ICANN), Berlin, Heidelberg, 2010, pp. 92–101.

[64] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv: 1409.1556, 2014.

[65] Q. Guo, F. Wang, J. Lei, D. Tu, G. Li, Convolutional feature learning and Hybrid CNN-HMM for scene number recognition, Neurocomputing 184 (2016) 78–90.

[66] J.H. Park, S.U. Park, M. Zia Uddin, M. Al-antari, M. Al-masni, T.-S. Kim, A single depth sensor based human activity recognition via convolutional neural network, in: Proceedings of 4th World Conference on Applied Sciences, Engineering & Technology, 2016, pp. 329–332.

[67] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic, in: Proceedings of 30th International Conference on Machine Learning (ICML), 2013, p. 3.

[68] X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, 2011, pp. 315–323.

[69] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (1) (2014) 1929–1958.

[70] S. Min, B. Lee, S. Yoon, Deep learning in bioinformatics, Briefings Bioinf. 18 (5) (2017) 851–869.

[71] Z. Jiao, X. Gao, Y. Wang, J. Li, A deep feature based framework for breast masses classification, Neurocomputing 197 (C) (2016) 221–231.

[72] H.-C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (5) (2016) 1285–1298.

[73] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen, H. Greenspan, Chest pathology identification using deep feature selection with non-medical training, Comput. Methods Biomech. Biomed. Eng. 13 (2016) 1–5.

[74] R.K. Samala, H. Chan, L. Hadjiiski, M.A. Helvie, J. Wei, K. Cha, Mass detection in digital breast tomosynthesis: deep convolutional neural network with transfer learning from mammography, Med. Phys. 43 (12) (2016) 6654–6666.

[75] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, How transferable are features in deep neural networks? in: Proceedings of Advances in Neural Information Processing Systems, 2014, pp. 3320–3328.

[76] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vision 115 (3) (2015) 211–252.

[77] E. Szymańska, E. Saccenti, A.K. Smilde, J.A. Westerhuis, Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies, Metabolomics 8 (1) (2012) 3–16.

[78] S. Smit, M.J.v. Breemen, H.C. Hoefsloot, A.K. Smilde, J.M. Aerts, C.G.D. Koster, Assessing the statistical validity of proteomics based biomarkers, Anal. Chim. Acta 592 (2) (2007) 210–217.

[79] "image-segmentation-keras," 2017. [Online]. Available: https://github.com/divamgupta/image-segmentation-keras . [Accessed 20 01 2018].

[80] M. Dong, X. Lu, Y. Ma, Y. Guo, Y. Ma, K. Wang, An efficient approach for automated mass segmentation and classification in mammograms, J. Digital Imaging 28 (5) (2015) 613–625.

[81] M.A. Al-antari, M.A. Al-masni, S.U. Park, J.H. Park, M.K. Metwally, Y.M. Kadah, S.M. Han, T.-S. Kim, An automatic computer-aided diagnosis system for breast cancer in digital mammograms via deep belief network, J. Med. Biol. Eng. (2017) 1–14.

[82] D. Zikic, Y. Ioannou, M. Brown, A. Criminisi, Segmentation of brain tumor tissues with convolutional neural networks, in: Proceedings MICCAI-BRATS, 2014, pp. 36–39.

[83] S. Pereira, A. Pinto, V. Alves, C.A. Silva, Brain tumor segmentation using convolutional neural networks in MRI images, IEEE Trans. Med. Imaging 35 (5) (2016) 1240–1251.

[84] D.M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, J. Mach. Learn. Technol. 2 (1) (2011) 37–63.

[85] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," in arXiv: 1703.00577, 2017.

[86] L. Bi, K. Jinman, E. Ahn and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," in arXiv: 1703.04197, 2017.