ORIGINAL PAPER

Skin Lesion Segmentation with Improved Convolutional Neural Network

Şaban Öztürk[1] · Umut Özkaya[2]

Abstract

Recently, the incidence of skin cancer has increased considerably and is seriously threatening human health. Automatic detection of this disease, where early detection is critical to human life, is quite challenging. Factors such as undesirable residues (hair, ruler markers), indistinct boundaries, variable contrast, shape differences, and color differences in the skin lesion images make automatic analysis quite difficult. To overcome these challenges, a highly effective segmentation method based on a fully convolutional network (FCN) is presented in this paper. The proposed improved FCN (iFCN) architecture is used for the segmentation of full-resolution skin lesion images without any pre- or post-processing. It is to support the residual structure of the FCN architecture with spatial information. This situation, which creates a more advanced residual system, enables more precise detection of details on the edges of the lesion, and an analysis independent of skin color can be performed. It offers two contributions: determining the center of the lesion and clarifying the edge details despite the undesirable effects. Two publicly available datasets, the IEEE International Symposium on Biomedical Imaging (ISBI) 2017 Challenge and PH2 datasets, are used to evaluate the performance of the iFCN method. The mean Jaccard index is 78.34%, the mean Dice score is 88.64%, and the mean accuracy value is 95.30% for the proposed method for the ISBI 2017 test dataset. Furthermore, the mean Jaccard index is 87.1%, the mean Dice score is 93.02%, and the mean accuracy value is 96.92% for the proposed method for the PH2 test dataset.

Keywords Skin lesion segmentation · CNN · FCN · Segmentation · Melanoma

Introduction

Melanoma is one of the most common types of cancer, resulting from the uncontrolled distribution of the skin cell called melanocytes [1]. According to the annual report of the American Cancer Society in the USA, 96,480 cases are diagnosed as new melanoma cases, and the estimated mortality rate is 7230 [2]. Compared with the report in 2017, the number of cases increased by 9370 [3]. Since melanoma is known to be the most lethal skin cancer, this increase is hazardous for human life. Despite this adverse situation, early detection and treatment significantly increase the chances of survival [4]. As a conventional method, dermatologists perform a visual examination for the detection of melanoma. This method is time-consuming, requires a well-trained expert, and suffers from inter-observer variation [5]. Besides, visual similarities between the lesions, variations in lesion shapes, and disruptive factors such as hair and markers are other main difficulties. At this point, technological advances are running with the help of dermatologists. Nowadays, skin lesion images can be obtained easily, thanks to the reach of standard cameras to mobile phones. However, these devices suffer from many problems, such as resolution and light. Much better quality and reliable results can be achieved by the non-invasive imaging tool dermoscopy [6]. Dermoscopy utilizes polarized light to provide an enlarged and illuminated view of the skin area for a more accurate diagnosis of skin lesions [7]. Thus, more in-depth details of the skin surface can be captured. Although dermoscopy is very useful, it still has drawbacks caused by the human factor. Making subjective decisions, being time-consuming, being irreproducible, requiring experience, and being prone to error are the most important ones. According to these studies, it is shown that inexperienced specialists have a success rate of up to 84% in the diagnosis of melanoma [8]. There

✉ Şaban Öztürk
saban.ozturk@amasya.edu.tr

[1] Technology Faculty, Electrical and Electronics Engineering, Amasya University, Amasya, Turkey

[2] Engineering and Natural Science Faculty, Electrical and Electronics Engineering, Konya Technical University, Konya, Turkey

is a high need for computer-aided diagnosis (CAD) methods to achieve higher success in early detection of melanoma, where early diagnosis is essential, and to protect human health. Fortunately, many problems can be solved successfully thanks to the technology developed in recent years and to the machine learning (ML) techniques advancing in this direction. When the development and application areas of ML techniques are examined, it is seen that they have been used actively in the solution of medical problems in recent years. This situation is developed in order to provide a quick solution to the time-consuming and specialist problems in the medical field, and it is also used as an advisory system.

By analyzing the appearance of a skin lesion using ML systems, the system can produce the result benign or melanoma. To achieve this judgment, pre-processing, segmentation, post-processing, feature extraction, and classification steps are applied. Although each step is essential, the segmentation step stands out because it provides visual image information. An unsuccessful segmentation approach plays a fundamental role in the inaccuracy of the classification result. In the detection of melanoma, the segmentation stage is the most challenging stage due to misleading factors such as the color of the lesions, edge information, hair, markers, bad frames, size, blood vessels, and air bubbles.

The methods used for skin lesion segmentation are classified according to various characteristics in the literature [9]. Usually, this number of classes is specified as five, including histogram thresholding methods, unsupervised clustering methods, edge-based methods, active contour methods, and supervised methods [1, 8]. These methods generally use pixel-level features. The success of these low-level features does not reach the desired level. The remarkable success of the deep learning approach in recent years has enabled it to enter almost every field. As it enters the area of skin lesion segmentation, the course of studies in this area has changed considerably. Therefore, in this study, skin lesion studies will be examined in two parts: before convolutional neural network (CNN) and after CNN.

The traditional methods used before the CNN architecture were often based on manual selection of features. Generally, the experience of the researcher affects the success of the process. Since each researcher has a different approach and experience, many approaches have been tried for skin lesion segmentation. In the first studies, histogram-based studies and thresholding methods were widely used [10–14]. Yüksel et al. [12] have proposed a new approach for skin lesion segmentation using fuzzy type 2. Their method was very effective in detecting the uncertainties in the edges of the lesions compared with other studies of the period. Çelebi et al. [13] used ensembles of thresholding methods to handle the wide variety of dermoscopy images. Their main idea is to find a solution with thresholding fusion to the variety of skin lesion images that allow limited success for a single-threshold coefficient.

Another remarkable study of threshold methods was performed by Peruch et al. [14]. They proposed a five-step method that mimics the behavior of a dermatologist. Accordingly, the methods include pre-processing, dimensionality reduction, blurring, thresholding, and post-processing. In the same period, in addition to thresholding methods, clustering methods, which work quite well in homogeneous regions, were also used frequently [15–19]. Suer et al. [16] introduced boundary-based clustering to reduce the number of neighborhood searches. Their method targets the problem of the excessive number of region queries fired in the clustering process. Kockara et al. [18] used a graph spanner approach for the clustering of melanoma. They represent skin images as graphs to obtain a color pattern. Xie and Bovik [19] performed clustering using an artificial neural network optimized by genetic algorithms. The method they developed has attracted much attention by producing higher scores for their periods than other algorithms. In the same period, segmentation was performed simply by determining the lesion region with edge detection algorithms. However, undesirable lighting, artifacts in the lesion, and fuzzy borders were a problem for edge detectors. Abbas et al. [20] proposed a robust method to remove artifacts and detect lesion borders in dermoscopy images. Firstly, lesion artifacts were removed, then the least-squares method was used to acquire edge points, then dynamic programming is used to find optimal boundaries. Abbas et al. [21] presented a partial solution to this problem with the algorithm they developed. They remove the residues on the lesion by first transforming the color space. Then, the boundaries of the lesion are determined by using the improved dynamic programming method. Region-based methods produce more successful results when the edges are irregular [22–24]. Iyatomi et al. [22] proposed a four-step dermatologist-like lesion segmentation technique for skin lesions. Accordingly, segmentation was performed with the region growing algorithm. Celebi et al. [23] used the statistical region merging (SRM) method for the detection of borders of melanoma. Their SRM method is a color image segmentation technique based on color information. Glaister et al. [24] classified regions in the skin images based on the occurrence of representative texture distribution. With the advancement of technology and the increase in methodological knowledge, the use of semi-automatic methods has started to increase. In this direction, progress has been made towards fully automatic detection with the active counter method [25–27]. Erkol et al. [25] proposed gradient vector flow (GVF) snakes to detect edges of skin lesions. Mete and Sirakov [26] proposed a novel active contour model for fast and accurate detection of skin lesion boundaries. Their method collaborates with the boundary-driven density-based algorithm. Ma and Tavares [27] used a geometric deformable model based on color space conversion. Recently, segmentation methods based on the classification of manually extracted properties with classifiers

have been used [28–30]. These methods are pixel-based methods that perform the classification process automatically and also work according to the extracted features depending on the researcher's experience. Wang et al. [28] proposed the edge object value (EOV) threshold method. They used a neural network classifier to increase lesion ratio estimation and the watershed segmentation algorithm. Winghton et al. [29] examined two different approaches, the first using independent pixel labeling using maximum a posteriori (MAP) estimation and the second using conditional random fields (CRFs). Sadri et al. [30] introduced a fixed-grid wavelet network for segmentation of dermoscopic images. They used R, G, and B values of images as the network input.

Following the introduction of the CNN architecture and then achieving dazzling results for almost all image processing problems, it has attracted the attention of skin lesion researchers. Although the classification process is done automatically in pre-CNN studies, the extracted features are based on the researcher's experience. The success of these manually extracted features is not the same for each skin lesion dataset. What is more, success in the same dataset with different light conditions and different skin structures was rather low. In order to solve these drawbacks, pre-CNN solutions have been developed, which are very complex and have high processing time. The CNN architecture automatically learns and classifies the features. This feature is more successful than other algorithms. Al-masni et al. [1] proposed a full-resolution CNN (FrCN) for the segmentation of skin lesions. FrCn learns features from full-resolution lesion images without the need for pre-processing. Unver et al. [8] combined You Only Look Once (YOLO) and GrabCut algorithms for effective segmentation. Their model has achieved a 90% sensitivity rate on the ISBI 2017 dataset. Alom et al. [31] proposed the NABLA-N network combined with feature fusion methods in decoding for melanoma segmentation. Li and Shen [5] proposed a framework that consists of two fully convolutional residual networks (FCRN) to address segmentation tasks, feature extraction tasks, and classification tasks. They developed a lesion index calculation unit to refine network results. Huang et al. [32] introduced an end-to-end object scale-oriented FCN (OSO-FCNs) for lesion segmentation. Their method is based on the fact that the scale of the lesion affects the segmentation result considerably. Bi et al. [33] developed generative adversarial networks (GANs) with stacked adversarial learning to learn robust features from skin lesion images. Jiang et al. [34] proposed developed GAN for great skin lesion feature representation. They used atrous convolution and concatenating residual layers as the generator. Liu et al. [35] proposed a deep metric learning enhanced neural network (DMLEN) to increase segmentation accuracy. The DMLEN method produces a MAE value of at least 3% lower than other methods. Khan et al. [36] used the transfer learning approach with a deep CNN (Resnet) for feature extraction. They used kurtosis-controlled principle component analysis to choose optimal features.

In this study, we present an effective FCN architecture that does not require any pre-processing for skin lesion segmentation. The proposed architecture is an improved FCN structure that is not affected by disturbing factors such as hair, ruler markers, indistinct boundaries, and illumination problems. It has two important contributions: 1) it is not affected by any disturbing factors and is stable to the change of color and light information, and 2) it is highly successful against the irregularity of the lesion edges. In addition to RGB color space, S component from HSV color space, I component from YIQ color space, Cb component from YCbCr color space, and Z component from XYZ color space are used to eliminate the effect of disturbing factors. Segmentation is performed without affecting the light, skin color, and lesion color conditions. To protect the edge information, a more intensive deconvolution technique is used in the decoder section of the CNN architecture. The success of the proposed method is tested using the IEEE International Symposium on Biomedical Imaging (ISBI) 2017 Challenge [37] and PH2 datasets [38]. The results obtained are more successful than other state-of-the-art studies in the literature. One of the most important reasons for this is that unique design is applied to skin lesion problems. The novelty of the proposed method is to support the residual structure of the FCN architecture with spatial information. This situation, which creates a more advanced residual system, enables more precise detection of details on the edges of the lesion, and an analysis independent of skin color can be performed.

The structure of this article is the following. In the "Methodology" section, the definition of skin lesion problems and the proposed method is presented. The dataset and experimental results are reported in the "Experiments and Experimental Results" section. Finally, the conclusion is given in the "Conclusion" section.

Methodology

Definition of Fundamental Skin Lesion Problems

Skin lesion segmentation is a challenging task due to the environment and structure of the lesion. These problems are very challenging for CAD systems. Although these problems are usually solved by pre-processing methods, it creates computational load for the whole system. In addition, when there is not enough experience, these negative factors decrease the success of the system. When the studies in the literature are examined, an additional pre-processing process is generally applied to remove these residues before the main

segmentation algorithm. Pre-processing algorithms are generally used to solve a single problem. A pre-processing algorithm is designed to eliminate hairs, a different pre-processing algorithm is designed to eliminate the effect of markers, and a different pre-processing algorithm is designed to react to non-dominant lesions. Various pre-processing algorithms are available in the literature for many such problems. However, when studies with a large dataset are available, all of the problems mentioned are present and more. Figure 1a shows a brown lesion that is quite distinct and easy to detect. Figure 1b shows a lesion surrounded by hairs. Figure 1c shows a lesion that is difficult to detect by black hair. Figure 1d shows the distorting effect of different colored markers. Figure 1e shows white acceptance on a highly indeterminate lesion. Figure 1f shows a highly indistinct lesion. In Figure 1g, there is a lesion with some areas not visible. Figure 1h shows an image of a lesion with a black substance in the marginal region. When many pre-processing algorithms are used to solve all these problems, process load and response time become unacceptable. An algorithm created without considering these problems produces undesirable results even if it is a deep learning method. The most appropriate solution for this problem is to solve these problems by creating attention in the algorithm. This proposed approach is a convenient approach that will reduce both processing load and processing time.

One of the most challenging parts of the segmentation of skin lesion images is the uncertainty at the boundaries of the lesion. While skin lesions are usually visible, some lesions are challenging to react with the eye. Dermoscopy is used for such lesions. Although dermoscopy images allow us to see a lot of information about lesion easily, it is difficult to obtain information about lesion boundaries. Figure 2a, b, c, and g shows images of the skin lesion obtained by dermoscopy. Although the lesion centers are quite clear, there are problems with the lesion boundaries. These problems are clearly understood when Fig. 2d, e, f, and i are examined. When these images are examined, it is seen that the edge information in ground truth images is quite irregular. However, this irregularity is not apparent in skin lesion images.

Another problem is understood when examining Fig. 2h and j. In the appearance of a very prominent and predominant skin lesion, the boundaries between the skin and the lesion are rather obscure. These two fundamental problems constitute a contradiction in itself. Edge estimation is made for the region where there is no tissue difference with healthy skin. This problem is the most fundamental problem to be overcome for today's CAD systems.

Proposed Method Overview

In this study, a single CNN architecture (called iFCN (improved FCN)) with end-to-end training capability is proposed for the solution of skin lesion segmentation problems. The proposed architecture is capable of dealing with the two main problems mentioned above at the same time. The iFCN architecture and layers are shown in Fig. 3. The proposed method is inspired by the classical FCN architecture [39]. For semantic segmentation, CNN architectures first extract image features and then derive the target image from these features. When divided into two parts, these parts are called encoder and decoder parts. In the section called encoder or subsampling, image sizes are slowly decreasing depending on filter sizes. The most important layers of this section are the convolution layer, pooling layer, ReLU, dropout, and concatenate layer. In terms of operation, the convolution layer is used to extract features from the image. Although it is generally used in 3×3 neighborhood size in the literature, the neighborhood size is changed according to the application type. Convolution layer is calculated as in Eq. 1.
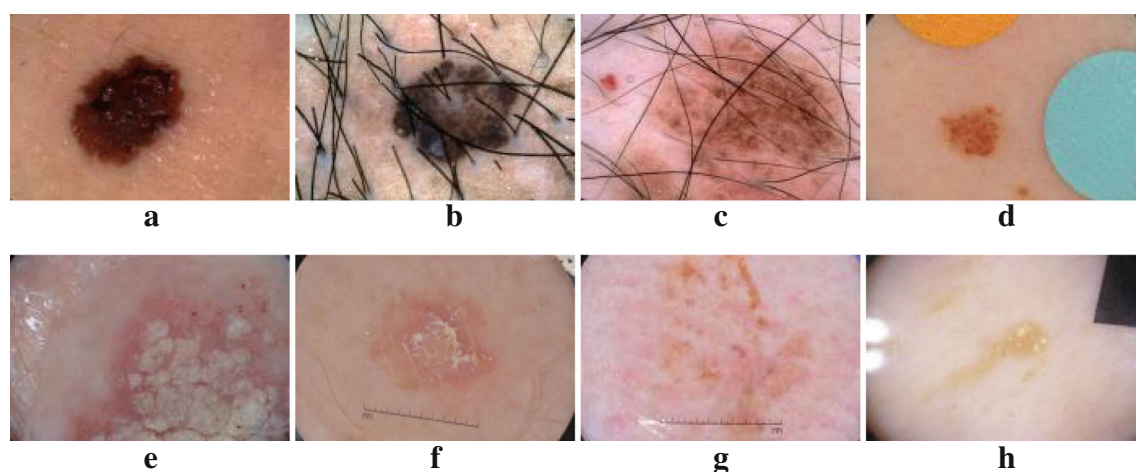
Fig. 1 Undesirable residues and some problems of skin lesion images. **a** Brown lesion. **b** Black hairs. **c** Black hairs with the brown lesion. **d** Markers. **e** White residues on the lesion. **f** White residues on the lesion. **g** White and indistinct lesion. **h** Hard scene for lesion

Fig. 2 Indistinct boundaries. a, b, c, g Lesion images. d, e, f, i Ground truth images of a, b, c, and g. h Boundaries of a skin lesion. j Ground truth of h image
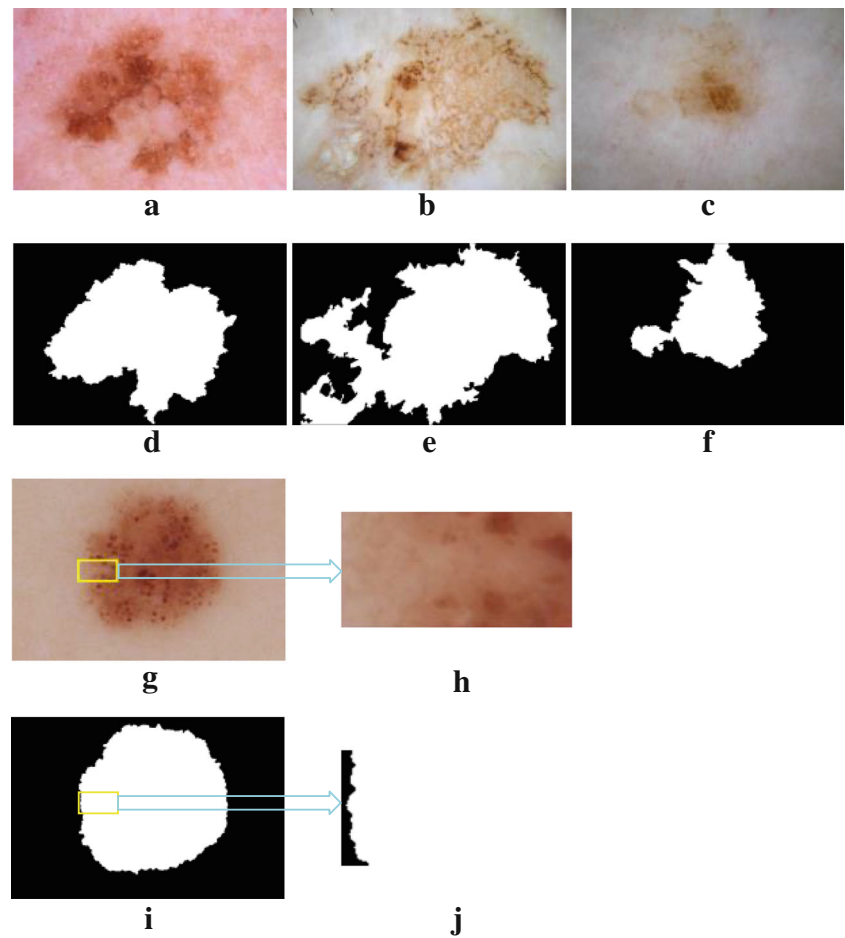
$$y_i^l = f\left(\sum_j x_j^{l-1} \otimes w_{ij}^l + b_i^l\right) \quad (1)$$

where f represents activation function, w represents weights of convolution operator, b represents bias, and x represents the input matrix of the convolutional layer. In the decoder section, the pooling layer has the most significant effect on reducing the size of the input image. This layer dramatically reduces the image size according to the selected neighborhood values. In this way, while transferring the most important features in the image to the next layers, it prevents the insignificant features from creating unnecessary processing load. Although this process is very efficient for CNN architectures used in classification tasks, it poses some problems for segmentation architectures. Pooling layers with high neighborhood values lead to the softening of the edge information. The max-pooling operation is calculated as in Eq. 2.

$$P_{jm} = \max_{k=1}^{r}\left(x_{j(m-1)n+k}\right) \quad (2)$$
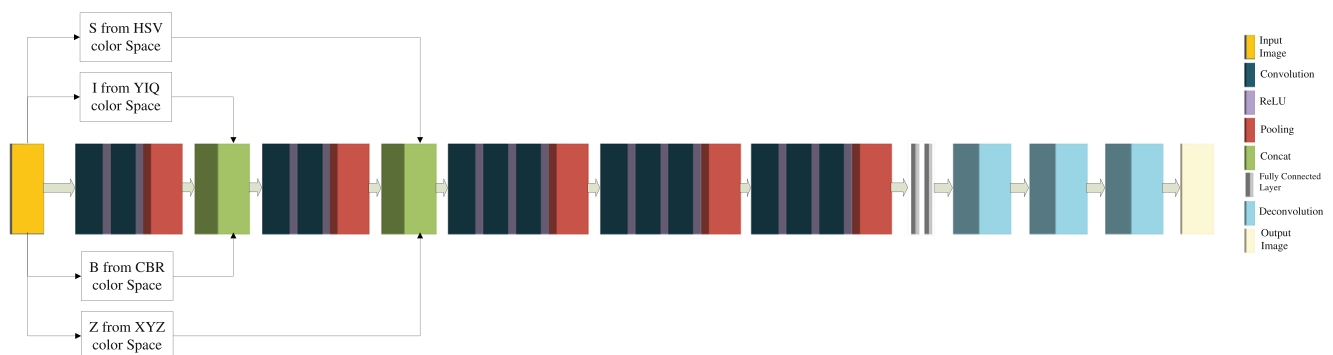
Fig. 3 The proposed iFCN architecture

To solve the memorization problem, which is the biggest problem of deep networks, the ReLU layer disrupts network linearity. The output of the ReLU layer is calculated as in Eq. 3.

$$R(i) = \max(0, i) = \begin{cases} i, \text{if } x \geq 0 \\ 0, \text{if } x < 0 \end{cases} \tag{3}$$

Another problem of deep networks is that neurons learn in cooperation and proportion to each other. The dropout layer is used to solve this problem. It excludes a certain amount of parameters from each training iteration. The concatenate layer is used for combining feature maps in residual architectures.

After the image feature maps are obtained in the decoder section, the target image must be obtained at the same resolution. The deep features obtained from the encoder section are processed and resized to the target image size in the decoder section. In the decoder or up-sampling section, new pixels or features are created from each pixel or feature point and transferred to the next layers. In this section, the feature map size gradually increases as the layers progress. There are various approaches in the literature for the up-sampling part, which is very important for determining the boundaries of objects. The best known of these approaches are the unpooling layer, deconvolution layer, atrous convolution layer, and bilinear interpolation [40, 41]. Details of the layers and parameters of the iFCN architecture are shown in Table 1.

The iFCN architecture offers two important contributions to skin lesion segmentation. The first is to determine the center of the lesion without being affected by environmental factors in the lesion images. Figure 1 shows various skin lesion problems, and the pre-processing algorithms used to solve these problems have been mentioned previously. These problems are usually caused by image or skin problems. The proposed architecture can detect the lesion area without the need for additional pre-processing. For this purpose, the elemental powers of different color spaces are utilized. Each color space provides different image representations based on information such as light information in the image and color intensity information. But the information in some channels of these color spaces is equal to or quite close to each other. For this purpose, the efficiency of color spaces for skin lesion segmentation is investigated. Figure 4 shows the R, G, and B components of the RGB color space, the S component from the HSV color space, I component from the YIQ color space, the C component from the YCbCr color space, and the Z component from the XYZ color space.

As can be seen from Fig. 4, component R is generally used for the detection of hair and dark markers. Lesion areas are mostly absent in component R. Component G is useful for fluctuation in lesion specimens, and component B is useful for locating the lesion. Component S clearly shows the lesion after hair removal. Component I is useful for locating dark markers on lesion images. The Cb component reveals parts

Table 1 Layers and parameters of iFCN

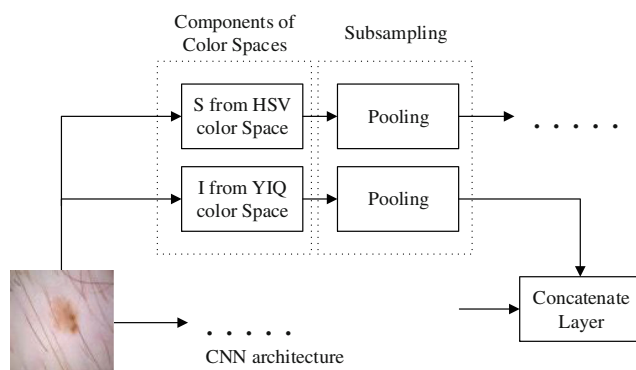|	Layer	Filter size, maps
Encoder	Conv_1	3 × 3, 3-64
	Relu_1	-
	Conv_2	3 × 3, 64-64
	Relu_2	-
	Pool_1	2 × 2
	Concat_1	-
	Conv_3	3 × 3, 64-128
	Relu_3	-
	Conv_4	3 × 3, 128-128
	Relu_4	-
	Pool_2	2 × 2
	Concat_2	-
	Conv_5	3 × 3, 128-256
	Relu_5	-
	Conv_6	3 × 3, 256-256
	Relu_6	-
	Conv_7	3 × 3, 256-256
	Relu_7	-
	Pool_3	2 × 2
	Conv_8	3 × 3, 256-512
	Relu_8	-
	Conv_9	3 × 3, 512-512
	Relu_9	-
	Conv_10	3 × 3, 512-512
	Relu_10	-
	Pool_4	2 × 2
	Conv_11	3 × 3, 512-512
	Relu_11	-
	Conv_12	3 × 3, 512–512
	Relu_12	-
	Conv_13	3 × 3, 512-512
	Relu_13	-
	Pool_5	2 × 2
	FCL_1	1 × 1, 4096-4096
	Dropout_1	0.5
	FCL_2	1 × 1, 4096-4096
	Dropout_2	0.5
Decoder	Deconv_1	16 × 16, 2-2
	Deconv_2	16 × 16, 2-2
	Deconv_3	16 × 16, 2-2
	Softmax	-

of the lesion that are not visible in more specific numbers. The Z component contains all the edge details of the lesion. When all these components are applied to the architecture at the same time, the iFCN architecture can automatically learn

Fig. 4 Color space effect on lesion images. **a** Original images. **b** R component from RGB. **c** G component from RGB. **d** B component from RGB. **e** S component from HSV. **f** I component from YIQ. **g** Cb component from YCbCr. **h** Z component from XYZ
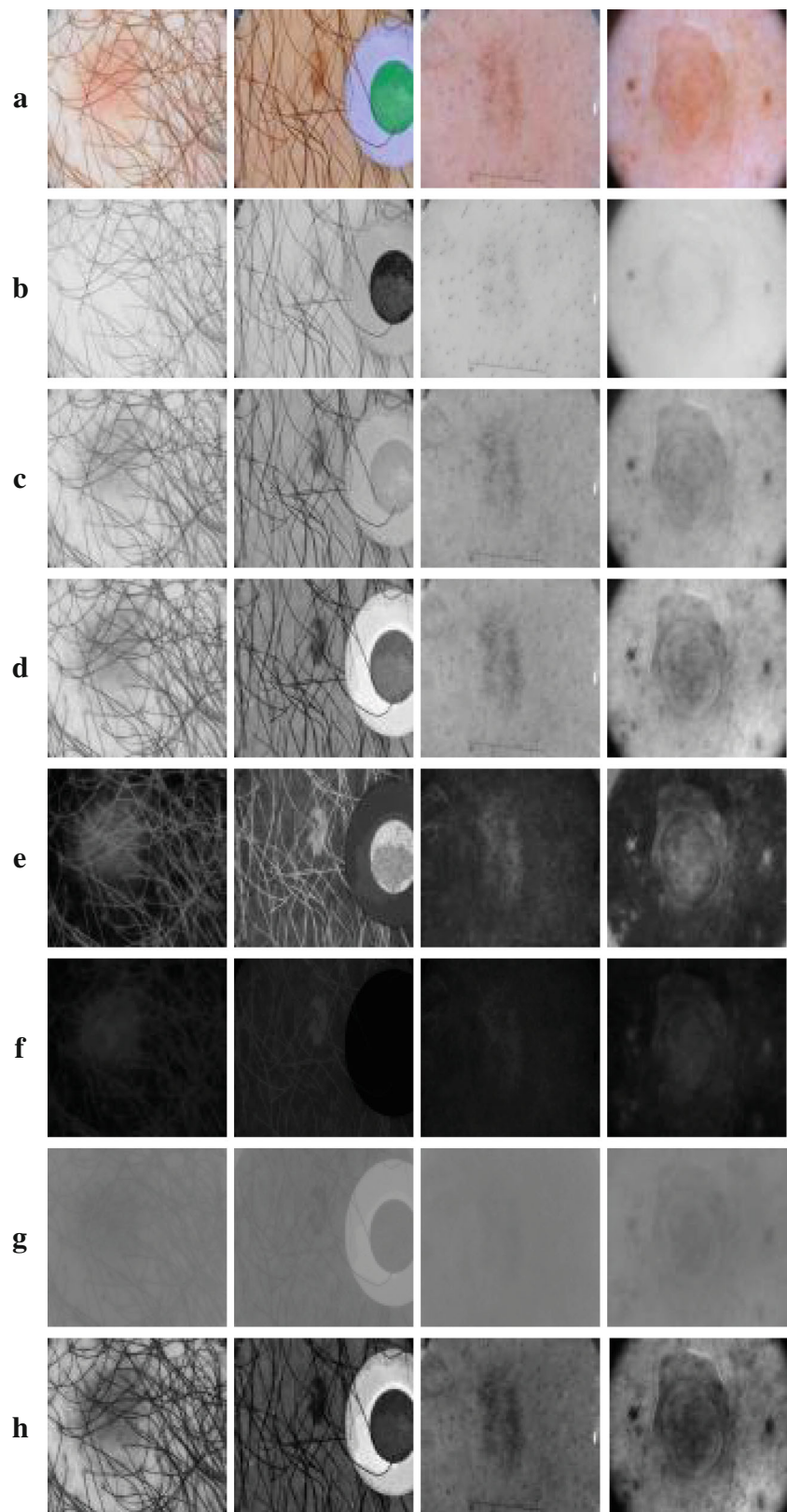
about their contributions. Besides, it can add new meanings to these components. The addition of these components to the architecture is sho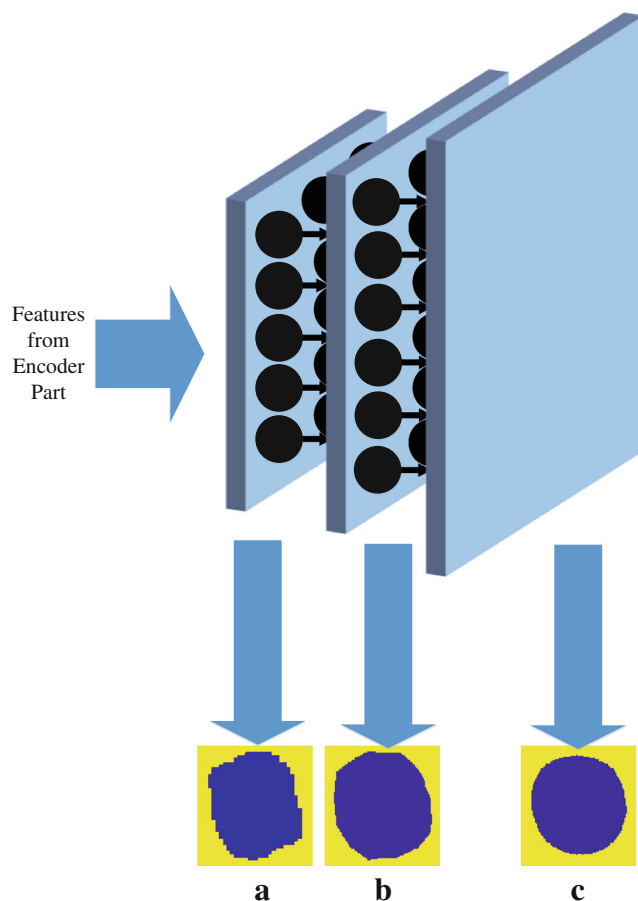wn in Fig. 5. The R, G, and B components are applied as standard to the input of the architecture. I and B components are applied to the first concatenate layer to learn the basic features. S and Z components are applied to the

Fig. 5 Adding color components to the iFCN architecture

Table 2 Data distribution of ISBI 2017 and PH2 datasets

		Dataset	
		ISBI 2017	PH2
Training	Benign nevi	1372	-
	Melanoma	374	-
	Seborrheic keratosis	251	-
	Total	2000	-
Validation	Benign nevi	78	-
	Melanoma	30	-
	Seborrheic keratosis	42	-
	Total	150	-
Test	Benign nevi	393	160
	Melanoma	117	40
	Seborrheic keratosis	90	-
	Total	600	200

second concatenate layer. In this way, unwanted factors are learned as the basic features, and the next layer of mid-level and high-level features are learned. A difference in size occurs when combining the color components S, I, B, and Z in the concatenate layers. To compensate for this image size difference, a subsampling is used after each color component layer.

Another contribution of the iFCN architecture is the ability to represent details on the edges of skin lesions vigorously. A straightforward but practical approach is used for this. It is known that target images are created from feature maps in

Fig. 6 The up-sampling effect. a One up-sampling. b Two up-sampling. c Three up-sampling

the decoder section. The design in this section affects the resolution of the objects in the target image. At the end of the encoder part, the magnification of the features represented by a minimal number of features at a time causes loss of details. For this reason, the up-sampling process is applied gradually to these features. Figure 6 shows the gradual up-sampling of the features and their effect on the object details. Figure 6a shows only the lesion after an up-sampling layer. The margin information of this lesion is rather uncertain. Figure 6b shows the results of two up-sampling layers. Lesion details are more apparent. Figure 6c shows the results obtained when three up-sampling layers are used. As it is seen, if the up-sampling process is made softer, the details are preserved.

Experiments and Experimental Results

Data

The segmentation performance of the proposed iFCN architecture is tested using two well-known and public datasets. The first of these is ISBI 2017 data called "Skin Lesion Analysis Towards Melanoma Detection" [37]. The ISBI 2017 dataset includes 2000 training images, 150 validation images, and 600 test images. It consists of 8-bit RGB lesion images of different sizes such as 540×722, 2048×1536, 1503×1129, and 4499×6748. Lesions in the ISBI dataset are labeled as benign, melanoma, and seborrheic keratosis. Also, each image in the dataset is annotated by expert dermatologists, and ground truth images are created. The other dataset is the skin lesion dataset called PH2 [38]. It was created at the
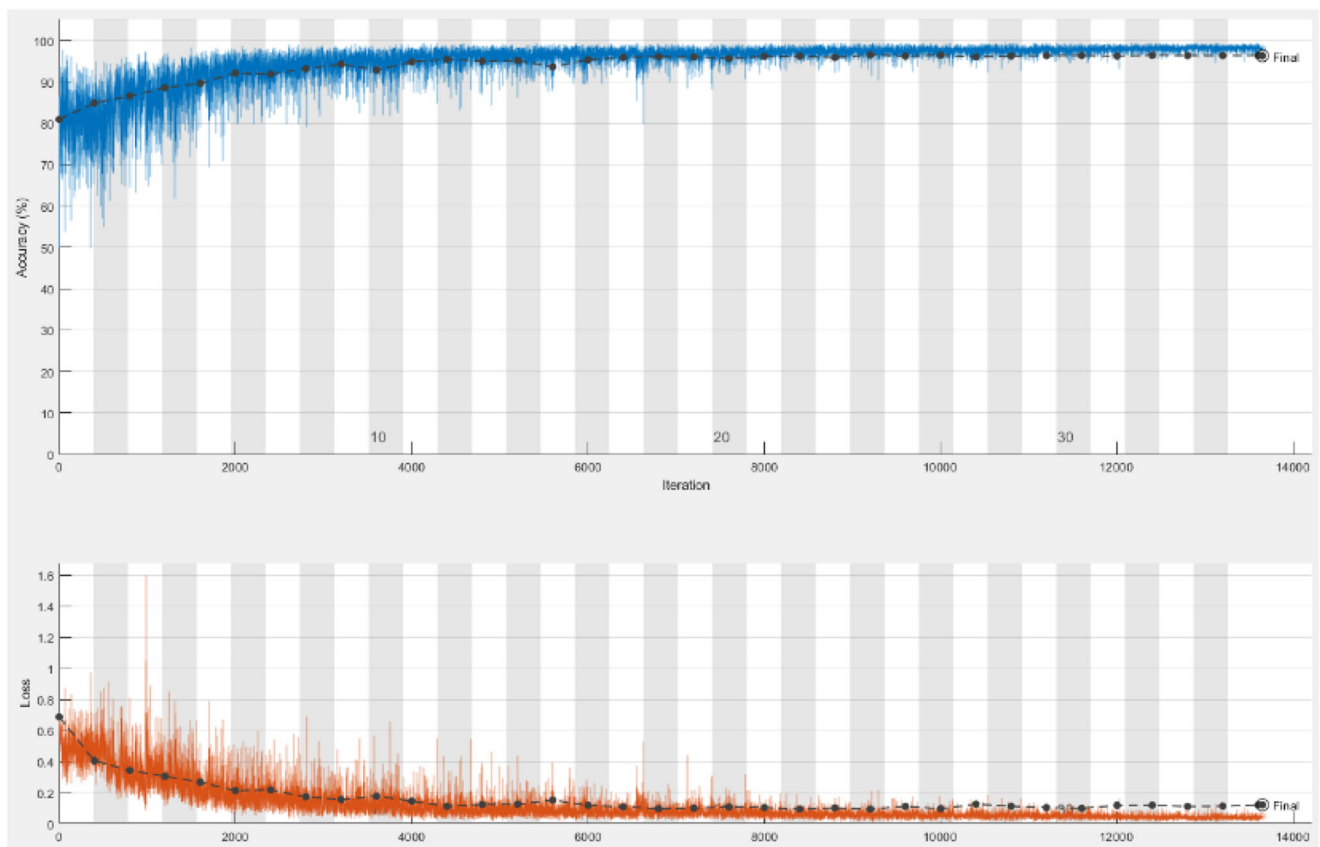
Fig. 7 Training and validation curves of the iFCN architecture

Hospital Pedro Hispano with the help of the research group Universidade do Porto, Técnico Lisboa in Matosinhos, Portugal. The PH2 dataset consists of 200 skin lesions in total. Eighty of them include atypical nevi cases, 80 of them include common nevi cases, and 40 of them include melanoma cases. Images in the PH2 dataset were captured under the same conditions, and the size of all images in this dataset was 768×560 pixels. Segmentation masks of this dataset were drawn by expert dermatologists. In this study, the PH2 dataset is used only for testing. Class information and distribution of data for both datasets are shown in Table 2.

Experimental Results

iFCN is trained on a computer with Intel Core i7-7700K CPU (4.2 GHz), 32-GB DDR4 RAM, and NVIDIA GeForce GTX 1080 graphic card.

To evaluate the success of the proposed iFCN method fairly, the same evaluation metrics are used as other studies in the literature. For this purpose, the sensitivity parameter is first calculated as in Eq. 4. The sensitivity parameter represents the ratio of accurately assigned lesion pixels in the image. The specificity parameter showing the false labeling rate of non-lesion pixels is

Table 3 Comparison of segmentation performance for the ISBI 2017 dataset

Methods	Sensitivity %	Specificity %	Accuracy %	Dice %	Jaccard %
U-Net	67.15	97.24	90.14	76.27	61.64
SegNet	80.05	95.37	91.76	82.09	69.63
FCN	79.98	96.66	92.72	83.83	72.17
Jianu et al. [42]	72	89	81	-	-
eVida [43]	86.9	92.3	88.4	76	66.5
Li et al. [5]	82	97.8	93.2	84.7	76.2
Unver and Ayan [8]	90.82	92.68	93.39	84.26	74.81
FrCN [1]	85.40	96.69	94.03	87.08	77.11
iFCN	85.44	98.08	95.30	88.64	78.34

Table 4 Comparison of segmentation performance for the PH2 dataset

Methods	Sensitivity %	Specificity %	Accuracy %	Dice %	Jaccard %
U-Net	81.63	97.76	92.55	87.61	77.95
SegNet	86.53	96.61	93.36	89.36	80.77
FCN	90.30	94.02	92.82	89.03	80.22
Unver and Ayan [8]	83.63	94.02	92.99	88.13	79.54
FrCN [1]	93.72	95.65	95.08	91.77	84.79
iFCN	96.88	95.31	96.92	93.02	87.1

calculated as in Eq. 5. The accuracy parameter is calculated as in Eq. 6. The accuracy parameter shows the success rate of the whole segmentation. The Dice coefficient is used to measure how effectively a segmentation process works. It only examines the relationship between the segmented field and the ground truth. The Dice coefficient is calculated as in Eq. 7. The Jaccard index calculates the intersection over union between index segmentation results and ground truth. Thus, it can be understood whether the proposed method shifts the lesion center or affects the image axes. The Jaccard index is calculated as in Eq. 8.

$$Sen = \frac{TP}{TP + FN} \qquad (4)$$

$$Spe = \frac{TN}{TN + FP} \qquad (5)$$

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \qquad (6)$$

$$Dice = \frac{2TP}{(2TP) + FP + FN} \qquad (7)$$

$$Jaccard = \frac{TP}{TP + FN + FP} \qquad (8)$$

where TP (true positive) represents correctly labeled lesion pixels, TN (true negative) represents correctly labeled non-lesion pixels, and FP (false positive) represents incorrectly labeled non-lesion pixels. In FP case, non-lesion pixels are labeled as lesions. FN (false negative) represents incorrectly labeled lesion pixels. In FN cases, lesion pixels are labeled as non-lesion.

Table 5 Performance of the training time and test time

| Methods	Training time per epoch	Test time per image
U-Net	486.8 s	8.1 s
SegNet	464.8 s	8.3 s
FCN	465.6 s	9.7 s
FrCN [1]	395.2 s	7.6 s
iFCN	432.3 s	8 s

The end-to-end iFCN architecture is trained for segmentation tasks with standard gradient descent with a 0.005 learning rate. The momentum of it is 0.9, and the L2 regularization factor is 0.005. The proposed model is trained with 35 epochs, and the batch size is 8. The number of iterations per epoch is 390, and the total number of iterations is 13,650. The main FCN architecture (backbone) at the center of the proposed method is a transfer learning-based FCN. In other words, it has been learned with low-level features (edges, Gabor, etc.) by training with a well-known dataset. Therefore, 35-epoch training with skin lesion images is quite sufficient. The learning rate is halved after every 10 epochs. Validation results are obtained every 400 iterations. The ISBI 2017 dataset training data is used for network training. All 2000 skin lesion images are used for training the iFCN architecture. The training lasts approximately 20 h. Only the specified epoch number is selected as the stop criterion.

The training and validation curves of the iFCN architecture are shown in Fig. 7. The blue curve at the top of Fig. 7 shows the accuracy during training. The black dots on this curve indicate validation accuracy. The red curve at the bottom of Fig. 7 shows the loss during training. The black dots on this curve indicate validation loss. As can be followed from the curve of Fig. 7, as the training process continues, the accuracy rate increases and the loss decreases accordingly. Preservation of the ratio between training and validation curves and the continuation of the increase is indicative that the network is not caught in the memorization problem. At the end of the training process, the training accuracy is 98.93% and the validation accuracy is 96.29%. At the end of the training, the training loss is 4.86% and the validation loss is 8.2%.

The test results of the iFCN architecture are calculated by the ISBI 2017 test dataset and the PH2 dataset images. Accordingly, 600 images in the ISBI 2017 datasets are used in the first test phase. Table 3 shows the test results for the ISBI 2017 dataset. Also, Table 3 lists the results of the most common algorithms for segmentation, the results of state-of-the-art methods in the literature, and the results of the iFCN architecture.

As shown in Table 3, the most successful segmentation results in the ISBI 2017 test data are obtained by the iFCN architecture. The proposed iFCN method performs better than

Fig. 8 The segmentation results of the proposed iFCN architecture. **a** Ground truth images. **b** iFCN results. **c** Plotting the proposed method and ground truth on the lesion
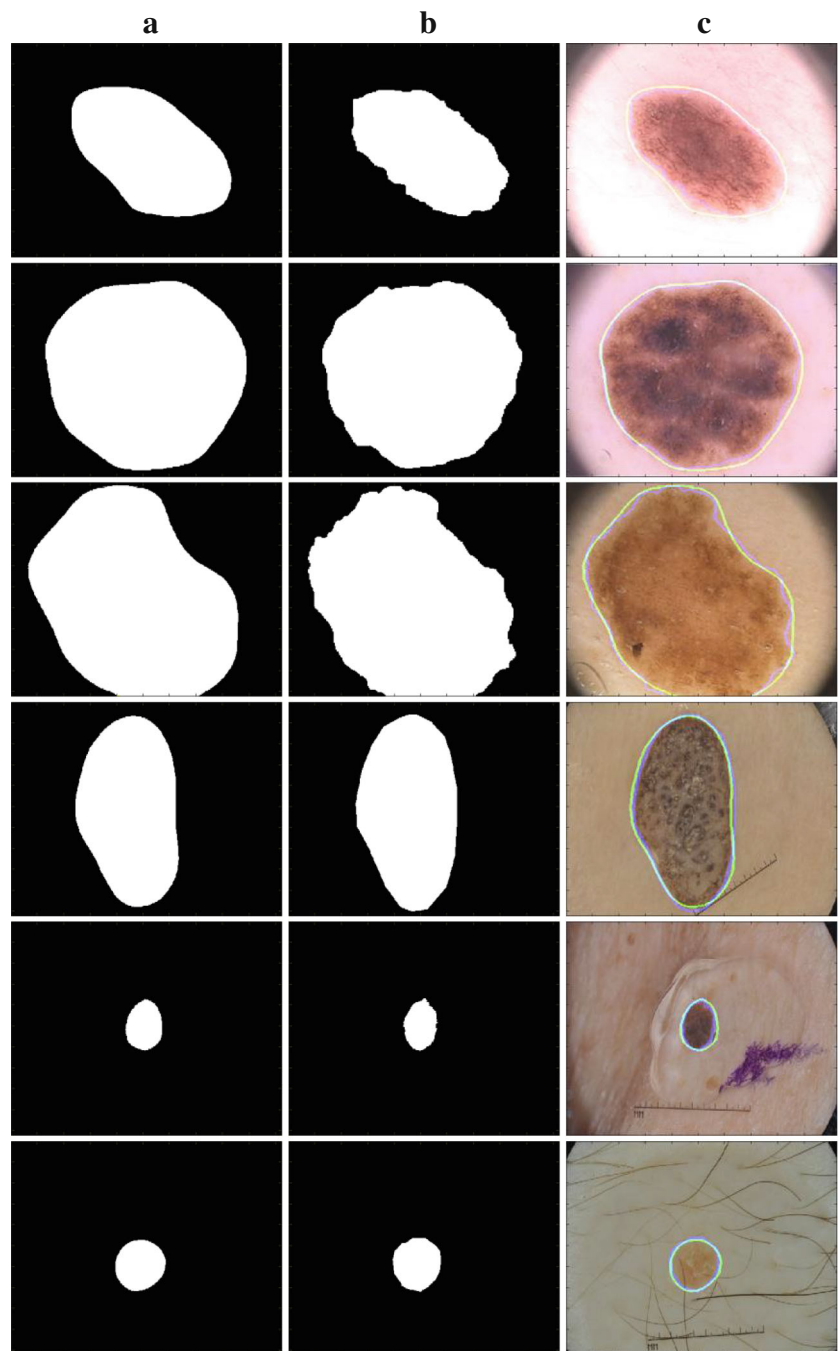
other techniques in the scope of sensitivity, specificity, accuracy, Dice, and Jaccard metrics. These metrics also show that the iFCN method matches a high resemblance to the background images. As shown in Table 3 for the ISBI 2017 test data, the obtained performance results with the proposed iFCN method are 85.44% sensitivity, 98.08% sensitivity, 95.30% accuracy, 88.64% Dice, and 78.34% Jaccard, respectively. The iFCN method has significant difference of 5.46% and 0.04% in sensitivity, 1.42% and 1.18% in specificity, 2.58% and 1.27% in accuracy, 4.81% and 1.56% in Dice,

6.17% and 1.23% in Jaccard than FCN and FrCN [1] respectively. In Table 3, the iFCN method obtained the highest performance in all metrics except for sensitivity. The iFCN method shows that the usability and consistency of the method is quite high in accordance with the obtained results.

Second experiments with the PH2 dataset test images are performed to prove the robustness of the iFCN architecture. The PH2 dataset test images are not used during the training and validation phase. The PH2 dataset test results are shown in Table 4. While considering the metric results of the proposed

iFCN method, it has provided an important performance advantage to the other studies in the literature. Although the FrCN [1] method is the most marginal study in the literature in terms of performance, the proposed iFCN method for the PH2 dataset test data obtained difference of 3.16% in sensitivity, 1.84% in accuracy, 1.25% in Dice, and 2.31% in the Jaccard index. According to the specificity of the FrCN [1] method, it showed 0.34% rate increase. Shortly, the iFCN method has 96.88% sensitivity, 95.31% specificity, 96.92% accuracy, 93.02% Dice, and 87.1% Jaccard metric values.

As can be seen from Tables 3 and 4, the proposed method produced more successful results in both datasets than other state-of-the-art methods. When the results produced together for two datasets are evaluated, it is seen that the proposed method is more robust than the other methods.

In addition, training and test time performances in the segmentation process are given in Table 5. When training times per epoch were analyzed, FrCN has the shortest time with 395.2 s. Then, the proposed iFCN method obtained second better performance with 432.3 s. There is a difference of 37.1 s per epoch as a training time difference. Test times per image is more valuable rather than training time per epoch. In the analysis of dermoscopic images, the test time should be less than 10 s. FrCN [1] achieved the highest performance with 7.6 s in terms of test time per image in Table 5. The proposed iFCN method has 8 s per image in test time. The difference between FrCN [1] and iFCN, 0.4 s, is not a big difference. Also, metric performance has a major priority. In summary, the proposed iFCN method has outperformed most studies in the literature due to metric performance in ISBI 2017 and PH2 datasets. When it was evaluated in terms of time, it showed remarkable performance.

Figure 8 illustrates the segmentation results of the proposed iFCN method. Besides, ground truth drawings are added to these images for comparison. Blue lines represent ground truth information, and green lines represent the results of the proposed method. In Fig. 8, the images in the first three lines were selected from the test images of the PH2 dataset, while the images in the last three lines were selected from the ISBI 2017 dataset test images. When all the images in Fig. 8 are examined, it is seen that the edges of the skin lesion are followed quite successfully by using the proposed method.

Conclusion

In this study, an effective segmentation method for skin lesion segmentation is presented. Since the FCN architecture produces results that are very close to the current studies in the literature, this architecture has been developed to create a robust CNN architecture. Unlike other state-of-the-art methods in the literature, the iFCN method includes specific solutions to skin lesion segmentation problems. For this purpose, a solution that can cope with undesirable residues (hair, ruler markers, etc.), indistinct boundaries, variable contrast, shape differences, and color differences in the lesion images is proposed. The ISBI 2017 and PH2 datasets, which are well known in the literature, are used to prove the robustness of our method. Thanks to the residual structure of the FCN architecture with proposed spatial information (iFCN), which creates a more advanced residual system and more precise detection of details on the edges of the lesion, and an analysis independent of the lightening conditions is performed. The results show that our method is more successful than the known methods in the literature, U-Net, classical FCN, and SegNet architectures. Also, it is more successful than the state-of-the-art methods proposed for skin lesion segmentation in the literature. The accuracy of the proposed method is approximately 1–2% higher than the most successful method in the literature. In future studies, it will be studied on a watershed network that is not affected by the color changes occurring at the edges of the lesion. The basis of decreasing the value depth towards the lesion edges will be taught to the CNN architecture.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Al-Masni MA, Al-Antari MA, Choi MT, Han SM, Kim TS: Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Comput Methods Programs Biomed 162: 221-231,2018
2. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2019. CA Cancer J Clin 69(1):7-34,2019
3. Siegel RL, Miller KD, Jemal A: Cancer statistics, 2017. CA Cancer J Clin 67(1):7-30,2017
4. Tsao H, Olazagasti JM, Cordoro KM, Brewer JD, Taylor SC, Bordeaux JS, Chren MM, Sober AJ, Tegeler C, Bhushan R, Begolka WS: Early detection of melanoma: reviewing the ABCDEs. J Am Acad Dermatol 72(4):717-23,2015
5. Li Y, Shen L: Skin lesion analysis towards melanoma detection using deep learning network. Sensors 18:2,2018
6. Oliveira RB, Papa JP, Pereira AS, Tavares JMRS: Computational methods for pigmented skin lesion classification in images: review and future trends Neural Comput Appl 29(3):613–636,2016
7. Pellacani G, Seidenari S: Comparison between morphological parameters in pigmented skin lesion images acquired by means of epiluminescence surface microscopy and polarized-light videomicroscopy Clin Dermatol 20(3):222–227,2002
8. Unver HM, Ayan E: Skin lesion segmentation in dermoscopic images with combination of YOLO and GrabCut algorithm Diagnostics (Basel) 9(3),2019
9. Pathan S, Prabhu KG, Siddalingaswamy PC: Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review. Biomed Sign Process Control 39:237–262,2018

10. Møllersen K, Kirchesch HM, Schopf TG, Godtliebsen F: Unsupervised segmentation for digital dermoscopic images: Skin Res Technol 16(4):401–407,2010

11. Gomez DD, Butakoff C, Ersboll BK, Stoecker W: Independent histogram pursuit for segmentation of skin lesions. IEEE Trans Biomed Eng, 55, 1, pp. 157-61, 2008.

12. Yuksel ME, Borlu M: Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic. IEEE Trans Fuzzy Syst 17(4):976–982,2009

13. Emre Celebi M, Wen Q, Hwang S, Iyatomi H, Schaefer G: Lesion border detection in dermoscopy images using ensembles of thresholding methods. Skin Res Technol 19(1)e252–e258,2013

14. Peruch F, Bogo F, Bonazza M, Cappelleri V-M, Peserico E: Simpler, faster, more accurate melanocytic lesion segmentation through MEDS. IEEE Trans Biomed Eng 61(2):557-565,2014

15. Zhou H, Schaefer G, Sadka AH, Celebi ME: Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images. IEEE J Select Top Sign Process 3(1):26–34,2009

16. Suer S, Kockara S, Mete M: An improved border detection in dermoscopy images for density based clustering. BMC Bioinformatics 12:10,2011

17. Schmid P: Segmentation of digitized dermatoscopic images by two-dimensional color clustering. IEEE Trans Med Imaging 18(2):164–171,1999

18. Kockara S, Mete M, Yip V, Lee B, Aydin K: A soft kinetic data structure for lesion border detection. Bioinformatics 26(12):i21–i28,2010

19. Xie F, Bovik AC: Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm: Pattern Recognit 46(3):1012–1019,2013

20. Abbas Q, Celebi ME, Fondón García I, Rashid M: Lesion border detection in dermoscopy images using dynamic programming. Skin Res Technol 17(1):91–100,2011

21. Abbas Q, Celebi ME, García IF: Skin tumor area extraction using an improved dynamic programming approach," Skin Res Technol 18(2):133–142,2012

22. Iyatomi H, Oka H, Celebi ME, Hashimoto M, Hagiwara M, Tanaka M, Ogawa K: An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm. Comput Med Imaging Graph 32(7):566–579,2008

23. Emre Celebi M, Kingravi HA, Iyatomi H, Alp Aslandogan Y, Stoecker WV, Moss RH, Malters JM, Grichnik JM, Marghoob AA, Rabinovitz HS, Menzies SW: Border detection in dermoscopy images using statistical region merging. Skin Res Technol 14(3):347–353,2008

24. Glaister J, Wong A, Clausi DA: Segmentation of skin lesions from digital images using joint statistical texture distinctiveness. IEEE Trans Biomed Eng 61(4):1220–1230,2014

25. Erkol B, Moss RH, Joe Stanley R, Stoecker WV, Hvatum E: Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes. Skin Res Technol 11(1):17–26,2005

26. Mete M, Sirakov NM: Lesion detection in demoscopy images with novel density-based and active contour approaches," BMC Bioinf 11:S6,2010

27. Ma Z, Tavares JMRS: A novel approach to segment skin lesions in dermoscopic images based on a deformable model. IEEE J Biomed Health Inform 20(2):615–623,2016

28. Wang H, Moss RH, Chen X, Stanley RJ, Stoecker WV, Celebi ME, Malters JM, Grichnik JM, Marghoob AA, Rabinovitz HS, Menzies SW, Szalapski TM: Modified watershed technique and post-processing for segmentation of skin lesions in dermoscopy images. Comput Med Imaging Graph 35(2):116–120,2011

29. Wighton P, Lee TK, Mori G, Lui H, McLean DI, Atkins MS: Conditional random fields and supervised learning in automated skin lesion diagnosis. Int J Biomed Imaging 2011:1–10,2011

30. Sadri AR, Zekri M, Sadri S, Gheissari N, Mokhtari M, Kolahdouzan F: Segmentation of dermoscopy images using wavelet networks IEEE Trans Biomed Eng 60(4):1134–1141,2013

31. Zahangir Alom M, Aspiras T, Taha TM, Asari VK: Skin cancer segmentation and classification with NABLA-N and inception recurrent residual convolutional networks. arXiv e-prints, https://ui.adsabs.harvard.edu/abs/2019arXiv190411126Z, [April 01, 2019]. 2019

32. Huang L, Zhao Y-g, Yang T-j: Skin lesion segmentation using object scale-oriented fully convolutional neural networks. SIViP 13(3):431–438,2019.

33. Bi L, Feng D, Fulham M, Kim J: Improving skin lesion segmentation via stacked adversarial learning. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp 1100–1103

34. Jiang F, Zhou F, Qin J, Wang T, Lei B: Decision-augmented generative adversarial network for skin lesion segmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp 447–450

35. Liu X, Hu G, Ma X, Kuang H: An enhanced neural network based on deep metric learning for skin lesion segmentation. In: 2019 Chinese Control And Decision Conference (CCDC), 2019, pp. 1633–1638.

36. Khan MA, Javed MY, Sharif M, Saba T, Rehman A: Multi-model deep neural network based features extraction and optimal selection approach for skin lesion classification. In: 2019 International Conference on Computer and Information Sciences (ICCIS), 2019, pp 1–7

37. Codella NCF, Gutman D, Emre Celebi M, Helba B, Marchetti MA, Dusza SW, Kalloo A, Liopyris K, Mishra N, Kittler H, Halpern A: 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 e-prints, https://ui.adsabs.harvard.edu/abs/2017arXiv171005006C, [October 01, 2017]. 2017

38. Mendonca T, Ferreira PM, Marques JS, Marcal ARS, Rozeira J: "PH^2" - A dermoscopic image database for research and benchmarking. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2013, pp 5437–5440

39. Shelhamer E, Long J, Darrell T: Fully Convolutional Networks for Semantic Segmentation," IEEE Trans Pattern Anal Mach Intell 39(4):640–651,2017.

40. Turchenko V, Chalmers E, Luczak A: A Deep Convolutional Auto-Encoder with Pooling - Unpooling Layers in Caffe. arXiv e-prints, https://ui.adsabs.harvard.edu/abs/2017arXiv170104949T, [January 01, 2017]. 2017

41. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL: DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Machine Intell 40(4)834–848,2018

42. Stefan Jianu SR, Ichim L, Popescu D: Automatic Diagnosis of Skin Cancer Using Neural Networks. In: 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE), 2019, pp 1–4

43. Garcia-Arroyo JL, Garcia-Zapirain B: Segmentation of skin lesions in dermoscopy images using fuzzy classification of pixels and histogram thresholding. Computer