



# Skin lesion segmentation using object scale-oriented fully convolutional neural networks

Lin Huang<sup>1,2</sup> · Yi-gong Zhao<sup>1</sup> · Tie-jun Yang<sup>2</sup>

Received: 11 August 2018 / Revised: 7 December 2018 / Accepted: 22 December 2018 / Published online: 28 January 2019  
© Springer-Verlag London Ltd., part of Springer Nature 2019

## Abstract

Melanoma is the deadliest form of skin cancer, and its incidence level is increasing. It is important to obtain a diagnosis at an early stage to increase the patient survival rate. Skin lesion segmentation is a difficult problem in medical image analysis. To address this problem, we propose end-to-end object scale-oriented fully convolutional networks (OSO-FCNs) for skin lesion segmentation. Given a single skin lesion image, the proposed method produces a pixel-level mask for skin lesion areas. We found that the scale of the lesions in the training dataset affects a large number of the segmentation results of the lesions in the testing phase, and thus, a training strategy called object scale-oriented (OSO) training is proposed. First, the pre-trained network of VGG-16 is adapted and is transformed into fully convolutional networks (FCNs). Second, after very simple preprocessing, skin lesion images with boundary-level annotations are fed into the FCNs for fine-tuning training based on the pre-trained model using OSO training. During the OSO training, the training dataset is divided into 2 subsets according to an index called the object occupation ratio, and then the whole training dataset and the 2 subsets are used to train 3 different scale-oriented FCNs. A dataset provided by the International Skin Imaging Collaboration (ISIC), ISIC2016, is used for training and testing. Our algorithm is compared with the state-of-the-art algorithms, and the experimental results demonstrate that the segmentation accuracy of our algorithm is higher or very close to the performances of the other algorithms.

**Keywords** Skin lesion · Melanoma · Fully convolutional neural networks · Object scale-oriented · Image segmentation

## 1 Introduction

Malignant melanoma is the deadliest form of skin cancer, and the number of invasive melanoma cases has recently been increasing rapidly [1]. Studies have shown that if melanomas are diagnosed before they spread to the surrounding lymph nodes or organs, the patient's 5-year survival rate is 98%; if they have spread to the lymph nodes, the survival rate is reduced to 63%; if they have spread to organs, the survival rate is only 16% [2]. Therefore, early diagnosis and treatment can be helpful in improving the cure rate of melanoma. Automated skin lesion segmentation is very important for the early diagnosis and treatment of melanoma.

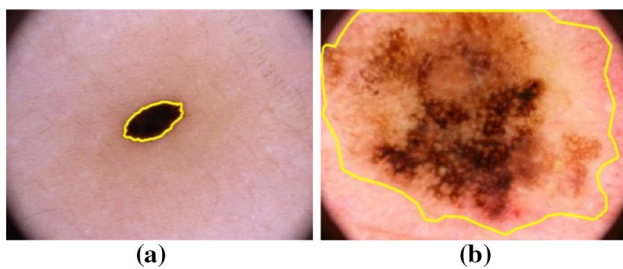
The main difficulties faced by the segmentation of pigmented skin lesion images include the following: low contrast between the skin lesion area and the surrounding skin; blurry boundary and irregular and non-fixed shapes; influence of artificial or other characteristics of the skin itself, such as hair and blood vessels; changing colors of the skin lesion area; and the possibility of being split into pieces by scars [3]. Figure 1 shows some dermoscopic images of pigmented skin lesions. Figure 1a shows a typical nevus; Fig. 1b presents an example of melanoma that has variegated coloring inside.

Commonly used segmentation techniques include image analysis methods, such as the threshold method [5–7], active contour model [8], clustering method [9, 10], histogram statistical method [11] and region-growing method [12–14], and supervised learning methods such as support vector machine (SVM) [15] and deep learning. These methods mainly use the object's own data (such as color and texture [16]) or learning from a labeled training dataset to perform the segmentation. For reviews of pigmented skin lesion segmentation methods, refer to [17].

✉ Yi-gong Zhao  
ygzhaoy@xidian.edu.cn

<sup>1</sup> Xidian University, Xi'an, Shanxi, China

<sup>2</sup> Guangxi Key Laboratory of Embedded Technology and Intelligent System, Guilin University of Technology, Guilin, Guangxi, China



**Fig. 1** Examples of dermoscopic images from the ISIC2016 [4] database: **a** a benign nevus; **b** a malignant melanoma with variegated coloring inside itself. In each figure, the yellow solid line indicates the ground truth of the segmentation of the lesion (color figure online)

In recent years, with the development of parallel processing capabilities using graphical processing units (GPUs) and the availability of large training datasets (such as ImageNet [18]), classification and recognition methods based on deep convolutional neural networks (CNNs) have attracted considerable attention. From the early LeNet [19] to the later AlexNet [20] and VGG-16 [21], for example, breakthroughs have been made in different fields. The main advantages of CNNs are that they can automatically learn effective features from a large number of training samples and have good scalability.

To explore end-to-end segmentation of skin lesions using CNNs, we perform the segmentation using FCNs [22]. There are no complicated preprocessing or post-processing steps in our method except for image size rescaling and format converting. First, we convert the VGG-16 model pre-trained on ImageNet Large-Scale Visual Recognition Challenge (ILSVRC2014) [18] into FCNs and train them for skin lesion segmentation by fine-tuning using the proposed object scale-oriented (OSO) training datasets provided by the International Skin Imaging Collaboration (ISIC), ISIC2016 [4]. Next, skips are added between the layers to fuse the hierarchically learned features. Finally, the public test dataset from ISIC2016 is evaluated using the performance indexes of Jaccard index, sensitivity, specificity, accuracy, precision, exclusive OR (XOR), and F measure. By comparing with the state-of-the-art methods, our method shows great simplicity, high accuracy, and efficiency.

The remainder of this paper is organized as follows: Sect. 2 discusses the FCN model. Then, the proposed method is described in detail in Sect. 3. Section 4 presents the experimental results and a discussion. Finally, Sect. 5 provides the conclusions.

## 2 Related work

The advent of CNN methods to medical image analysis has addressed this issue by learning hierarchical features of shape

and appearance based on annotated training sets. A review of the literature on CNNs is available in [23]. CNNs were first used for skin lesion classification by Codella et al. [24] and then Yoshida et al. [25]. Recently, CNNs have also been used in feature extraction [26] and segmentation [27]. Currently, CNN-based image segmentation methods can be roughly divided into two categories. For the first category, all of the methods are hybrid and use other methods with the CNNs, combining with full use of automatic extraction and classification features of the CNNs. The image segmentation is realized by combining other methods either before or after the CNNs. For the second category, all of the methods directly complete end-to-end image segmentation in the CNN networks.

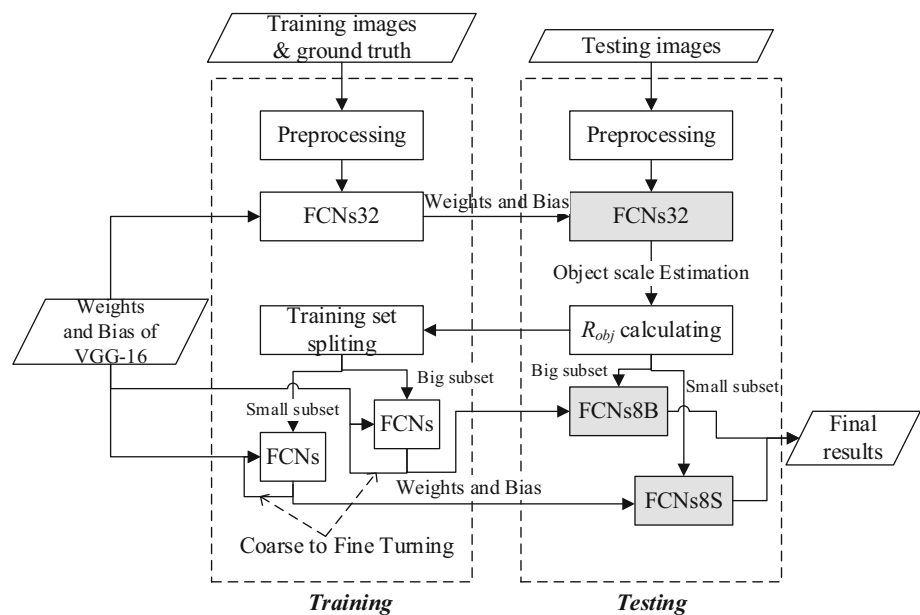
### 2.1 CNN-based combination models

In terms of CNN-based hybrid methods, Chen et al. [28] first extracted high-resolution depth features based on CNNs and then combined these features with conditional random fields (CRF) to perform semantic segmentation. Because CNNs have poor performance in precise object segmentation, patch-like methods are introduced in [27] for the purpose of lesion segmentation. Jafari et al. [27] prepared preprocessed patches of input images and fed them into a CNN for training and testing. After the segmentation probability map was obtained in the testing stage of the CNN, the post-processing steps, including largest connecting component selection and dilation, were performed. The main problems of this type of method are the extra cost for data exchange due to the separation between the CNN and the other image segmentation methods and the quality of the proposed region, which is basically determined by the later segmentation and classification results.

### 2.2 End-to-end models

End-to-end segmentation methods directly complete the image segmentation in the CNN network. To integrate the ability of object segmentation into CNNs, Long et al. [22] proposed fully convolutional networks (FCNs) and applied them to spatially dense prediction tasks (semantic segmentation). Contemporary networks are adapted into FCNs and have the ability to perform object segmentation in CNNs. By combining different hierarchical features, FCNs are capable of semantic segmentation by using the approach of end-to-end training and testing. Unlike traditional CNNs, this method discards the fully connected layer used for classification and uses a backward convolution (sometimes called deconvolution) layer [22] to restore the position information of the object. However, due to frequent downsampling in the forward transfer process, the position information is severely lost, and the position is still very blurred even after recovery.

**Fig. 2** Block diagram of the proposed method, which consists of three major processes: preprocessing, training, and testing based on FCNs



Based on an FCN, Yuan et al. [29] proposed a pigmented skin lesion segmentation method that uses an improved loss function. However, after obtaining the output result from the FCN, this method must still perform manual threshold segmentation, and thus, it is not strictly an end-to-end approach.

Currently, CNNs have made great progress in the area of pattern recognition. However, the deeper the CNN layers are, the larger the feature size that they extract, which runs in the opposite direction to the pixel accuracy required by segmentation. In addition, the performance of CNNs is highly dependent on having a large dataset, especially a labeled oversized training set.

### 3 Method

In this section, the proposed method is described in detail. The proposed method, as shown in the block diagram in Fig. 2, consists of three major processes: preprocessing, training, and testing based on FCNs. Prepared images are fed into the second stage, which is the training process of the FCNs. The details of these stages are explained in the following.

#### 3.1 Object scale of FCNs

CNNs are able to achieve high-accuracy object recognition based on automatically extracted multi-level features. However, the deeper the network, the larger the field of view, and the more details are lost. For pigmented skin lesions without fixed geometry, if a CNN can learn the size scale, the accuracy of segmentation will be improved. To verify this problem, we conducted the following experiment.

**Table 1** Mean object occupation ratio for training and testing

Training set	Train	Test
All	0.27	0.28
Train_big	0.48	0.32
Train_small	0.13	0.24

First, we define the object occupation ratio as follows:

$$R_{obj} = N_{obj}/N, \quad (1)$$

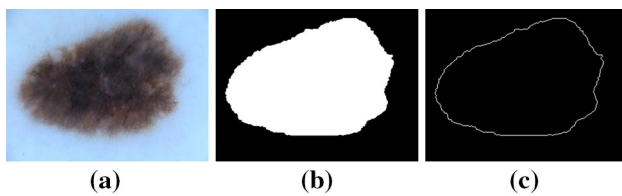
where  $N_{obj}$  is the number of pixels that belong to a lesion, and  $N$  is the total number of all of the pixels in a given image.

For the ISIC2016 training set, its mean object occupation ratio was 0.27 as calculated by the following equation:

$$\bar{R}_{obj} = \sum_v R_{obj}^i / K,$$

where  $R_{obj}^i$  is the object occupation ratio  $R_{obj}$  of the  $i$ th sample in a given dataset and  $K$  is the total number of samples. Based on this mean object occupation ratio  $\bar{R}_{obj}$ , we divided the training set into two subsets: Train\_big and Train\_small. All of the samples in Train\_big satisfied  $R_{obj} \geq 0.27$ , and the  $R_{obj}$  of all of the samples in Train\_small was  $< 0.27$ . Then, we used the entire training set, Train\_big, and Train\_small, to perform the training with FCN32s [22]. After the training, the ISIC2016 test set was used for testing. The mean object occupation ratio was calculated according to the segmentation results, as given in Table 1.

As given in Table 1, when the  $\bar{R}_{obj}$  of the training set is relatively large (the  $\bar{R}_{obj}$  of Train\_big = 0.48), the  $\bar{R}_{obj}$  of the segmentation results obtained from the test set is also relatively large (0.32). When the  $\bar{R}_{obj}$  of the training set is



**Fig. 3** Examples of preprocessing. **a** An original image after resizing. The sizes of the image before and after rescaling were 1022\*767 and 342\*256, respectively. **b** The segmentation ground truth of (a). **c** The ground truth after format conversion of (b)

relatively small (the  $\bar{R}_{obj}$  of Train\_small = 0.13), the  $\bar{R}_{obj}$  of the segmentation results also becomes smaller (0.24). When the  $\bar{R}_{obj}$  of the training set = 0.27, the  $\bar{R}_{obj}$  of the segmentation results = 0.28. Obviously, during the training process of the segmentation of the pigmented skin lesions, FCNs automatically learn the object scale of the training set.

Based on the above experimental results, we used the proposed OSO training method to improve the segmentation performance during the training phase. A specific description is given in Sect. 3.4.

### 3.2 Preprocessing

In the preprocessing stage, only image resizing and reformatting are needed. As shown in Fig. 2, transfer learning is performed based on VGG-16 pre-trained on ILSVRC2014. First, to be consistent with the size of the training samples used by VGG-16, the height of the original image is scaled to 256 pixels, and the width is adaptively scaled while maintaining the aspect ratio of the original image. An example of image resizing is demonstrated in Fig. 3a.

In addition, the images of the segmentation ground truth must be reformatted as follows: The values of the background pixels are set to be 0 s, those of the lesion pixels are set to be 1 s, and the pixels of the edges that separate the background from the lesions are set to be 255 s. The pixels of the edges are ignored during the training because it is too complicated to discriminate between the normal skin and lesions in the areas of the edges [22]. An example of the ground truth image resizing and reformatting is depicted in Fig. 3b, c.

### 3.3 Model architecture

The network architecture is depicted in Fig. 4; the networks comprise six stacks of convolution layers, denoted by  $C_{mn}$ , and a stack of deconvolution layers [22],  $DC_x$ . For the lower five stacks of convolution layers, each convolution layer has a small size (3 by 3) for the convolution filters and is followed by a rectified linear function (relu) layer. Between each stack of convolution layers, there is a max pooling layer  $P_i$ . After the fifth pooling layer  $P_5$ , the three fully connected layers

of VGG-16 are transformed into three convolutional layers, where the last 1000-way classification layer from VGG-16 is replaced with a 2-way convolutional layer for the task of skin lesion segmentation. The final stack consists of one or more deconvolution layers.

To improve the ability of the pixel-level prediction, we adopt the strategy in [22], which combines the final convolutional layer ( $C_{63}$ ) with lower pooling layers ( $P_4$  and  $P_3$ ). Because the proposed FCNs consist of five max pooling layers that lead to a  $32 \times$  reduction in the output resolution compared to the original image, we construct three FCNs, from coarse to fine, of FCNs32, FCNs16, and FCNs8, for the skin lesion segmentation.

### 3.4 Object scale-oriented (OSO) training

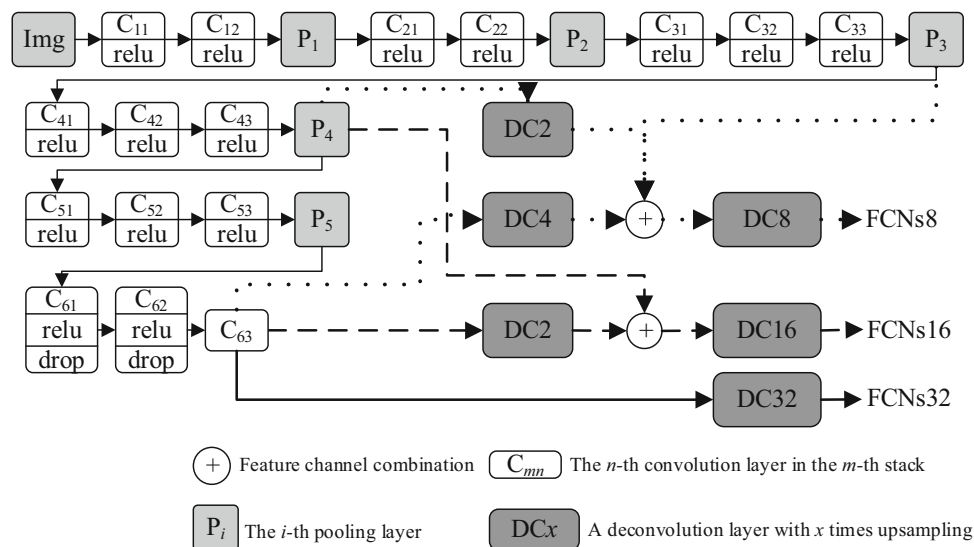
During the training phase, we proposed an OSO training method. The OSO training process is specifically described as follows:

1. Similar to the traditional training method, we first used all of the training set samples for training; the network structure used was FCNs32, and the trained network was called FCNs32A.
2. Then, the FCNs32A was used to perform the first round of tests on the test set: scale estimation test. Through the scale estimation test, the mean object occupation ratio  $\bar{R}_{obj}$  of the test set could be calculated.
3. In turn, based on  $\bar{R}_{obj}$ , the training set was divided into two subsets, Train\_big and Train\_small. Train\_big contained samples with values of  $\bar{R}_{obj}$  greater than or equal to  $\bar{R}_{obj}$ , while Train\_small contained samples with values of  $\bar{R}_{obj}$  less than  $\bar{R}_{obj}$ .
4. The two subsets, Train\_big and Train\_small, were used to train the network (the same as the one in 1), and the training method was similar to that in Ref. [22]. First, FCNs32, FCNs16, and FCNs8 networks were used to perform training from coarse to fine, respectively, and finally, two trained networks, FCNs8B and FCNs8S, were trained. The training process was then complete.

Through this training method, we obtained three trained FCN networks, where FCNs32A was used to estimate the object scale, and FCNs8B and FCNs8S were used for large-scale and small-scale object image segmentation (Fig. 2). During training, the loss is the sum of the cross-entropy terms over each pixel in the output layer.

### 3.5 Testing

During the testing phase, we adopted the proposed “estimate-test” two-step test method (shown in Fig. 2). In the first step, FCNs32A was used to estimate the scale of a tested sample,



**Fig. 4** Architecture of the proposed FCNs based on VGG-16. The networks comprise six stacks of convolution layers, denoted by  $C_{mn}$ , and a stack of deconvolution layers,  $DCx$ . For the lower five stacks of convolution layers, each convolution layer has a small size (3 by 3) for the convolution filters and is followed by a rectified linear function (relu) layer. Between each stack of convolution layers, there is a max pooling layer  $P_i$ . After the fifth pooling layer  $P_5$ , the three fully connected

layers of VGG-16 are transformed into three convolution layers, where the last 1000-way classification layer from VGG-16 is replaced with a 2-way convolutional layer for the task of skin lesion segmentation. The final stack consists of one or more deconvolution layers. We adopt the strategy in [22] that combines the final convolution layer ( $C_{63}$ ) with the lower pooling layers ( $P_4$  and  $P_3$ ) to construct three FCNs, from coarse to fine: FCNs32, FCNs16, and FCNs8

in other words, to calculate the  $R_{obj}$  of a single test sample according to the segmentation result. Then, based on the size relationship between  $R_{obj}$  and  $\bar{R}_{obj}$ , FCNs8B (if  $R_{obj} \geq \bar{R}_{obj}$ ) or FCNs8S (if  $R_{obj} < \bar{R}_{obj}$ ) was further selected as a test classifier, and the final segmentation results were obtained.

## 4 Experiments

Our experiments were performed on a PC with a 2.5 GHz Intel Core2 Q8300 processor and 8 GB of 800 MHz DDR2 RAM using Caffe [30] on a single NVIDIA GeForce GTX Titan X. We used a public database of skin lesions from ISIC2016 for segmentation accuracy comparisons with other state-of-the-art methods.

### 4.1 Dataset

A public skin lesion database was used to train and test the segmentation algorithms. The dataset was released by the International Skin Imaging Collaboration (ISIC) International Symposium on Biomedical Imaging (ISBI) challenges in 2016 and is called ISIC2016. ISIC2016 is divided into two subsets: The training subset contains 900 dermoscopic images, and the testing set contains 379 images. Each image has an associated ground truth.

### 4.2 Accuracy evaluation

The segmentation results were compared with a manually segmented ground truth, as determined manually by experienced dermatologists. The algorithms were compared numerically by calculating the sensitivity (recall), specificity, accuracy, precision, Jaccard index ( $J$ ),  $F$  measure (also called dice coefficient), and XOR measure, with respect to the correct classification of each pixel as normal skin or a lesion. The formulas of these metrics are given in Eqs. (2)–(8), respectively.

$$SE = TP / (TP + FN), \quad (2)$$

$$SP = TN / (FP + TN), \quad (3)$$

$$ACC = (TP + TN) / (TP + FP + TN + FN), \quad (4)$$

$$PR = TP / (TP + FP), \quad (5)$$

$$J = TP / (TP + FP + FN), \quad (6)$$

$$F = 2(PR \cdot SE) / (PR + SE), \quad (7)$$

$$F = 2(PR \cdot SE) / (PR + SE), \quad (8)$$



where TP is the number of true positive pixels, FP the number of false positives, TN is the number of true negatives, and FN is the number of false negatives.

### 4.3 Training

For the ISIC2016 dataset, we used the proposed OSO method for training. For the three networks that need training (FCNs32, FCNs16, and FCNs8), the same parameters were set: The networks were optimized with a stochastic gradient with a mini-batch size of 1 image, with a dropout of 0.5 at all dropout layers, and an initial momentum of 0.99 with a weight decay of 0.0005. A mini-batch size of 1 with high momentum allows more accurate models to be trained in less time [22]. Different values were used for the learning rate. The learning rates of FCNs32, FCNs16, and FCNs8 were  $1e-10$ ,  $1e-12$ , and  $1e-14$ , respectively, and all had a fixed learning rate decay policy. We trained a network for a total of 100,000 iterations; the total training time for one network was approximately 8 h.

### 4.4 Testing and results

In the testing stage, the testing set from ISIC2016 was used for a segmentation performance comparison. We compared several state-of-the-art methods: the top three performers in the ISBI 2016 Challenge (EXB, CUMED, and Mahmudur [31]), JD-FCNs [29], MS-FCNs [32], and the FCN model from the present study trained on all training images, called ALL-FCNs.

Table 2 presents a summary of the segmentation experiments evaluated on the testing set of ISIC2016. The proposed OSO-FCNs yield the highest accuracy (95.8%) and sensitivity (92.5%). In addition, OSO-FCNs produced a Jaccard index of 0.845, which is very close to that of JD-FCNs. However, from the perspective of modeling, the segmentation model of JD-FCNs is more complex. After obtaining the FCNs output, it is necessary to perform two rounds of manual threshold segmentation and to integrate the output of 6 FCNs to obtain the best results.

The OSO-FCN proposed by this paper is a fully automatic end-to-end method that requires no post-processing, and it is simpler and more efficient. JD-FCNs and MS-FCNs have the best  $F$  measure value of 0.912, and OSO-FCN is nearly the same with a value of 0.911. ALL-FCNs perform worse than OSO-FCNs on all indices.

Figure 5 shows the segmentation results of 6 representative test images. According to the relationship between  $R_{obj}$  (the object scale estimated by FCNs32A) and  $\bar{R}_{obj}$ , the six images can be divided into 3 cases:  $R_{obj} \geq \bar{R}_{obj}$ , including (a) and (b);  $R_{obj} < \bar{R}_{obj}$ , including (c) and (d); and  $R_{obj} \approx \bar{R}_{obj}$ , including (e) and (f). Among them,  $\bar{R}_{obj} = 0.284$ , and the  $R_{obj}$  of Fig. 5a–f are 0.569, 0.813, 0.123, 0.011, 0.295, and

**Table 2** Results for the testing dataset of ISIC2016 comparing our model with state-of-the-art methods (%)

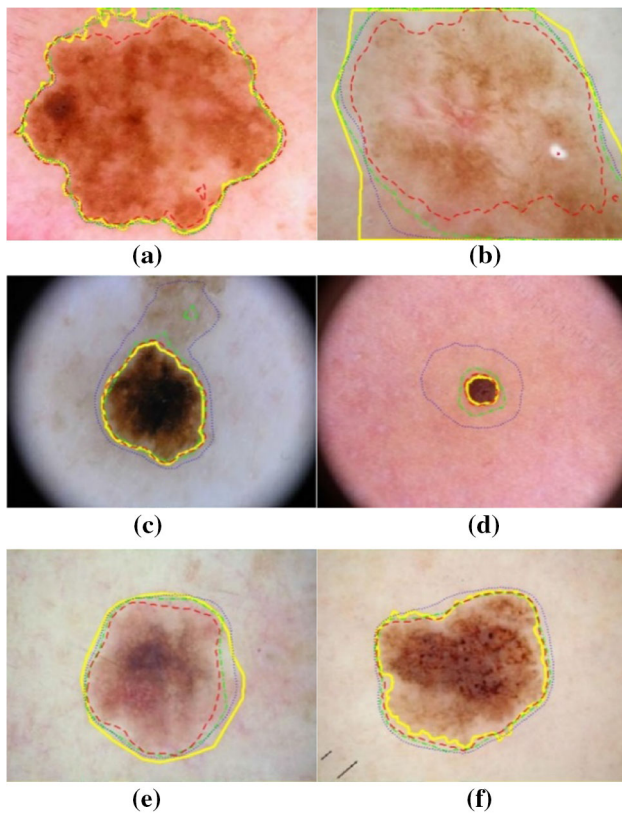
Algorithm	SE	SP	ACC	XOR	PR	$J$	$F$
EXB	91.0	96.5	95.3	N/A	N/A	84.3	91.0
CUMED	91.1	95.7	94.9	N/A	N/A	82.9	89.7
Mahmudur	88.0	<b>96.9</b>	95.2	N/A	N/A	82.2	89.5
JD-FCNs	91.8	96.6	95.5	N/A	N/A	<b>84.7</b>	<b>91.2</b>
MS-FCNs	92.2	96.5	95.5	N/A	N/A	84.6	<b>91.2</b>
ALL-FCNs	92.4	96.0	95.2	22.0	89.9	83.0	89.9
OSO-FCNs	<b>92.5</b>	96.4	<b>95.8</b>	<b>18.2</b>	91.3	84.5	91.1

Bold values indicate the best indices

0.279, respectively. When the estimated object scale satisfies  $R_{obj} \geq \bar{R}_{obj}$ , FCNs8B will be selected to complete the final segmentation (Fig. 5a, b). At this time, compared with the FCNs8S segmentation result, the segmentation result of FCNs8B is closer to the ground truth (indicated by the blue dotted line), while the segmentation result of FCNs8S is more compact, concentrating on the “protruding” part of the object (indicated by the red dash line). When the estimated object scale satisfies  $R_{obj} < \bar{R}_{obj}$ , FCNs8S is selected to complete the final segmentation (Fig. 5c, d). At this time, the difference between the segmentation results of FCNs8S and FCNs8B is even more distinct. The result of FCNs8S is basically consistent with the ground truth, while the result of FCNs8B contains a much larger area than the object. We also compared the  $R_{obj} \approx \bar{R}_{obj}$  case.  $R_{obj} = 0.295$  in Fig. 5e, which is slightly larger than  $\bar{R}_{obj}$ , and the result of FCNs8B is closer to the ground truth than that of FCNs8S.  $R_{obj} = 0.279$  in Fig. 5f, which is slightly smaller than  $\bar{R}_{obj}$ , and the result of FCNs8S is more accurate than that of FCNs8B.

Figure 6 shows a comparison of the performance curves of the two proposed OSO classification networks FCNs8B, FCNs8S, and ALL-FCNs. The histogram in the figure presents the number of images equally divided intervals within  $[0, 1]$ . When  $0 < R_{obj} \leq 0.3$ , the average  $J$  value of the FCNs8S and ALL-FCNs segmentation is greater than 80%, while the average  $J$  value of FCNs8B is less than 70%. When  $0.3 < R_{obj} \leq 1$ , the average  $J$  value of FCNs8B increases significantly and remains above 80%, while the average  $J$  value of FCNs8S and ALL-FCNs keeps decreasing with the increase in  $R_{obj}$ , and the minimal value is close to 60%. Notably, the change in the segmentation performance of FCNs8S and FCNs8B occurs only within  $0.2 < R_{obj} < 0.3$ , which is exactly the interval that  $\bar{R}_{obj} = 0.284$  is in. Therefore, the OSO-FCNs proposed by this paper adaptively selected the higher-performance classification network based on the object scale to perform segmentation and achieved satisfactory results.

In terms of the segmentation calculation time, the average time cost for each test sample is approximately 100 ms; in

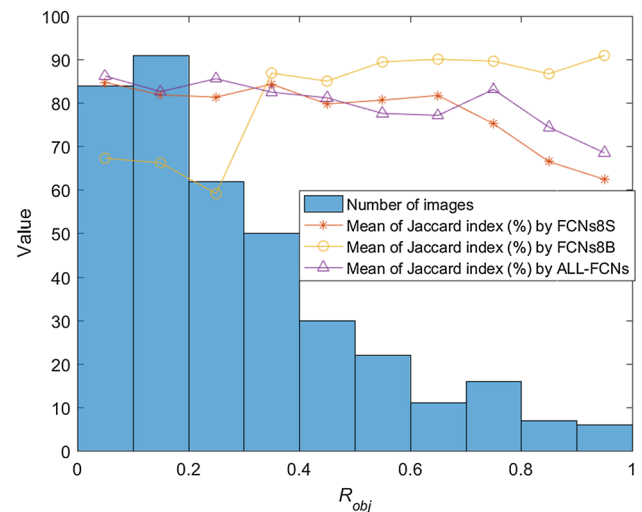


**Fig. 5** Segmentation results for six examples in the ISBI 2016 testing dataset, where **a, b** are examples with  $R_{obj} \geq \bar{R}_{obj}$ , **c, d** are examples with  $R_{obj} < \bar{R}_{obj}$ , and **e, f** are examples with  $R_{obj} \approx \bar{R}_{obj}$ . In each figure, the yellow solid line indicates the ground truth as segmented manually by dermatologist, the green dash-dotted line is the result of the estimated segmentation by FCNs32A, the blue dotted line is the result of the final segmentation by FCNs8B, while the red dash line is the result of the final segmentation by FCNs8S (color figure online)

other words, a segmentation speed of 10 samples/second can be achieved. This finding is mainly due to the modeling of a simpler, end-to-end segmentation model that does not require additional post-processing.

## 5 Conclusions

In this paper, we proposed an OSO fully convolutional neural network segmentation method, which was applied to the segmentation of pigmented skin lesions and achieved satisfactory segmentation results. Compared with the existing advanced methods, the characteristics of this method are automatic segmentation and simple preprocessing, the segmentation model is constructed using only one set of FCNs, and there is no post-processing. First, the coarse classification network that had been trained by the entire training set was used to estimate the scale of the object, which improved the accuracy of the estimation of the scale of the test object.



**Fig. 6** Comparison of the FCNs8B, FCNs8S, and ALL-FCNs performance curves. The yellow solid line indicates the mean Jaccard index produced by FCNs8B, the red solid line is the mean Jaccard index produced by FCNs8S, the purple solid line is the mean Jaccard index produced by ALL-FCNs, and the histogram indicates that for the number of images in a certain interval that corresponds to the object scale  $R_{obj}$  estimated by FCNs32, there is a total of 10 equally divided intervals within [0 1] (color figure online)

Second, based on the estimated object scale, the fine classification networks that had been trained by the training samples with a certain range of scale were selected to perform the final segmentation, which improved the accuracy of the segmentation of the object. On the ISIC2016 database, we compared its performance with the current advanced methods, and our method was simple and achieved satisfactory segmentation results.

**Acknowledgements** This research was partly supported by the Guangxi Natural Science Foundation (2018JJB170004), the Guangxi Basic Ability Promotion Project for Young and Middle-aged Teachers (2017KY0247), the Project of Cultivating a Thousand Young and Middle-aged Teachers in Guangxi Universities, the Guangxi Key Laboratory Fund of Embedded Technology and Intelligent System (2018A-07), and the Guangxi Universities Key Laboratory Fund of Embedded Technology and Intelligent Information Processing (2017-1-1, 2017-2-4). Additionally, we would like to thank NVIDIA for providing the Titan X GPU used in this research.

## References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2018. *CA Cancer J. Clin.* **68**(1), 7–30 (2018)
2. American Cancer Society What are the key statistics about melanoma skin cancer? (2015). <http://www.cancer.org/cancer/skincancer-melanoma/detailedguide/melanoma-skin-cancer-key-statistics>. Accessed 16 Aug 2015
3. Celebi, M.E., Iyatomi, H., Schaefer, G., Stoecker, W.V.: Lesion border detection in dermoscopy images. *Comput. Med. Imaging Graph.* **33**(2), 148–153 (2009)

4. Gutman, D., Codella, N.C.F., Celebi, E., Helba, B., Marchetti, M., Mishra, N., Halpern, A.: 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). In: IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (2016). [arXiv:1605.01397](https://arxiv.org/abs/1605.01397) [cs.CV]
5. Garnavi, R., Aldeen, M., Celebi, M.E., Varigos, G., Finch, S.: Border detection in dermoscopy images using hybrid thresholding on optimized color channels. *Comput. Med. Imaging Graph.* **35**(2), 105–115 (2011)
6. Celebi, E.M., Quan, W., Sae, H., Hitoshi, I., Gerald, S.: Lesion border detection in dermoscopy images using ensembles of thresholding methods. *Skin Res. Technol.* **19**(1), e252–e258 (2013)
7. 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 (2013)
8. Ma, Z., Tavares, J.M.R.S.: 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)
9. Zhou, H., Schaefer, G., Sadka, A.H., Celebi, M.E.: Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images. *IEEE J. Sel. Top. Signal Process.* **3**(1), 26–34 (2009)
10. Ashour, A.S., Hawas, A.R., Guo, Y., Wahba, M.A.: A novel optimized neutrosophic k-means using genetic algorithm for skin lesion detection in dermoscopy images. *Signal Image Video Process.* **12**(7), 1311–1318 (2018)
11. Gomez, D.D., Butakoff, C., Ersboll, B.K., Stoecker, W.: Independent histogram pursuit for segmentation of skin lesions. *IEEE Trans. Biomed. Eng.* **55**(1), 157–161 (2008)
12. Iyatomi, H., Oka, H., Celebi, M.E., 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)
13. Celebi, M., Kingravi, H.H., Aslandogan, Y., Stoecker, W., Moss, R., Malters, J., Grichnik, J., Marghoob, A., Rabinovitz, H., Menzies, S.: Border detection in dermoscopy images using statistical region merging. *Skin Res. Technol.* **14**(3), 347–353 (2008)
14. Celebi, M.E., Asl, Y.A., Stoecker, W.V., Iyatomi, H., Oka, H., Chen, X.: Skin research and technology unsupervised border detection in dermoscopy images. *Skin Res. Technol.* **13**(4), 377–384 (2007)
15. Arakeri, M.P., Reddy, G.R.M.: Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images. *Signal Image Video Process.* **9**(2), 409–425 (2015)
16. An, N.-Y., Pun, C.-M.: Color image segmentation using adaptive color quantization and multiresolution texture characterization. *Signal Image Video Process.* **8**(5), 943–954 (2014)
17. Celebi, M.E., Mendonca, T., Marques, J.S. (eds.): A state-of-the-art survey on lesion border detection in dermoscopy images. In: *Dermoscopy image analysis*, pp. 97–129. CRC Press, Boca Raton, FL (2015). <https://www.taylorfrancis.com/books/9781482253269>
18. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.: ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**(3), 211–252 (2015)
19. LéCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
20. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. *Adv. Neural. Inf. Process. Syst.* **25**(2), 1106–1114 (2012)
21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014). [arXiv.org/abs/1409.1556](https://arxiv.org/abs/1409.1556)
22. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **79**(10), 1337–1342 (2015)
23. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
24. Codella, N., Cai, J., Abedini, M., Garnavi, R., Halpern, A., Smith, J.R.: Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In: *International Workshop on Machine Learning in Medical Imaging*, pp. 118–126 (2015)
25. Yoshida, T., Celebi, M.E., Schaefer, G., Iyatomi, H.: Simple and effective pre-processing for automated melanoma discrimination based on cytological findings. In: *2016 IEEE International Conference on Big Data (Big Data)*, pp. 3439–3442. (2016)
26. Barata, A.C.F., Celebi, E.M., Marques, J.: A survey of feature extraction in dermoscopy image analysis of skin cancer. *IEEE J. Biomed. Health Inform.* (2018). <https://doi.org/10.1109/jbhi.2018.2845939>
27. Jafari, M.H., Nasresfahani, E., Karimi, N., Soroushmehr, S.M.R., Samavi, S., Najarian, K.: Extraction of skin lesions from non-dermoscopic images using deep learning (2016). [arXiv.org/abs/1609.02374](https://arxiv.org/abs/1609.02374)
28. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2018)
29. Yuan, Y., Ming, C., Lo, Y.C.: Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Trans. Med. Imaging* **36**(9), 1876–1886 (2017)
30. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J.: Caffe: convolutional architecture for fast feature embedding. In: *Proceedings of the 22nd ACM International Conference on Multimedia*, pp. 675–678. Orlando, FL (2014)
31. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans. Med. Imaging* **36**(4), 994–1004 (2017)
32. Bi, L., Kim, J., Ahn, E., Kumar, A., Fulham, M., Feng, D.: Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Trans. Biomed. Eng.* **64**(9), 2065–2074 (2017)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.