

RESEARCH

Open Access



Automatic lesion segmentation using atrous convolutional deep neural networks in dermoscopic skin cancer images

Ranpreet Kaur^{1*}, Hamid GholamHosseini¹, Roopak Sinha¹ and Maria Lindén²

Abstract

Background: Melanoma is the most dangerous and aggressive form among skin cancers, exhibiting a high mortality rate worldwide. Biopsy and histopathological analysis are standard procedures for skin cancer detection and prevention in clinical settings. A significant step in the diagnosis process is the deep understanding of the patterns, size, color, and structure of lesions based on images obtained through dermatoscopes for the infected area. However, the manual segmentation of the lesion region is time-consuming because the lesion evolves and changes its shape over time, making its prediction challenging. Moreover, it is challenging to predict melanoma at the initial stage as it closely resembles other skin cancer types that are not malignant as melanoma; thus, automatic segmentation techniques are required to design a computer-aided system for accurate and timely detection.

Methods: As deep learning approaches have gained significant attention in recent years due to their remarkable performance, therefore, in this work, we proposed a novel design of a convolutional neural network (CNN) framework based on atrous convolutions for automatic lesion segmentation. This architecture is built based on the concept of atrous/dilated convolutions which are effective for semantic segmentation. A deep neural network is designed from scratch employing several building blocks consisting of convolutional, batch normalization, leakyReLU layer, and fine-tuned hyperparameters contributing altogether towards higher performance.

Conclusion: The network was tested on three benchmark datasets provided by International Skin Imaging Collaboration (ISIC), i.e., ISIC 2016, ISIC 2017, and ISIC 2018. The experimental results showed that the proposed network achieved an average Jaccard index of 90.4% on ISIC 2016, 81.8% on ISIC 2017, and 89.1% on ISIC 2018 datasets, respectively which is recorded as higher than the top three winners of the ISIC challenge and other state-of-the-art methods. Also, the model successfully extracts lesions from the whole image in one pass in less time, requiring no pre-processing step. The conclusions yielded that network is accurate in performing lesion segmentation on adopted datasets.

Keywords: Skin cancer, Lesion segmentation, CNN, Deep learning

Introduction

Skin cancer is caused by the growth of cancerous cells that proliferate in an abnormal and uncontrolled manner in the topmost layer of the skin called the epidermis. The primary reason for the occurrence of skin cancer is direct exposure to ultraviolet sun rays for longer hours producing a pigment known as melanin in the upper skin layer [1]. Moreover, fair complexion, sunburn, genetic

*Correspondence: ranpreet.kaur@aut.ac.nz

¹ School of Engineering, Computer, and Mathematical Sciences, Auckland University of Technology, 55 Wellesley street, 1010 Auckland, New Zealand
Full list of author information is available at the end of the article



history, and weak immune system are other risk factors that contribute to the formation of skin cancer [2]. There are different types of skin cancer, such as squamous cell carcinoma, basal cell carcinoma, and melanoma [3], where melanoma is the most aggressive form of cancer comparatively. According to the statistics reported by the American skin cancer society [4], melanoma is the 19th most commonly found problem worldwide, and 100,350 new cases of melanoma were anticipated in the USA, 16221 in Australia, and 2500 in New Zealand in 2020 [5]. In terms of pricing, it has been estimated that the cost of treating skin cancer is 3.3 billion per year [6, 7]; thus, it is the most expensive procedure for the health systems.

Moreover, melanoma has become a critical public health concern for clinicians and researchers who emphasize reducing the mortality rate with early diagnosis. Detecting melanoma early can increase the survival rate. In clinical settings, trained specialists such as dermatologists commonly diagnose melanoma from dermoscopic lesion images based on Asymmetry, Border, Color, Diameter, and Evolution (ABCDE) [8] characteristics which is a very time-consuming process. Other methods used by dermatologists for performing visual examination are biopsy and histopathological analysis. The major problem in these traditional diagnostic procedures is time, high cost, and variation inaccuracy. Thus, computer-aided design (CAD) systems are widely adopted for the timely detection of melanoma, where cancer image segmentation is the most crucial process in CAD for the detailed analysis of lesion structure. The heterogeneous appearance of the lesion area in terms of color, size, shape, location, and texture makes the segmentation task very challenging,

as shown in image samples of Fig. 1. Therefore, there is a need for an automatic segmentation approach to assist dermatologists in understanding the nature and pattern of the lesion area. Also, this method is significant for automatically generating the ground truth images which were previously annotated manually by dermatologists.

Recent advances in artificial intelligence, particularly deep learning, offered numerous automatic and accurate identification advantages. Therefore, for extracting accurate lesion patterns in skin cancer images, a novel deep neural framework is proposed that proved to be a suitable segmentation approach based on the obtained results for the given datasets. We also embedded the atrous convolutions [9] in the feature extraction layers of the network, which improved performance and maintained the spatial resolution of the segmented images. In addition to this, the design of a deeper network allows the extraction of shallow features and in-depth features for performing more accurate segmentation. The designed network was evaluated on a benchmark dataset collected from the last three years (2016, 2017 and 2018) by ISIC [10–13].

Our contributions are listed as: (a) a new design of the CNN network organized into five building blocks for extracting low-level features to high-level features to allow auto segmentation on the whole image rather than processing a patch or subset, (b) incorporating an atrous dilation rate in each convolution feature layer to capture lesion and image at different scales, (c) a careful placement of many leakyReLU activation functions in each block instead of standard ReLU because the former has a small slope for negative values that makes the network learns faster and is more balanced.

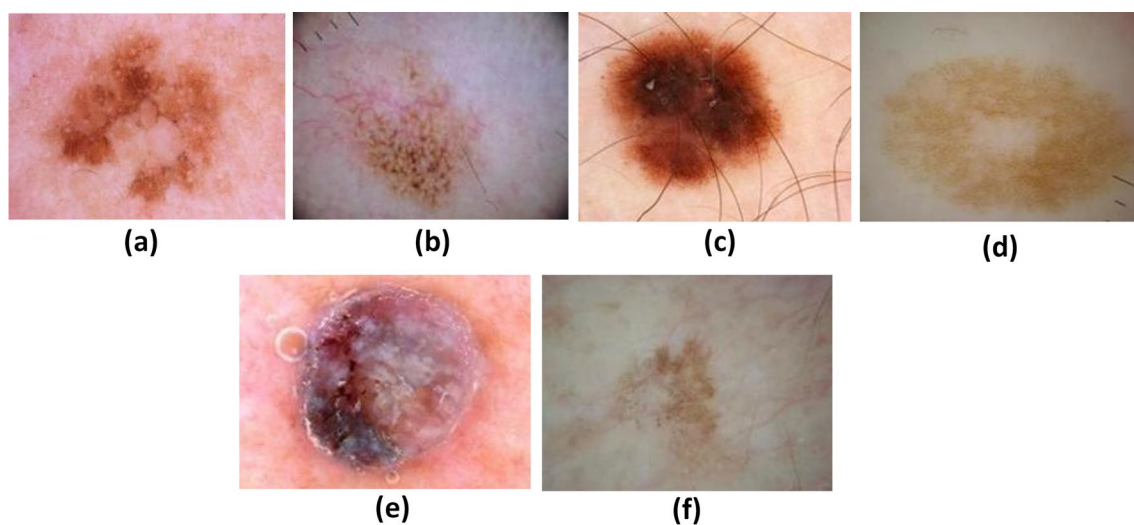


Fig. 1 Examples of noise artifacts. **a** Irregular boundaries, **b** blood vessels, **c** hairlines, **d** color illumination, **e** bubbles, **f** low contrast

This framework is designed to meet two main challenges: (a) lesion segmentation with high performance on images containing irregularities and non-uniform borders, (b) achieving low inference time, making the network faster.

Background

Researchers have made several attempts to develop image segmentation techniques for accurate lesion extraction. In the literature, segmentation approaches have been categorized into six categories: Edge-based methods [14], Thresholding-based methods [15], Clustering [16], Active contour [17], and Supervised approaches such as ANN [18]. Deep learning has been widely used for classification and object detection, where the idea of applying it to semantic segmentation has been an active area of research over the past few years.

Deep neural networks are effectively applied in semantic segmentation tasks to label each pixel with a class of object or non-object region. Some of the most popular CNNs proposed by the researchers to attempt segmentation task were fully convolutional neural network (FCN) [19], deconvolution networks (DeconvNet) [20] as an extension to FCN networks, SegNet [21], and UNet [22]. They are purely data-dependent, so their performance differs depending on the problem domain. A few limitations exist in the design of these networks, such as FCN uses pooling layers that reduce resolution and discards essential image information. However, semantic segmentation requires exact knowledge of class maps and needs to preserve 'where' information. Similarly, the encoder-decoder networks (SegNet and UNet) are efficient in maintaining output image resolution for some problem domains; however, networks become heavy and take high execution time due to many sampling and downsampling layers. In contrast, the concept of atrous convolutions introduced by Chen et al. [23] allows direct control of the resolution to preserve feature map information computed in the deep convolutional layers but have a high computation time due to its large network design.

Many studies proposed an extended version of these networks to perform segmentation of melanoma. For example, Kawahara et al. [24] extended VGG16 architecture to perform lesion segmentation by eliminating its fully connected layers and resizing the final feature map so that it matches with the output size. Bi et al. [25] proposed an FCN based ResNet model that learns the visual features of the lesion corresponding to each class based on their probability. This network is significantly different from other networks as it segments images according to their category and learns more in-depth features. Another work of Yu et al. [26] presented a very deep residual network having 50 layers to calculate rich and

more significant features for accurate recognition. Residual learning was applied to prevent overfitting and degradation problems of the convolutional neural networks. For lesion segmentation using a convolutional network, Al-Masni et al. [27] presented the full resolution convolutional network (FRCN) without any preprocessing. The VGG16 layers inspired the network by removing their subsampling layers to preserve pixels in their full resolution. According to work presented in [28] a convolutional multistage segmentation network is highly efficient in obtaining lesions from skin cancer images. In this, multiple stages of the network integrated outputs from different blocks combined with other steps. To create a fine segmentation mask, the network used pixel classification. In a study by Hassan et al. [29] dermoscopic skin network (DSNet) was used to segment lesions. A deeper view of the feature maps was obtained by using depth-wise separable convolutions instead of standard convolutions. A similar type of lesion segmentation work can be seen in the [30–33] that given different approaches for lesion segmentation either inspired by existing networks or extending them using transfer learning.

Recent developments have heightened the need for a melanoma detection system, and considerable literature has grown up around the theme of lesion segmentation. It has been analyzed that the most popular network's choices for semantic segmentation are U-Net, FCN, and SegNet, containing sampling and upsampling layers to maintain the spatial resolution of the output. The major problem in these networks is that they suffer from a low spatial resolution output due to the repetitive use of max-pooling and striding at consecutive layers. Moreover, the optimized, extended, and customized frameworks proposed in the literature for lesion segmentation tasks still exhibit difficulties such as low performance or high execution time due to complex architectures. These networks have not given satisfactory performance on the adopted skin cancer dataset. Hence, there is probably room for further improvement in performance by designing a new architecture. In the proposed work, we employ atrous convolutions that effectively maintain the segmented image's resolution. The network is designed from scratch by organizing different layers of the network, fine-tuning hyperparameters, and using a suitable loss function. This network has the advantage of being a small network with low execution time, minimal learning parameters, and high performance.

Methods and materials

This section explains the preparation of the dataset used for training, validation, and test, architectural details of the atrous convolution-based deep neural network, and discussion of metrics used for performance evaluation.

Datasets

In this study, three benchmark datasets were acquired from open-access dermatology repositories, ISIC archive [10–13], containing dermoscopic images of different skin cancer types such as Basal cell carcinoma (BCC), Melanoma, Squamous, and Nevus, including their ground truths which were used for training, validation, and testing purposes. The ISIC 2016 dataset contains 900 training and 379 testing images, ISIC 2017 has 2000 training, 150 validation, 366 testing images, and ISIC 2018 consist of 2594 training and 100 validation. The ISIC 2016 challenge has not provided external validation data; hence, the training set was divided in the ratio of 7:3 into training (630 images) and validation sets (270 images). Similarly, in ISIC 2018 challenge, the ground truth for test data was not provided; hence, the 30% of training data was used for testing purposes. The details of data available based on three years divided into different sets are given in Table 1.

The proposed approach targets dermoscopic images, which were produced by dermatoscopes. Available images are 8-bit with size ranges from 540 × 722 to

4499 × 6748 pixels. These images cannot be used directly for the network’s training due to their large size; thus, images are downsized to 192 × 256 × 3 dimensions using the nearest-neighbor interpolation technique. The principle idea used in this resizing method is to have an original image as a reference image based on which a new rescaled image is constructed. The constructed image results in a smaller size maintaining the aspect ratio and resolution as the original image. Different image sizes were taken into consideration during experiments, such as 224 × 224, 227 × 227, 256 × 297 but 192 × 256 × 3 had given the best performance for the proposed network. Data augmentation with rotation at an angle between [15°, 20°], scaling, and translation with factors [−6, 5] is applied to enlarge the training dataset and to overcome the problem of underfitting or overfitting that occurs in neural networks owing to the insufficient availability of data.

Proposed DilatedSkinNet architecture

With the invention of atrous convolutions in CNN architecture by Chen et al. [9, 23] to achieve a wider field of view, research on the development of dilated CNN networks continued for different applications, and a high execution time is one of the significant challenges required to achieve. In this paper, we designed an end-to-end trainable deep neural network architecture having 16 convolutional layers with different dilation factors, as shown in Fig. 2. The structure of DilatedSkinNet is interpreted in two primary steps: feature extraction and pixel classification. Using multiple scaling rates, the network

Table 1 Details of the ISIC challenge data

Dataset name	Training	Validation	Test
ISIC 2016	630	270	379
ISIC 2017	2000	150	366
ISIC 2018	1816	100	780
Total	4446	520	1525

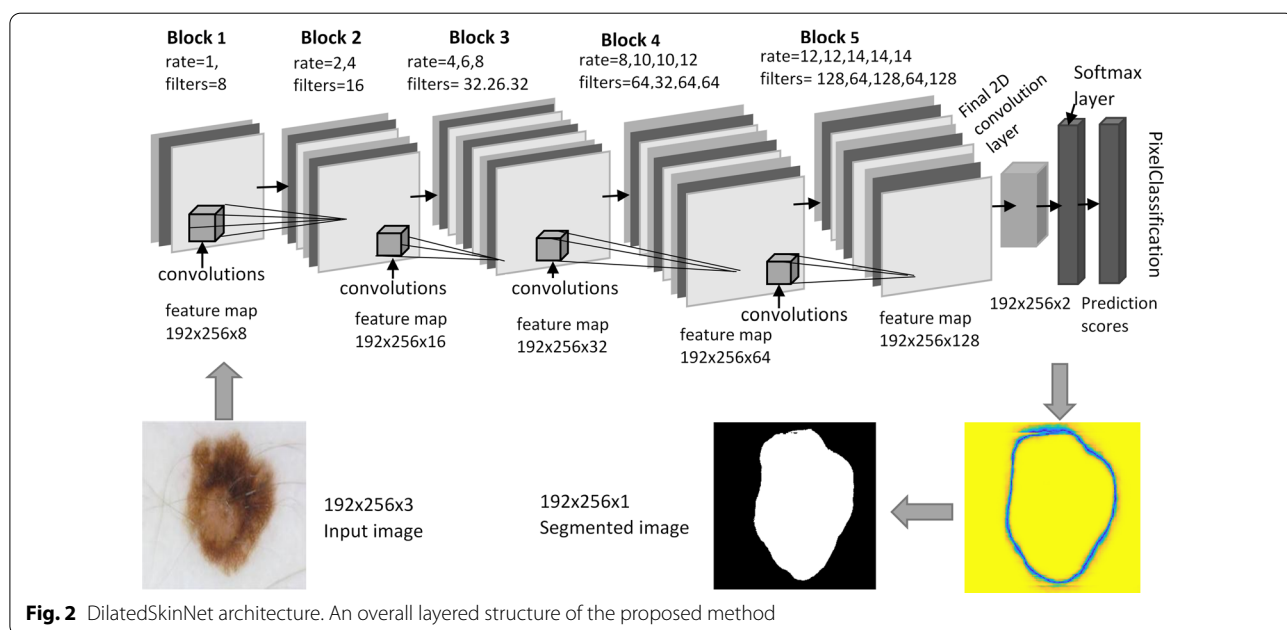


Fig. 2 DilatedSkinNet architecture. An overall layered structure of the proposed method

used atrous convolutions to enlarge the filter’s view. The pooling operation used in other semantic networks reduces the dimensional size of output feature vectors and the loss of information. In contrast, dilated convolutions expand the receptive field’s view to obtain in-depth information without using pooling operation and are suitable for maintaining the spatial resolution of the segmented image. This section discusses the working of networks divided among various layers such as feature extraction, use of atrous convolutions, normalization, activation function, and classification.

Feature extraction

The convolutional layer in the network is responsible for performing the feature calculation process. The initial image is divided into multiple sub-blocks based on a certain size and those sub-blocks compute features of the input image. The feature calculation is a process where a small matrix called kernel or filter slides over an image and transforms the image pixel’s values as:

$$C(I,K)_{x,y} = \sum_{i=1}^{m_h} \sum_{j=1}^{m_w} \sum_{k=1}^{m_c} I_{x+i-1,y+j-1,k} * K_{i,j} \tag{1}$$

Here m_h is the height, m_w is width, and m_c is the number of channels of an input image I. The number of channels of a kernel needs to be the same as the channels of an input image. The other parameters that we set artificially in the convolution layer are stride and padding. However, the dilation rate is set in each convolutional layer, so the stride and padding are zero in the network. Multiple filters are used in each convolutional layer to generate feature maps (see Table 2). The advantage of using multiple filters is that each filter convolves over the whole image separately to calculate many different features and produces the rich feature map used by the next layers.

Atrous/dilation CNN model

The use of atrous convolutions with different dilation rates at each convolutional layer helps extract a rich feature map. It allows increasing the view of the filter’s field (the space of the input vector that a layer can see) to assimilate a larger context. Therefore, it provides an efficient methodology to determine the best trade-off between correct localization and context absorption without increasing the computational parameters.

In the convolutional layers, the convolution operation is performed by sliding a template over an image for extracting features. We used atrous convolutions instead of general convolutions, which are decisive for extracting more contextual information. The standard convolutional operation is described in (2) with the dilation rate always ‘1’.

$$C[i] = \sum_{s=1}^S I[i + s] * K[i] \tag{2}$$

whereas, (3) describes the atrous convolution operation when the dilation rate is more than ‘1’.

$$C[i] = \sum_{s=1}^S I[i + s.r] * K[i], \quad r \geq 2 \tag{3}$$

We used two kernel sizes, 3×3 and 1×1 , in the atrous convolutional layers with different dilation rates at each layer to overcome the ‘gridding effect’ that occurs due to the use of the same dilation rate. The dilation factor increased at a rate of ‘2’ at each successive 2-dimensional convolutional layer. Figure 3 represents the impact of using dilation rate ‘2’, ‘2’ and ‘4’ on filter’s view over an input image.

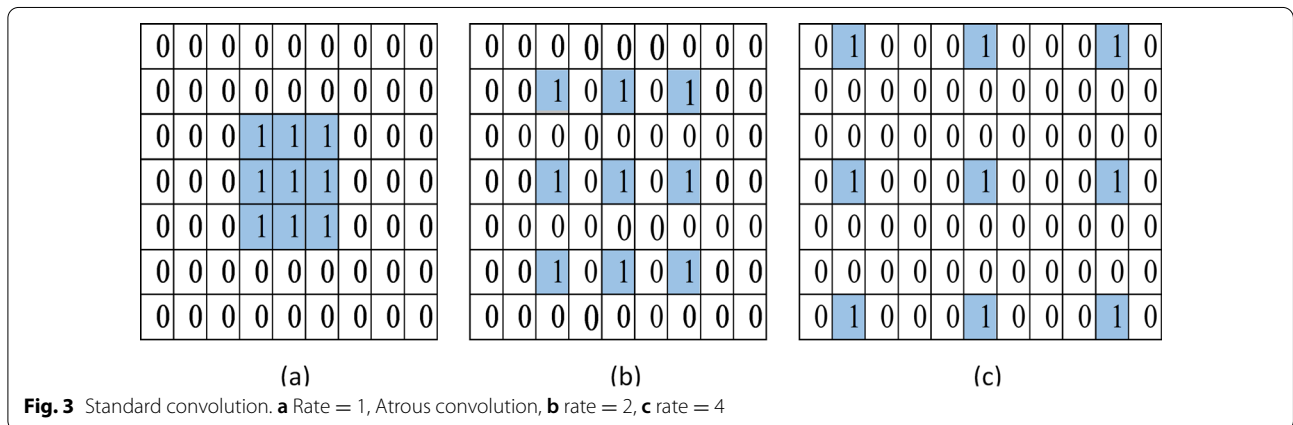


Table 2 Architecture details of the DilatedSkinNet

Block	Layer	Kernel size, Feature maps	Atrous dilation Rate	Output size
Input layer	192 × 256 × 3	–	–	192 × 256 × 3
Block 1	Conv1	3 × 3, 8	1	192 × 256 × 8
Block 2	Conv2_1	3 × 3, 16	2	192 × 256 × 16
	Conv2_2	3 × 3, 16	4	192 × 256 × 16
Block 3	Conv3_1	3 × 3, 32	4	192 × 256 × 32
	Conv3_2	1 × 1, 16	6	192 × 256 × 16
	Conv3_3	3 × 3, 32	8	192 × 256 × 32
Block 4	Conv4_1	3 × 3, 64	8	192 × 256 × 64
	Conv4_2	1 × 1, 32	10	192 × 256 × 32
	Conv4_3	3 × 3, 64	10	192 × 256 × 64
	Conv4_4	3 × 3, 64	12	192 × 256 × 64
Block 5	Conv5_1	3 × 3, 128	12	192 × 256 × 128
	Conv5_2	1 × 1, 64	12	192 × 256 × 64
	Conv5_3	3 × 3, 128	14	192 × 256 × 128
	Conv5_4	1 × 1, 64	14	192 × 256 × 64
	Conv5_5	3 × 3, 128	14	192 × 256 × 128
	final_Conv	1 × 1, 2	–	192 × 256 × 2
	Softmax	–	–	192 × 256 × 2
OutPutMap	Pixel classification	Cross entropy	Loss function	192 × 256 × 2

Data normalization

The batch normalization is used to speed up the process of network learning using internal normalization values. The word ‘batch’ refers to the group or set of data processed at a time, where mini-batch size is a variation of the gradient descent algorithm that splits the training dataset into small batches to calculate network error and update network coefficients. This layer in the network is used to normalize each input according to the mini-batch size. The batch normalization layer performs a series of operations to normalize the data. Firstly, the standardization process converts the batch of input data so that their mean is zero and the standard deviation is one. The mean (μ) is computed using (4) by summing up all inputs (x_i) of the batch (B) and divided by the total number of inputs (n). The resultant vector contains each input sample’s mean value. The variance (σ_B^2) (5) is obtained by squaring the standard deviation of the input. It is computed by taking the square of each input sample (x_i) in the current batch (B) subtracted from the mean (μ_B).

$$\text{mean, } \mu_B = \frac{1}{N} \sum_{i=1}^N x_i \tag{4}$$

$$\text{variance, } \sigma_B^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_B)^2 \tag{5}$$

In the following operation (6), the mean value of the current batch is subtracted from each input instance and divided by the square root of addition between standard deviation and smoothing term ϵ . This term is set as ‘0.00005’ to avoid division by a zero number in the calculations.

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \tag{6}$$

Finally, the output of the batch normalization process is obtained by re-scaling γ and offsetting β of the input values using (7). These two parameters were learned during the training process and optimized to ensure accurate normalization.

$$\text{BN}_{\gamma\beta}(x_i) \quad \text{or} \quad y_i = \gamma \hat{x}_i + \beta \tag{7}$$

The batch normalization is generally placed between the convolutional and ReLU layers to stabilize the learning process and reduce the number of epochs. After each convolutional layer, the batch normalization is to reduce the internal covariate shift, which significantly improves the network’s learning efficiency.

Activation function

The activation function used in the network is leakyReLU, which applies the threshold function to each element in the input and multiplies all negative values by a fixed scalar

'a' This layer passes the output element as the input to the next layer directly if it is positive; otherwise, it outputs to a value multiplied by 'a' as given:

$$\text{leakyReLU} = \begin{cases} a * x, & x < 0, \\ x, & x \geq 0 \end{cases} \quad (8)$$

In the proposed CNN, the scalar 'a' value is set as '0.1'. The activation function is responsible for transforming the summed weighted input from the node into the node's activation.

Classification

In the second part of the proposed network, instead of fully connected layers, a 1 × 1 convolutional layer was created to flatten the feature maps and minimize the number of channels. Then, the SoftMax layer was used, accepting the sparse feature sets as input for classifying each pixel of the image into two defined classes, i.e., lesion and background. The SoftMax function normalizes the weighted sum feature values to probability scores between 0 and 1. The output of this layer was the probability of each pixel mapped to each class. The last layer of the network was the pixel classification layer that produces a categorical label (background or lesion) for each pixel based on the probability score generated by the SoftMax layer. This layer also uses a loss function to calculate the network's prediction error rate. Over several iterations, the network's training is repeated to minimize the loss function's value. The structural details of the network illustrating a design of layers, kernel size, number of filters, and dilation rate are explained in Table 2.

Model training

The network is trained and optimised depending upon the loss function that measures the error between the prediction score P and target T. In this paper, the weighted cross-entropy loss [34] function was employed to measure the error as:

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^K \sum_{n=1}^N w_i T_{n_i} \log(P_{n_i}) \quad (9)$$

Here, N is the number of observations, K is the number of classes, and w is a vector of weights determined by the network for each class. The stochastic gradient descent algorithm is used to update the network weights and biases to reduce the loss value by applying small changes in the direction of optimization.

$$\theta_{i+1} = \theta_i - \alpha \nabla L(\theta_i) \quad (10)$$

Here i is the number of iterations, α > 0 is the learning parameter (set as '0.01'), θ is a parameter vector, and ∇L(θ_i) is the gradient of the loss function. The algorithm

evaluates the gradient at each iteration and updates parameters over a mini-batch set. The larger weight values can cause a network to be stuck into the local minima; thus momentum term is added in the gradient descent algorithm to reduce the oscillations as given in (11). The values of these hyperparameters set for the network's training are shown in Table 3.

$$\theta_{i+1} = \theta_i - \alpha \nabla L(\theta_i) + \gamma(\theta_i - \theta_{i+1}) \quad (11)$$

Evaluation metrics

The performance of the proposed DilatedSkinNet is evaluated quantitatively using performance metrics such as accuracy (ACC), Jaccard index (JAC), and Dice-coefficient (DICE). The value of these parameters was calculated for the test dataset and is expected to be higher for good segmentation results. The ACC parameter indicates the number of corrected pixels identified over the total number of pixels. A statistical measure to determine the similarity ratio between the ground truth and predicted label is known as the JAC index. The DICE computes the boundary contour matching index between the predicted and accurate segmentation in the ground truth.

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (12)$$

$$\text{JAC} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (13)$$

$$\text{DICE} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \quad (14)$$

The parameters TP, TN, FP, and FN, denotes the true positives, true negatives, false positives, and false negatives, respectively, which are elements of the confusion matrix. TP represents those pixels that are segmented correctly, whereas incorrectly segmented pixels are considered FN.

Table 3 Optimised hyperparameters during training

Parameter	Values
Input Image size	192 × 256 × 3
Batch size	16
Learning parameter, α	0.01
L2Regularization	0.005
Momentum, γ	0.9
Loss function, E(θ)	Weighted cross entropy loss
Optimiser	SGDM

On the other hand, non-lesion pixels, if classified correctly then considered TN; otherwise, FP.

Results and discussion

The proposed network is trained for three years' datasets (ISIC 2016–2018) separately, having 4446 training images, 520 validation, and 1525 test images. The proposed network is implemented in Matlab 2020a with GeForce GTX 1080 Ti hardware configuration with a computation capacity '7.5'. To show the impact of using data augmentation, atrous convolutions, leakyReLU activation function, and use of sigmoid layer on the performance of the model are displayed in Table 4. It illustrated that the proposed model with augmentation, atrous convolutions, leakyReLU, and softmax achieved higher ACC, JAC, and DICE index with low training time than the network without them.

In order to illustrate the generalization of the proposed model, we trained the network on the ISIC 2018 set and evaluated it on the PH^2 dataset, ISIC 2016, and ISIC 2017 test sets as illustrated in Table 5. The network trained on ISIC 2018 and tested on ISIC 2016 and 2017 test sets showed higher performance with a margin of ($\pm 1\%$) than the network trained individually on ISIC 2016–2018 datasets and evaluated on their respective test sets. For example, the ACC increased from 94.0 to 95.0%, 87.9 to 88.8% on ISIC 2016 and 2017 test sets, respectively. Additionally, a dataset PH^2 [35] which is widely used in literature studies is included and evaluated using ISIC 2018 trained network.

The visual outputs predicted by the proposed model for a few samples are shown in Fig. 4 that closely resembles the expected ground truths. Further, in Table 6, the results of the given model are demonstrated in comparison to the existing semantic segmentation networks named UNet, SegNet, and DeepLabv3+. The proposed network showed better performance as compared to the existing segmentation frameworks. The networks were trained by fine-tuning them on the same datasets for

Table 5 Evaluation of DilatedSkinNet trained with ISIC 2018 and tested on PH^2 , ISIC 2016 and 2017 tests sets

Datset	ACC	JAC	DICE
PH^2	0.900	0.891	0.708
ISIC 2016 test set	0.950	0.904	0.949
ISIC 2017 test set	0.888	0.818	0.884

conducting a fair comparison. The pixel classification block of these networks was replaced by the new layers segmenting an image into two classes; lesion and background. The same hyperparameter configuration is used, and networks are trained from end to end for training these networks.

The performance of these networks is recorded on the individual test sets of three years. The proposed network achieved an average ACC 95.0%, JAC 90.4%, and DICE 94.9% on ISIC 2016, ACC 88.8%, JAC 81.8%, DICE 88.4% on ISIC 2017, and ACC 94.2%, JAC 89.1%, DICE 94.2% on ISIC 2018 dataset. In contrast to this, the JAC index computed by the UNet was 79.8%, 68.7%, and 79.3%, SegNet computed JAC of 81.3%, 67.9%, 73.0%, and DeepLabv3+ gained JAC of 89.2%, 73.0%, and 88.8% on ISIC 2016, 2017 and 2018 datasets respectively. The results show that the proposed network generalized well on the test sets compared to the state-of-the-art semantic segmentation networks. Moreover, the primary advantage of the proposed networks is that it yields high performance in less inference time comparatively. The graphs in Figs. 5 and 6 shows higher accuracy achieved by the DilatedSkinNet on ISIC 2016–2018 test and validation sets, respectively, in comparison to the SegNet, UNet, and DeepLabv3+ networks. Moreover, the box plots in Fig. 7 demonstrate that DilatedSkinNet is efficient in extracting lesion information with a high JAC score as compared to the other models.

To prove the robustness of the network, we compared our model with the top winners of the ISIC challenge

Table 4 Performance comparison of the DilatedSkinNet with its modified architectures

Methods	ISIC 2016				ISIC 2017				ISIC 2018			
	ACC	JAC	DICE	Time (s)	ACC	JAC	DICE	Time (s)	ACC	JAC	DICE	Time (s)
DilatedSkinNet	0.940	0.887	0.940	10	0.879	0.817	0.874	9	0.942	0.891	0.942	14
Without atrous	0.853	0.735	0.513	12	0.758	0.556	0.460	16	0.825	0.635	0.510	21
Without leakyReLU	0.935	0.855	0.622	11	0.874	0.784	0.578	14	0.863	0.736	0.556	19
Without augmentation	0.945	0.884	0.679	11.8	0.885	0.788	0.559	12.6	0.942	0.854	0.643	16.5
Sigmoid layer	0.450	0.329	0.496	27	0.465	0.338	0.506	27	0.499	0.331	0.498	51

Time is in seconds on test sets

The higher values are marked in bold

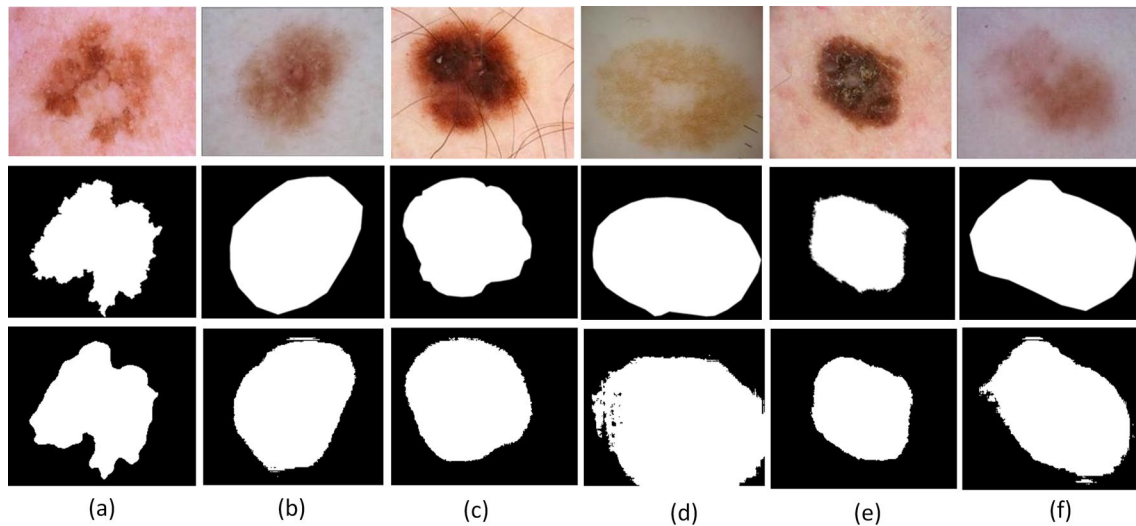


Fig. 4 Exemplary pairs of the segmentation result using DilatedSkinNet. The first row shows original images, second row are gold images and the last row are segmented images **a** irregular boundaries, **b** blood vessels, **c** hairlines, **d** color illumination, **e** bubbles, **f** low contrast

Table 6 Performance comparison of the DilatedSkinNet with UNet, SegNet, and DeepLabv3+ on ISIC 2016–2018 test sets

Methods	ISIC 2016				ISIC 2017				ISIC 2018			
	ACC	JAC	DICE	Time (s)	ACC	JAC	DICE	Time (s)	ACC	JAC	DICE	Time (s)
U-Net [22]	0.854	0.798	0.832	11	0.764	0.687	0.696	10	0.842	0.793	0.815	24
SegNet [21]	0.908	0.813	0.907	16	0.822	0.679	0.818	16	0.880	0.730	0.879	33
DeepLabv3+ [23]	0.952	0.892	0.952	16	0.878	0.730	0.881	19	0.939	0.888	0.941	20
DilatedSkinNet	0.950	0.904	0.949	10	0.888	0.818	0.884	9	0.942	0.891	0.942	14

Time is in seconds on test sets

The higher values are marked in bold

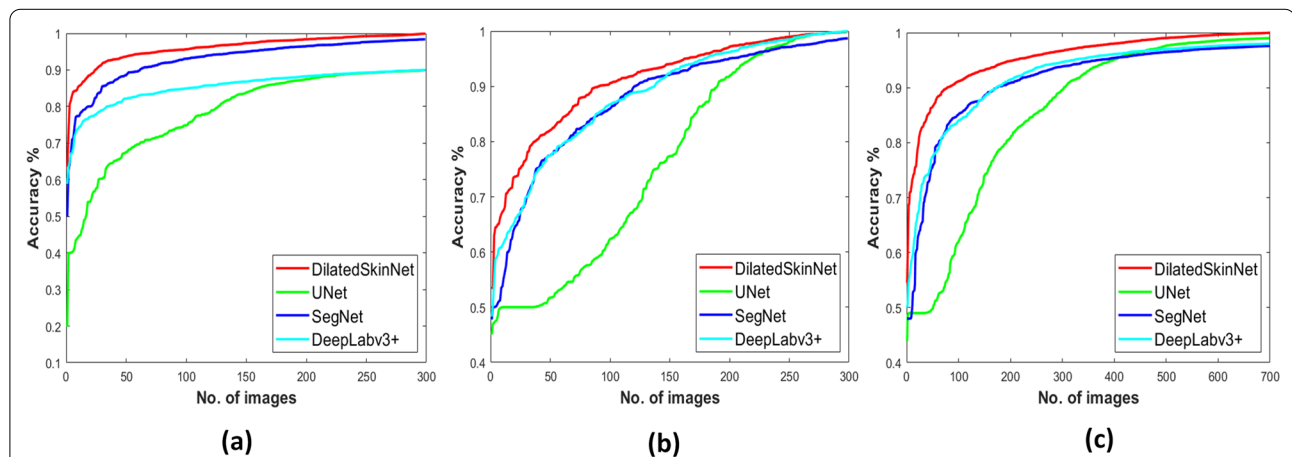


Fig. 5 Comparison of networks. Accuracy versus test samples on the ISIC **a** 2016, **b** 2017, **c** 2018 test sets

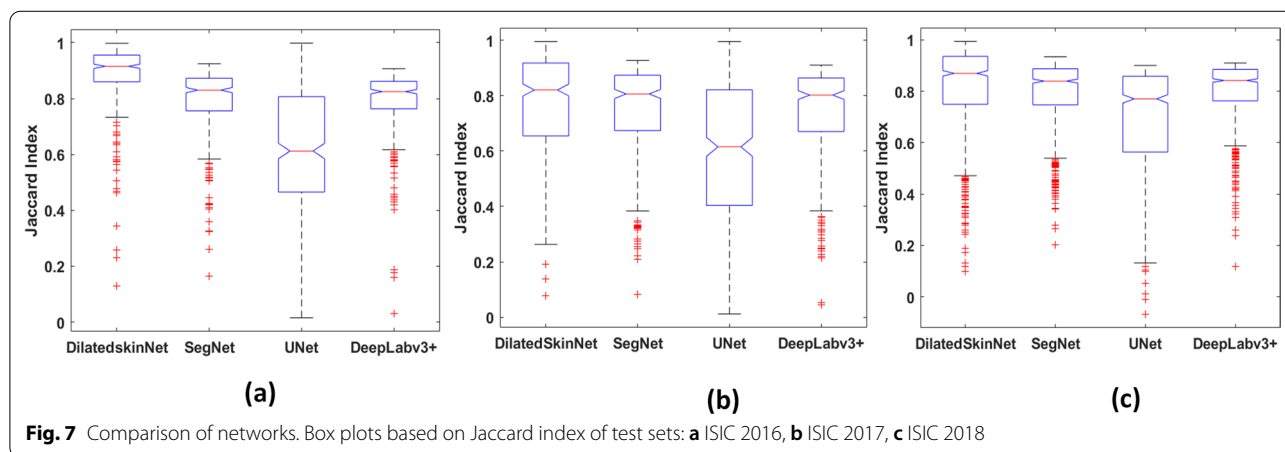
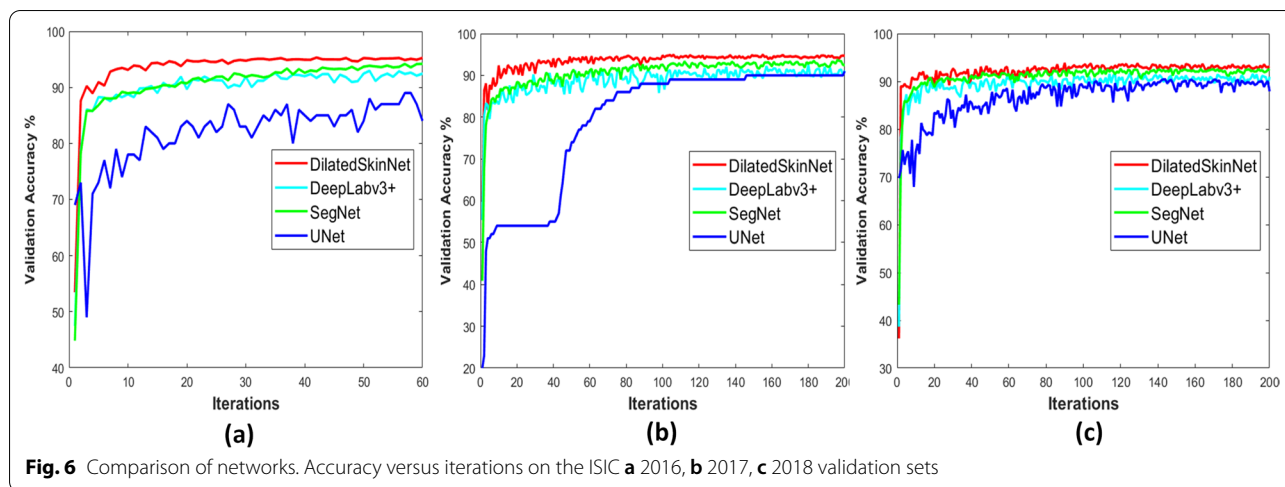


Table 7 Performance comparison of the DilatedSkinNet with winners of ISIC 2016–2018 challenge and some recent studies

ISIC 2016		ISIC 2017		ISIC 2018	
References	JAC	References	JAC	References	JAC
DilatedSkinNet	0.904	DilatedSkinNet	0.818	DilatedSkinNet	0.891
U. Sanchez [36]	0.843	Yuan [37]	0.765	Qian [38]	0.802
Yu [26]	0.829	Berseth [39]	0.762	Hao [40]	0.799
Rahman [41]	0.822	Bi [42]	0.760	Ji [43]	0.799
Huang [44]	0.811	Menegola [45]	0.754	Yuan [46]	0.798
Xie [47]	0.858	Zafar [48]	0.772	Ali [49]	0.735
Hassan [29]	0.859	Pour [30]	0.782	Lei [50]	0.824
Ashraf [51]	0.859	Liu [33]	0.794	Chu [52]	0.835
Tong [53]	0.845	Ashraf [51]	0.800	–	–
–	–	Tong [53]	0.742	–	–

2016–2018, as given in Table 7. The critical parameter used by the ISIC challenge to announce the winner was the JAC index, for which DilatedSkinNet achieved the highest value of 90.4% on ISIC 2016, 81.8% on ISIC 2017, and 89.1% on the ISIC 2018 dataset. Moreover, the last two rows displayed the comparison of DilatedSkinNet with the recent studies published in the years 2019–2021. It is presented that DilatedSkinNet outperformed all given studies by offering higher JAC scores.

The experimental results proved that the proposed lightweight network could calculate the expected information efficiently rather than designing a heavy network such as DeepLabV3+. The number of learnable parameters for SegNet, UNet, and DeepLabV3+ is 29M, 31M, 20M, respectively, whereas, for DilatedSkinNet,

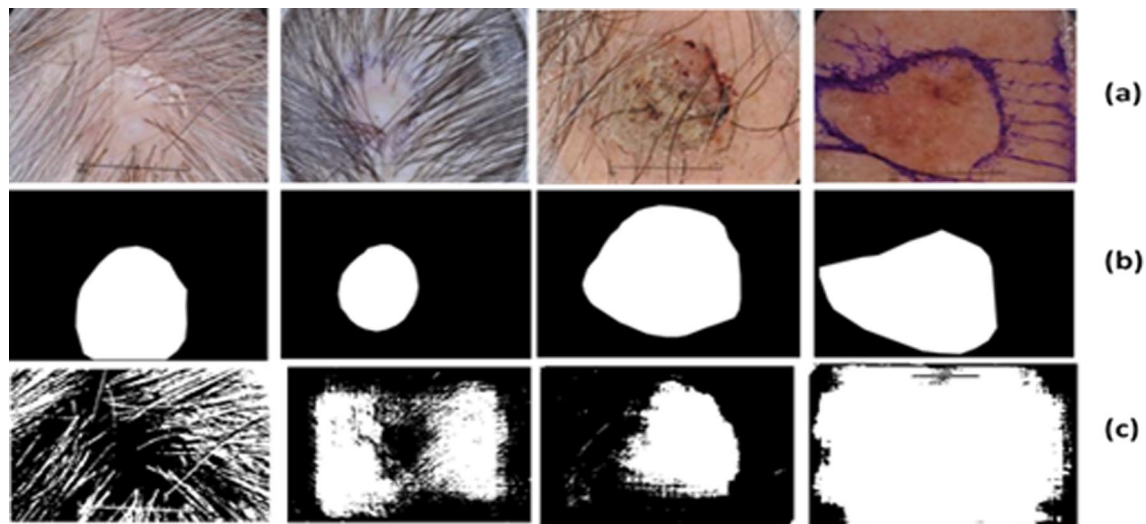


Fig. 8 Failure cases. Poor segmentation results using the proposed DilatedSkinNet: **a** Original images, **b** expected segmentation mask, **c** segmented outputs, respectively

the number is 3.27K. Additionally, the inferencing time of DilatedSkinNet is also less than others, as shown in Table 6.

Due to the structural dissimilarities, it was challenging to obtain an accurate border of the lesion region. There were some challenging samples, as given in Fig. 8 which were not properly segmented by the given method. The reason for this failure is the presence of noise elements such as dense hairlines and dark ink projections that cause impediments in extracting an accurate region of interest. The network partially segmented these images but was not as accurate as of the ground truth images. Thus, in the future, the design of any pre-processing technique will be taken into consideration for removing noisy elements, primarily hairlines, from images.

Conclusion

This paper presented a method for the segmentation of lesions in dermoscopic skin cancer images. The proposed network design is a CNN method based on the use of atrous convolutions to replace pooling layers. The atrous dilations can expand the receptive field of the input vector without using pooling layers. These allow each convolution output to contain a wide range of information without extra computations and lose the image's resolution. The network achieved higher performance by minimizing the cross-entropy loss across mini-batches. The network successfully extracted relevant features from the different dermoscopic skin cancer images and generated segmented image maps. We observed through the experimental results that the proposed network successfully segmented accurate

lesion areas that would aid future research work to develop a highly efficient CAD system to classify melanoma and non-melanoma. The network was successful in segmenting the majority of challenging cases, such as irregular boundaries, gel bubbles, low contrast, and color illumination, as given in Fig. 4. However, a few challenging samples, mainly containing dense hairlines, were not accurately segmented illustrated in Fig. 8, which will be considered in the future scope of this research. Based on the higher performance of the DilatedSkinNet, we will focus on its application areas to make it a more general approach, including automatic segmentation and tracking over multiple image sequences.

Abbreviations

CNN: Convolutional neural network; ISIC: International skin imaging collaboration; ReLU: Rectified linear unit; BN: Batch normalization.

Acknowledgements

Not applicable.

Author contributions

Conceptualization: RK and HG; Methodology: RK and HG; Software: RK; Validation: RK, HG and RS; Data collection: RK; Writing—original draft preparation: RK; Writing—review and editing: HG, RS and ML; Supervision: HG; ML. Proofreading and revision: RS and ML. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

The benchmark dataset is used publicly available at (<https://challenge.isic-archive.com/data>).

Declarations

Ethics approval and consent to participate

This article does not contain any studies with human participants performed by any of the authors.

Competing interests

The authors declare no competing interest.

Consent for publication

Not applicable.

Author details

¹School of Engineering, Computer, and Mathematical Sciences, Auckland University of Technology, 55 Wellesley street, 1010 Auckland, New Zealand.

²School of Innovation Design and Engineering, Mälardalen University, Västerås, Sweden.

Received: 28 February 2021 Accepted: 13 April 2022

Published online: 29 May 2022

References

- Elwood JM, Jopson J. Melanoma and sun exposure: An overview of published studies. *Int J Cancer*. 1997;73(2):198–203.
- Bogo F, Peruch F, Fortina A, Peserico E, Celebi M, Mendonca T, Marques J. Where's the lesion? Variability in human and automated segmentation of dermoscopy images of melanocytic skin lesions. Boca Raton: CRC Press; 2015.
- Schadendorf D, van Akkooi AC, Berking C, Griewank KG, Gutzmer R, Hauschild A, Stang A, Roesch A, Ugurel S. Melanoma. *Lancet*. 2018;392(10151):971–84.
- Massey University of New Zealand: Environmental Health Indicators New Zealand. <https://www.ehinz.ac.nz/indicators/uv-exposure/melanoma/>. Accessed 10 July 2020.
- The Skin Cancer Foundation. Skin cancer facts and statistics. <https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/>. Accessed Jan 2021.
- Ruiz ES, Morgan FC, Zigler CM, Besaw RJ, Schmults CD. Analysis of national skin cancer expenditures in the United States medicare population, 2013. *J Am Acad Dermatol*. 2019;80(1):275–8.
- Guy GP Jr, Machlin SR, Ekwueme DU, Yabroff KR. Prevalence and costs of skin cancer treatment in the US, 2002–2006 and 2007–2011. *Am J Prev Med*. 2015;48(2):183–7.
- Rigel DS, Friedman RJ, Kopf AW, Polsky D. ABCDE: an evolving concept in the early detection of melanoma. *Arch Dermatol*. 2005;141(8):1032–4.
- Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. 2017. arXiv preprint. [arXiv:1706.05587](https://arxiv.org/abs/1706.05587).
- Gutman D, Codella N-C, 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). 2016. arXiv preprint. [arXiv:1605.01397](https://arxiv.org/abs/1605.01397).
- Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, Kalloo A, Liopyris K, Mishra N, Kittler H, et al. Skin lesion analysis toward melanoma detection: a challenge at the 2017 international symposium on biomedical imaging (ISIC), hosted by the international skin imaging collaboration (ISIC). In: IEEE 15th international symposium on biomedical imaging (ISBI 2018). IEEE; 2018. p. 168–172.
- Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, Helba B, Kalloo A, Liopyris K, Marchetti M, et al. Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC). 2019. arXiv preprint. [arXiv:1902.03368](https://arxiv.org/abs/1902.03368).
- Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data*. 2018;5: 180161.
- Lee I, Du X, Anthony B. Hair segmentation using adaptive threshold from edge and branch length measures. *Comput Biol Med*. 2017;89:314–24.
- Garnavi R, Aldeen M, Celebi ME, Varigos G, Finch S. Border detection in dermoscopy images using hybrid thresholding on optimized color channels. *Comput Med Imaging Graph*. 2011;35(2):105–15.
- Aitnouri E, Ouali M. Performance evaluation of clustering techniques for image segmentation. *Comput Sci J Mold*. 2010;54(3):271–302.
- Vasconcelos FFX, Medeiros AG, Peixoto SA, Reboucas Filho PP. Automatic skin lesions segmentation based on a new morphological approach via geodesic active contour. *Cogn Syst Res*. 2019;55:44–59.
- Moghaddam MJ, Soltanian-Zadeh H. Medical image segmentation using artificial neural networks. *Artif Neural Netw Methodol Adv Biomed Appl*. 2011;6:121–38.
- Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 3431–3440.
- Hong S, Noh H, Han B. Decoupled deep neural network for semi-supervised semantic segmentation. 2015. arXiv preprint. [arXiv:1506.04924](https://arxiv.org/abs/1506.04924).
- Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(12):2481–95.
- Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2015. p. 234–241.
- 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 Mach Intell*. 2017;40(4):834–48.
- Kawahara J, BenTaieb A, Hamarneh G. Deep features to classify skin lesions. In: IEEE 13th international symposium on biomedical imaging (ISBI). IEEE; 2016. p. 1397–1400.
- Bi L, Kim J, Ahn E, Kumar A, Feng D, Fulham M. Step-wise integration of deep class-specific learning for dermoscopic image segmentation. *Pattern Recogn*. 2019;85:78–89.
- 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*. 2016;36(4):994–1004.
- Al-Masni MA, Al-Antari MA, Choi M-T, Han S-M, Kim T-S. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput Methods Programs Biomed*. 2018;162:221–31.
- Bi L, Kim J, Ahn E, Kumar A, Fulham M, Feng D. Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Trans Biomed Eng*. 2017;64(9):2065–74.
- Hasan MK, Elahi MTE, Alam MA, Jawad MT. DermoExpert: skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation. *medRxiv*. 2021.
- Pour MP, Seker H. Transform domain representation-driven convolutional neural networks for skin lesion segmentation. *Expert Syst Appl*. 2020;144: 113129.
- Vesal S, Ravikumar N, Maier A. SkinNet: a deep learning framework for skin lesion segmentation. In: IEEE nuclear science symposium and medical imaging conference proceedings (NSS/MIC). IEEE; 2018. p. 1–3.
- Yuan Y, Chao M, Lo Y-C. Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance. *IEEE Trans Med Imaging*. 2017;36(9):1876–86.
- Liu L, Tsui YY, Mandal M. Skin lesion segmentation using deep learning with auxiliary task. *J Imaging*. 2021;7(4):67.
- Bishop CM. *Pattern recognition and machine learning*. New York: Springer; 2006.
- Mendonça T, Ferreira PM, Marques JS, Marcal AR, Rozeira J. PH²-A dermoscopic image database for research and benchmarking. In: 35th Annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2013. p. 5437–5440.
- U. Sanchez: ISBI 2016 challenge results. 2016. <https://challenge.kitware.com/submission/56fe2b60cad3a55ecee8cf74>.
- Yuan Y. Automatic skin lesion segmentation with fully convolutional-deconvolutional networks. 2017. arXiv preprint. [arXiv:1703.05165](https://arxiv.org/abs/1703.05165).
- Qian C, Liu T, Jiang H, Wang Z, Wang P, Guan M, Sun B. A detection and segmentation architecture for skin lesion segmentation on dermoscopy images. 2018. arXiv preprint. [arXiv:1809.03917](https://arxiv.org/abs/1809.03917).
- Berseth M. ISIC 2017-skin lesion analysis towards melanoma detection. CoRR. [abs/1703.00523](https://arxiv.org/abs/1703.00523). 2017.

40. Du H, Seok JY, Ng Ngiam D, Yuan K, Feng M. Team Holiday Burned at ISIC Challenge 2018. 2018. <https://challenge.isic-archive.com/leaderboards/2018>.
41. Rahman M. ISBI 2016 challenge results. 2016. <https://challenge.kitware.com/submission/56fbfa1bcad3a54f8bb809bf>.
42. Bi L, Kim J, Ahn E, Feng D. Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. 2017. arXiv preprint. [arXiv:1703.04197](https://arxiv.org/abs/1703.04197).
43. Ji Y, Li X, Zhang G, Lin D, Chen H. Automatic skin lesion segmentation by feature aggregation convolutional neural network. Technical report. 2018.
44. Huang L, Zhao Y-G, Yang T-J. Skin lesion segmentation using object scale-oriented fully convolutional neural networks. *Signal Image Video Process*. 2019;13(3):431–8.
45. Menegola A, Tavares J, Fornaciali M, Li L.T, Avila S, Valle E. Recod titans at ISIC challenge 2017. 2017. arXiv preprint. [arXiv:1703.04819](https://arxiv.org/abs/1703.04819).
46. Xue Y, Gong L, Peng W, Huang X, Zheng Y. Automatic skin lesion analysis with deep networks. 2018. <https://challenge.isic-archive.com/leaderboards/2018>.
47. Xie F, Yang J, Liu J, Jiang Z, Zheng Y, Wang Y. Skin lesion segmentation using high-resolution convolutional neural network. *Comput Methods Programs Biomed*. 2020;186: 105241.
48. Zafar K, Gilani SO, Waris A, Ahmed A, Jamil M, Khan MN, Sohail Kashif A. Skin lesion segmentation from dermoscopic images using convolutional neural network. *Sensors*. 2020;20(6):1601.
49. Ali R, Hardie RC, De Silva MS, Kebede TM. Skin lesion segmentation and classification for ISIC 2018 by combining deep CNN and handcrafted features. 2019. arXiv preprint [arXiv:1908.05730](https://arxiv.org/abs/1908.05730).
50. Lei B, Xia Z, Jiang F, Jiang X, Ge Z, Xu Y, Qin J, Chen S, Wang T, Wang S. Skin lesion segmentation via generative adversarial networks with dual discriminators. *Med Image Anal*. 2020;64: 101716.
51. Ashraf H, Waris A, Ghafoor MF, Gilani SO, Niazi IK. Melanoma segmentation using deep learning with test-time augmentations and conditional random fields. *Sci Rep*. 2022;12(1):1–16.
52. Chu T, Li X, Vo H.V, Summers R.M, Sizikova E. Improving weakly supervised lesion segmentation using multi-task learning. In: *Medical imaging with deep learning*. PMLR; 2021. p. 60–73.
53. Tong X, Wei J, Sun B, Su S, Zuo Z, Wu P. ASCU-Net: attention gate, spatial and channel attention U-Net for skin lesion segmentation. *Diagnostics*. 2021;11(3):501.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

