Automatic Lesion Segmentation Using Atrous Convolutional Deep Neural Networks in Dermoscopic Skin Cancer Images

Ranpreet Kaur* Hamid GholamHosseini and Roopak Sinha

Abstract

Background: Among skin cancers, melanoma is the most dangerous and aggressive form, exhibiting a high mortality rate worldwide. Biopsy and histopathological analysis are common procedures for skin cancer detection and prevention in clinical settings. A significant step involved in the diagnosis process is the deep understanding of 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 which makes its prediction challenging. Moreover, at the initial stage, it is difficult to predict melanoma 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 high attention in recent years due to their remarkable performance, therefore, in this work, we proposed a novel, end-to-end atrous spatial pyramid pooling based convolutional neural network (CNN) framework for automatic lesion segmentation. This architecture is built based on the concept of atrous dilated convolutions which are effective for semantic segmentation. A dense deep neural network is designed using several building blocks consisting of convolutional, batch normalization, leaky ReLU layer with fine-tuning of hyperparameters contributing towards higher performance.

Conclusion: The network was tested on three benchmark datasets by International Skin Imaging Collaboration, i.e. ISIC 2016, ISIC 2017, and ISIC 2018. The experimental results showed that the proposed network achieved an average Jaccard index of 86.5% on ISIC 2016, 81.2% on ISIC 2017, and 81.2% on ISIC 2018 datasets, respectively which is recorded as higher than the top three winners of the ISIC challenge. Also, the model successfully extracts lesions from the whole image in one pass, requiring no pre-processing process. The conclusions yielded that network is accurate in performing lesion segmentation on skin cancer images.

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 long hours which produces a pigment known as melanin in the upper skin layer [1]. Moreover, fair complexion, sunburn, genetic history, and weak immune
system are other risk factors that contribute towards 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 key 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 ABCDE (Asymmetry, Border, Color, Diameter, and Evolution) metrics 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 in accuracy. Thus, computer-aided systems are widely adopted for the timely detection of melanoma, where cancer image segmentation is the most significant process for the detailed analysis of lesion structure. The heterogeneous appearance of lesion area in terms of color, size, shape, location, and texture makes the segmentation task very challenging as shown in samples of Figure 1. Therefore, there is a need for an automatic segmentation approach to assist dermatologists in understanding the nature and pattern of the lesion area.

![Figure 1 Examples of noise artifacts](image)

Recent advances in artificial intelligence in particular deep learning offered numerous advantages towards automatic and accurate identification. Therefore, for the accurate extraction of lesion pattern in skin cancer images, we proposed a novel deep neural framework which is suitable for segmentation as compared to traditional approaches. We also embedded the atrous convolutions in the feature extraction layer of the network, which improved performance and maintained the spatial
resolution of the image. In addition to this, dense layers are designed that allow shallow features and in-depth features to be combined for more accurate segmentation. The designed network was evaluated on a benchmark dataset collected from the last three years (2016, 2017 & 2018) by International Skin Imaging Collaboration (ISIC) [8, 9, 10, 11].

Our contributions are listed as: (a) CNN layers organized in a new fashion divided into 5 building blocks for extracting low-level features to high-level features which allow auto segmentation on the whole image rather than processing a patch or subsets, (b) incorporating an atrous dilation rate in the convolution layer to have a wider view of the field at same computational cost as simple convolutions, (c) use of leaky ReLU activation function instead of ReLU as it has a small slope for negative values which makes it learn faster and more balanced. This framework is designed to meet two main challenges of segmentation: (a) segmenting accurate lesion area including its irregularities and non-uniform borders, (b) training with a large dataset to makes it more consistent.

Background

Several attempts have been made by researchers to develop image segmentation techniques for accurate lesion extraction. In the literature, segmentation approaches have been categorized into six categories: Edge-based methods [12], Thresholding-based methods [13], Clustering [14], Active contour [15], and Supervised approaches such as ANN [16]. For the past few years, deep learning has been widely used for classification and object detection, where the idea of applying it to semantic segmentation is a new area to explore. Deep neural networks label each pixel with a class of object or non-object region and are effective in semantic segmentation applications. In 2015, Long et al. [17] proposed an end-to-end fully convolutional neural network (FCN) to perform pixel-wise segmentation. They adapted AlexNet, VGGNet, and GoogleNet classification models for semantic segmentation by replacing convolutional layers with fully connected layers to generate segmentation maps and were faster compared to the patch classification approach. Hong et al. [18] proposed a semi-supervised learning approach, deconvolution networks (DeconvNet), as an extension of FCN networks. This architecture decoupled classification and segmentation tasks by adding bridging layers to construct class-oriented feature maps that reduce training and inferencing time. The main problem with FCN was pooling layers, that increased field of view and discarded the essential information. However, semantic segmentation requires exact knowledge of class maps and needs to preserve ‘where’ information. Then in 2017, an encoder-decoder based architecture was presented by Badrinarayanan et al. called SegNet [19] that has a network layout similar to VGG16. In this network, the encoder section reduces spatial dimensionality with pooling layers and the decoder gradually regains the object information and spatial details using pooling indices computed by the max-pooling layers. UNet is another famous network inspired by encoder-decoder architecture which is successfully applied for the segmentation of medical images [20]. Although the encoder-decoder networks are efficient in maintaining output image resolution, networks become heavy and take high execution time. In contrast, the concept of atrous convolutions introduced by Chen et al. [21] in 2018 allows controlling resolution explicitly to
Some studies proposed techniques for melanoma segmentation, for example, Jeremy Kawahara et al. [22] 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. Lei Bi et al. [23] 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 deep features. Another work is given by Lequan Yu et al. [24] that 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. A similar type of lesion segmentation work can be seen in the [25, 26, 27] that given different approaches for lesion segmentation either inspired from 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 choices of networks 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 dense networks is the repetitive use of max-pooling and striding at consecutive layers that reduce the spatial resolution of the output feature map, which is not suitable for segmentation. Moreover, the optimized, extended, and customized frameworks proposed in the literature for lesion segmentation task still exhibit some difficulties such as low performance or high execution time due to complex architectures. There is probably room for further improvement in terms of accuracy by designing new architectures. Therefore, we used the idea of employing atrous convolutions in neural networks proposed by Chen et al. [21]. We designed a new network for lesion extraction by organizing layers of the network from scratch, fine-tuning hyperparameters, and use of suitable loss function. In the proposed network we have not used max-pooling layers because pooling layers reduce the resolution of the segmentation map. Instead, we employed atrous convolutions to increase the field of view over an image.

**Methods and Materials**

This section explains the preparation of the dataset used for training and testing of the network, architectural details of the atrous convolution [28] based deep neural network, and performance evaluation metrics.

**Database Preparation**

In this study, three benchmark datasets are acquired from open-access dermatology repositories, International skin Imaging Collaboration dermoscopic archive [8, 9, 10, 11], containing approximately 15000 dermoscopic images of Basal cell carcinoma (BCC), Melanoma, Squamous, and Nevus skin cancer types including their ground truths which were used for training and testing purposes. The ISIC 2016 dataset
contains 900 training and 379 testing images, ISIC 2017 has 2000 training and 366 testing images and ISIC 2018 consists of 2594 training and 1000 testing images. Extensive care was taken while dividing the whole database into training and testing sets. The proposed approach targets dermoscopic images, which were produced by the dermatoscopes. Available images are 8-bit with size ranges from 540×722 to 4499×6748 pixels. These images cannot be used directly for the training of the network due to their large size; thus, images are downsized to 192×256×3. Different image sizes were taken into consideration, such as 224×224,227×227,256×297 but 192×256×3 had given the best performance. 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**

After the invention of atrous convolutions in CNN architecture by Chen et al. [28] to achieve a wider field of view, research on this concept is continued, and high execution time is one of the problems. In this paper, we designed an end-to-end trainable dense neural network architecture having 16 convolutional layers employing dilation factor as shown in Figure 2. The structure of DilatedSkinNet is interpreted in two steps: feature extraction, and classification. The network used atrous convolutions to get a larger receptive field, as pooling operation tends to reduce the dimensional size of output vectors along with the loss of information. Dilated convolutions expand the receptive field’s view to obtain deep information without using pooling operation and are suitable for maintaining the spatial resolution of the segmented image.

![Figure 2](image)

**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:

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

(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 need 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. Stride defines the step size taken by the kernel while sliding over an input image, usually taken as ‘1’. The size of the output feature map reduces after each convolutional operation, thus, to maintain the dimensions of the output map as of the input image, we used padding that adds zeros for obtaining symmetry. The above equation is modified as follows after including these parameters:

\[
C(I, K) = \left( \left[ \frac{m_h + 2p - f}{s} + 1 \right] , \left[ \frac{m_w + 2p - f}{s} + 1 \right] \right) ; s > 0
\]  

(2)

\[
C(I, K) = (m_h + 2p - f, m_w + 2p - f) ; s = 0
\]  

(3)

The terms \( p, s, f \) defines padding, stride, and filter size respectively. When stride is more than one then (2) is used to compute output feature map, otherwise (3) is applied for stride value 'zero'. To extract many features, multiple filters are applied simultaneously in the convolutional layer. The output can be calculated in case of multiple filters as:

\[
[m, m, m_c] \ast [f, f, m_c] = \left( \left[ \frac{m + 2p - f}{s} + 1 \right] , \left[ \frac{m + 2p - f}{s} + 1 \right] , m_f \right)
\]  

(4)

In this equation, \( m \) is the input image dimensions, \( m_c \) is the number of channels of an input image, \( f \) denotes the filter size, and \( p, s, m_f \) represents padding, stride, and number of filters respectively. Figure 3 presents the convolution operation with multiple channels and filter size 3x3.

**Atrous/Dilation CNN model**

The use of atrous convolutions with different dilation rates at each convolutional layer helps to extract a dense feature map. It allows to increase the view of the filter’s field (the space of the input which 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 spatial pyramid pooling instead of general convolutions which is a powerful tool to extract more contextual information. The standard convolutional operation is described in (5)
with the dilation rate as ‘1’.

\[ C[i] = \sum_{s=1}^{S} I[i + s] \ast K[i] \] (5)

whereas, (6) describes the atrous convolution operation when the dilation rate is more than ‘1’.

\[ C[i] = \sum_{s=1}^{S} I[i + s \times r] \ast K[i], r \geq 2 \] (6)

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

**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 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 (\( \mu \)) is computed using (7) by summing up all inputs (\( x_i \)) of the
The resultant vector contains each input sample's mean value. The variance ($\sigma^2_B$) is obtained by squaring the standard deviation of the input as given in (8). It is computed by taking the square of each input sample ($x_i$) in the current batch (B) subtracted from the mean ($\mu_B$).

$$\text{mean}, \mu_B = \frac{1}{N} \sum_{i=1}^{N} x_i$$  \hspace{1cm} (7)

$$\text{variance}, \sigma^2_B = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_B)^2$$  \hspace{1cm} (8)

In the following operation (9), 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 $\epsilon$. This term is set as ‘0.00005’ to avoid division by a zero number in the calculations.

$$\tilde{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma^2_B + \epsilon}}$$  \hspace{1cm} (9)

Finally, the output of the batch normalization process is obtained by re-scaling $\gamma$ and offsetting $\beta$ of the input values using (10). These two parameters are learned during the training of the CNN network and should be optimal to ensure accurate normalization.

$$\text{BN}_{\gamma,\beta}(x_i) \quad y_i = \gamma \tilde{x}_i + \beta$$  \hspace{1cm} (10)

The batch normalization is generally placed between the convolutional and ReLU layer to stabilize the learning process and to reduce the number of epochs. The batch normalization is after each convolutional layer to reduce the internal covariate shift that greatly improves the learning efficiency of the network.

**Activation function**

The activation function used in the network is leakyReLU that applied the threshold function to each element in the input and multiplied all negative values by a fixed
This layer passes the output element as the input to the next layer directly if it is positive; otherwise, it output to a value multiplied by a fixed scalar using the following formula:

\[
PReLU = \begin{cases} 
  \text{scalar} \times x, & x < 0, \\
  x, & x \geq 0 
\end{cases}
\]  

(11)

In a CNN, the activation function is responsible for transforming the summed weighted input from the node into the activation of the node.

**Classification**

In the second part of the proposed network, 1×1 convolutional layer was placed to minimize the number of channels and to reduce the computational cost. Then, the SoftMax layer obtained sparse feature sets to classify each pixel in the image corresponding to the classes defined, i.e. lesion and background. SoftMax assigns decimal probabilities to each class in the range between 0 and 1, and the sum of all the probabilities is equal to one.

**Model training**

The network is trained and optimized depending upon the loss function. The loss function measures the error between the prediction score \( P \) and target \( T \). The backpropagation algorithm then performs backward learning by transferring errors from the last layer to the first layer for updating weights again. The network continues to perform backward and forward learning until the loss function is minimum. In this paper, the weighted cross-entropy loss [29] function was employed to measure the error as:

\[
\text{Loss} = \frac{1}{N} \sum_{i=1}^{K} \sum_{n=1}^{N} w_{i} T_{n} \log(P_{n})
\]  

(12)

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 gradient descent algorithm updates the network weights and biases to reduce the loss value by applying small changes in the direction of optimization.

\[
\theta_{i+1} = \theta_{i} - \alpha \nabla L(\theta_{i})
\]  

(13)

Here \( i \) is the number of iterations, \( \alpha > 0 \) is the learning parameter, \( \theta \) is a parameter vector, and \( \nabla L(\theta_{i}) \) is the gradient of the loss function. At each iteration, the algorithm evaluates the gradient and update 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 follows:

\[
\theta_{i+1} = \theta_{i} - \alpha \nabla L(\theta_{i}) + \gamma(\theta_{i} - \theta_{i+1})
\]  

(14)
The momentum value was set manually as ‘0.9’ while training the network. The structural details of the designed network in terms of layers, kernel size, number of filters, and dilation rate are explained in Figure 5.

![Figure 5 Architecture details of the DilatedSkinNet](image)

**Evaluation Metrics**

The performance of the proposed DilatedSkinNet is evaluated quantitatively using performance metrics such as accuracy, Jaccard index, and BF-Score. The value of these parameters was calculated for the test dataset and is expected to be higher for good segmentation results. Accuracy (ACC) indicates the number of correctly identified pixels over the total number of pixels. The Jaccard index (JAC) is the ratio of correctly classified pixels to the total number of pixels that are assigned a class by the ground truth and the predictor. The BF-Score computes the Boundary F1 contour matching index between the predicted segmentation and the accurate segmentation in the ground truth.

\[
\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}
\]

\[
\text{JAC} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}
\]

\[
\text{BF-Score} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}
\]

The parameters TP, TN, FP, 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 as FN. On the other hand, non-lesion pixels if classified correctly then considered as TN otherwise, FP.
Results and Discussion

We trained the proposed network for 5494 images and tested on 1745 images of the ISIC 2016, ISIC 2017, and ISIC 2018 datasets. In Table 1, the results of the given model are demonstrated in comparison to other pre-trained segmentation networks under the same hyperparameter conditions. Our network showed better performance as compared to existing pre-trained frameworks such as UNet, SegNet, and DeepLabv3+. Our network achieved an average ACC 93.9%, JAC 86.5%, and BF-Score 64.6% on ISIC 2016, ACC 85.7%, JAC 77.0%, BF-Score 58.5% on ISIC 2017 and ACC 88.1%, JAC 81.2%, BF-Score 54.2% on ISIC 2018 dataset. In contrast to this, the JAC index computed by the UNet was 64.3%, 64.0%, and 71.2%, SegNet computed JAC of 85.8%, 72.0%, 72.4% and DeepLabv3+ gained JAC of 78%, 68% and 59% on ISIC 2016, 2017 and 2018 datasets respectively which is much lower than our proposed network. It can be analyzed that atrous dilated convolutions are efficient in extracting lesion information in comparison to previously proposed pre-trained networks for semantic segmentation. Moreover, the experimental results proved that a small network can efficiently calculate the expected information rather than designing a heavy network such as DeepLabv3+.

| Methods          | ISIC 2016 ACC | ISIC 2016 JAC | ISIC 2016 BF-Score | ISIC 2017 ACC | ISIC 2017 JAC | ISIC 2017 BF-Score | ISIC 2018 ACC | ISIC 2018 JAC | ISIC 2018 BF-Score |
|------------------|---------------|---------------|-------------------|---------------|---------------|-------------------|---------------|---------------|-------------------|
| DilatedSkinNet   | 0.939         | 0.865         | 0.646             | 0.957         | 0.770         | 0.585             | 0.981         | 0.812         | 0.542             |
| U-Net [20]       | 0.735         | 0.643         | 0.473             | 0.726         | 0.640         | 0.589             | 0.789         | 0.712         | 0.624             |
| SegNet [19]      | 0.933         | 0.858         | 0.665             | 0.831         | 0.720         | 0.644             | 0.875         | 0.724         | 0.700             |
| DeepLabv3+ [21]  | 0.874         | 0.780         | 0.614             | 0.790         | 0.685         | 0.593             | 0.699         | 0.593         | 0.561             |

To prove the robustness of the network, we compared our model with the top winners of the ISIC challenge of three years 2016, 2017, and 2018 as given in Table 2. It is presented that DilatedSkinNet outperformed all given studies by offering higher accuracy and Jaccard index. The important parameter used by the ISIC challenge to announce the winner was the Jaccard index for which DilatedSkinNet achieved the highest value as 86.5% on ISIC 2016, 77.0% on ISIC 2017, and 81.2% on ISIC 2018 dataset. However, it is important to note that the proposed model achieved slightly lower values for BF-Score. This parameter is used to measure correctly classified pixels at boundary contour of the lesion region.

| References       | ISIC 2016 JAC | ISIC 2017 References | JAC | ISIC 2018 References | JAC |
|------------------|---------------|----------------------|-----|----------------------|-----|
| DilatedSkinNet   | 0.865         | DilatedSkinNet       | 0.770 | DilatedSkinNet       | 0.812 |
| U. Sanchez [30]  | 0.843         | Y. Yuan [31]         | 0.765 | C. Qian [32]         | 0.802 |
| L. Yu [24]       | 0.829         | M. Berseth [33]      | 0.762 | H. Du [34]           | 0.799 |
| M. Rahman [35]   | 0.822         | L. Bi [36]           | 0.760 | Y. Ji [37]           | 0.799 |
| L. Huang [38]    | 0.811         | Menegola [39]        | 0.754 | Xue [40]             | 0.798 |

The graphs in Figure 6 illustrate the performance of the DilatedSkinNet on three datasets. It clearly shows that accuracy ratio increased gradually and reached above 90% whereas, loss values fell off against the iterations of the network. Figure 7 shows higher accuracy achieved by the DilatedSkinNet on ISIC 2017 dataset in comparison
to other pre-trained networks. The visual outputs of a few samples predicted by the proposed model are shown in Figure 8 that closely resembles to ground truth images.

Due to the structural similarities, it was a challenging task to obtain accurate border of the lesion region. There were some challenging samples containing a high level of noise such as dense hairlines, and dark ink projections as given in Figure 9 which
Figure 8 Exemplary pairs of the segmentation result using proposed DilatedSkinNet: (a) Original images. (b) Segmentation mask annotated by the dermatologist experts (c) lesion boundaries marked by the method, and (d) Segmented outputs.

were not properly segmented by the given method. The reason for this failure is that presence of noise elements that causes impediment in determining accurate region of interest. The network partially segmented these images but not as accurately as 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.

Figure 9 Failure cases of segmentation using the proposed DilatedSkinNet: (a) Original images. (b) Expected segmentation mask. (c) Segmented outputs, respectively.

Conclusion
The proposed segmentation method employed an atrous dilation CNN architecture for generating a larger field of view of the input vector. Also, the feature maps from the intermediate network layers were concatenated and the network achieved higher performance by minimizing the cross-entropy loss across mini-batches. The
network successfully extracted features from different types of dermoscopic skin lesions images and mapped their probabilities respective to each class to ensure visual consistency. We observed through experiments that the proposed network was successful in segmenting accurate lesion area that aids in developing a highly efficient pixel classifier. Moreover, the network was successful in segmenting the majority of challenging cases such as irregular boundaries, low contrast, and color illumination. However, few challenging samples especially containing dense hairlines were not accurately segmented, 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.

**Availability of data and materials**
The dataset used are benchmark data available publicly.

**Acknowledgements**
N/A.

**Funding**
N/A.

**Ethics approval and consent to participate**
This article does not contain any studies with human participants performed by any of the authors.

**Competing interests**
All authors have read and agreed to the published version of the manuscript.

**Consent for publication**
The authors declare no conflict of interest.

**Authors’ contributions**
Conceptualization: R.K and H.G; Methodology: R.K; Software: R.K; Validation: R.K, H.G and R.S; Data collection: R.K; Writing—original draft preparation: R.K; Writing—review and editing: H.G and R.S; Supervision: H.G and R.S.

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

**References**
1. Elwood, J.M., Jopson, J.: Melanoma and sun exposure: an overview of published studies. International journal of cancer 73(2), 198–203 (1997)
2. Bogo, F., Peruch, F., Fortina, A., Peserico, E., Celebi, M., Mendonça, 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)
3. Schadendorf, D., van Akkooi, A.C., Berking, C., Griewank, K.G., Gutzmer, R., Hauschild, A., Stang, A., Roesch, A., Ugurel, S.: Melanoma. The Lancet 392(10151), 971–984 (2018)
4. Massey University of New Zealand: Environmental Health Indicators New Zealand. https://www.ehinz.ac.nz/indicators/uv-exposure/melanoma/ Accessed 10 July 2020
5. The Skin Cancer Foundation: Skin cancer facts and statistics. https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/ Accessed Jan 2021
6. Ruiz, E.S., Morgan, F.C., Zigler, C.M., Besaw, R.J., Schmults, C.D.: Analysis of national skin cancer expenditures in the united states medicare population, 2013. Journal of the American Academy of Dermatology 80(1), 275–278 (2019)
7. Guy Jr, G.P., Machlin, S.R., Ekwueme, D.U., Yabroff, K.R.: Prevalence and costs of skin cancer treatment in the us, 2002-2006 and 2007-2011. American journal of preventive medicine 48(2), 183–187 (2015)
8. 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). arXiv preprint arXiv:1605.01397 (2016)
9. Codella, N.C., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kolloo, 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 (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 168–172 (2018). IEEE
10. Codella, N., Rotenberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kolloo, A., Liopyris, K., Marchetti, M., et al.: Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). arXiv preprint arXiv:1902.03368 (2019)
11. Tschandl, P., Rosendahl, C., Kottler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data 5, 180161 (2018) (2018)
12. Lee, I., Xu, D., Anthony, B.: Hair segmentation using adaptive threshold from edge and branch length measures. Computers in biology and medicine 89, 314–324 (2017)
13. Garnavi, R., Aldeen, M., Celebi, M.E., Varigos, G., Finch, S.: Border detection in dermoscopy images using hybrid thresholding on optimized color channels. Computerized Medical Imaging and Graphics 35(2), 105–115 (2011)
14. Aitnouri, E., Ouali, M.: Performance evaluation of clustering techniques for image segmentation. Computer Science Journal of Moldova 54(3), 271–302 (2010)
15. Vasconcelos, F.F.X., Medeiros, A.G., Peixoto, S.A., Reboucas Filho, P.P.: Automatic skin lesions segmentation based on a new morphological approach via geodesic active contour. Cognitive Systems Research 55, 44–59 (2019)
16. Moghaddam, M.J., Soltanian-Zadeh, H.: Medical image segmentation using artificial neural networks. Artificial Neural Networks-Methodological Advances and Biomedical Applications, 121–138 (2011)
17. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
18. Hong, S., Noh, H., Han, B.: Decoupled deep neural network for semi-supervised semantic segmentation. arXiv preprint arXiv:1506.04924 (2015)
19. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 39(12), 2481–2495 (2017)
20. 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, pp. 234–241 (2015). Springer
21. 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 transactions on pattern analysis and machine intelligence 40(4), 834–848 (2017)
22. Kawahara, J., BenTaieb, A., Hamanenn, G.: Deep features to classify skin lesions. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1397–1400 (2016). IEEE
23. 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 recognition 85, 78–89 (2019)
24. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.-A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE transactions on medical imaging 36(4), 994–1004 (2016)
25. Al-Masni, M.A., Al-Antari, M.A., Choi, M.-T., Han, S.-M., Kim, T.-S.: Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. Computer methods and programs in biomedicine 162, 221–231 (2018)
26. Vesal, S., Ravikumar, N., Maier, A.: Skinnet: A deep learning framework for skin lesion segmentation. In: 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), pp. 1–3 (2018). IEEE
27. Yuan, Y., Chao, M., Lo, Y.-C.: Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE transactions on medical imaging 36(9), 1876–1886 (2017)
28. Chen, L.-C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017)
29. Bishop, C.M.: Pattern Recognition and Machine Learning. springer, New York (2006)
30. U. Sanchez: ISBI 2016 challenge results (2016). https://challenge.kitware.com/submission/56fe2b60cad3a55ecee68cf74
31. Yuan, Y.: Automatic skin lesion segmentation with fully convolutional-deconvolutional networks. arXiv preprint arXiv:1703.05165 (2017)
32. 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. arXiv preprint arXiv:1809.03917 (2018)
33. Berseth, M.: ISIC 2017 - skin lesion analysis towards melanoma detection. CoRR abs/1703.00523 (2017)
34. Hao Du, Jeon Young Seok, Dianwen Ng Ngiam, Kee Yuan, Mengling Feng: Team HolidayBurned at ISIC CHALLENGE 2018 (2018). https://challenge.isic-archive.com/leaderboards/2018
35. M. Rahman: ISBI 2016 challenge results (2016). https://challenge.kitware.com/submission/56fbfa1bcad3a54f8bb809bf
36. Bi, L., Kim, J., Ahn, E., Feng, D.: Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks. arXiv preprint arXiv:1703.04197 (2017)
37. Ji, Y., Li, X., Zhang, G., Lin, D., Chen, H.: Automatic skin lesion segmentation by feature aggregation convolutional neural network. Technical report (2018)
38. Huang, L., Zhao, Y.-g., Yang, T.-j.: Skin lesion segmentation using object scale-oriented fully convolutional neural networks. Signal, Image and Video Processing 13(3), 431–438 (2019)
39. Menegola, A., Tavares, J., Fornaciali, M., Li, L.T., Avila, S., Valle, E.: Recod titans at isic challenge 2017. arXiv preprint arXiv:1703.04819 (2017)
40. Yuan Xue, Lijun Gong, Wei Peng, Xiaolei Huang, Yefeng Zheng: Automatic Skin Lesion Analysis with Deep Networks (2018). https://challenge.isic-archive.com/leaderboards/2018