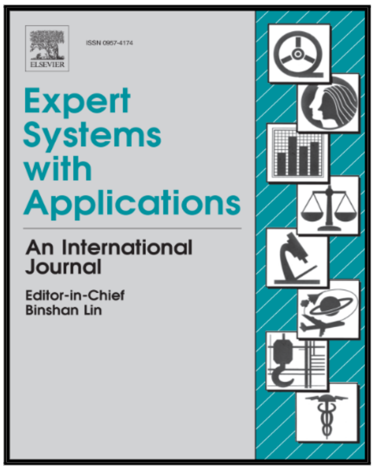
Journal Pre-proof



Transform domain representation-driven convolutional neural networks for skin lesion segmentation

Mansoureh Pezhman Pour , Huseyin Seker

PII: S0957-4174(19)30846-2
DOI: https://doi.org/10.1016/j.eswa.2019.113129
Reference: ESWA 113129

To appear in: *Expert Systems With Applications*

Received date: 10 August 2019
Revised date: 2 November 2019
Accepted date: 8 December 2019

Please cite this article as: Mansoureh Pezhman Pour , Huseyin Seker , Transform domain representation-driven convolutional neural networks for skin lesion segmentation, *Expert Systems With Applications* (2019), doi: https://doi.org/10.1016/j.eswa.2019.113129

Highlights

- An automatic skin lesion segmentation based on deep convolution network designed.
- Image representations from transform domain used to improve the performance.
- Despite of small dataset, pre-processing and excessive data augmentation not applied.
- The analysis proved the efficiency of the proposed method on tested images.

Transform domain representation-driven convolutional neural networks for skin lesion segmentation

Mansoureh Pezhman Pour[a*] , Huseyin Seker[b]

[a]*Department of Computer and Information Sciences, Northumbria University, Newcastle Upon Tyne, NE1 8ST, UK*
[b]*Department of Computer and Information Sciences, Northumbria University, Newcastle Upon Tyne, NE1 8ST, UK*

Abstract

Automated diagnosis systems provide a huge improvement in early detection of skin cancer, and consequently, contribute to successful treatment. Recent research on convolutional neural network has achieved enormous success in segmentation and object detection tasks. However, these networks require large amount of data that is a big challenge in medical domain where often have insufficient data and even a pretrained model on medical images can be hardly found. Lesion segmentation as the initial step of skin cancer analysis remains a challenging issue since datasets are small and include a variety of images in terms of light, color, scale, and marks which have led researchers to use extensive augmentation and preprocessing techniques or fine tuning the network with a pretrained model on irrelevant images. A segmentation model based on convolutional neural networks is proposed in this study for the tasks of skin lesion segmentation and dermoscopic feature segmentation. The network is trained from scratch and despite the small size of datasets neither excessive data augmentation nor any preprocessing to remove artifacts or enhance the images are applied. Alternatively, we investigated incorporating image representations of the transform domain to the convolutional neural network and compared to a model with more convolutional layers that resulted in 6% higher Jaccard index and has shorter training time. The model improved by applying CIELAB color space and the performance of the final proposed architecture is evaluated on publicly available datasets from ISBI challenges in 2016 and 2017. The proposed model has resulted in an improvement of as much as 7% for the segmentation metrics and 17% for the feature segmentation, which demonstrates the robustness of this unique hybrid framework and its future applications as well as further improvement.
Keywords: Convolutional Neural Network; Dermoscopic features; Melanoma; Skin lesion segmentation;Transform domain

1. Introduction

Skin cancer is a prevalent kind of cancer worldwide with fast increment in incidence and number of deaths over the past decade as cancer research statistics reported (Siegel, Miller, & Jemal, 2018). Nevertheless, there is a high chance of cure if the cancer is diagnosed in a primary stage before other parts of the body get invaded. Dermoscopic imaging has significantly assisted dermatologists to detect malignant melanoma that is the deadliest type of skin cancer. However, expert clinicians are still needed to distinguish the disease. Segmentation has an important role in cancer diagnosis as primary stage to detect the affected area. Research towards automated computer-based detection systems for melanoma cancer have increased in the past few years to assist clinicians, lessen the workload, keep monitoring high risk patients and more, to reduce the costs of diagnosis and treatment (Alamdari, et al., 2017). Moreover, such algorithms can be used to improve embedded systems, robots or even mobile software to make an easy user interface as part of an automated diagnosis system.

* Corresponding author,
E-mail addresses: mansoureh.pezhmanpour@northumbria.ac.uk, Huseyin.seker@northumbria.ac.uk.

Lesion segmentation is the initial step in a melanoma diagnosis system. It aims to separate the relevant pixels to melanoma tumors in medical images. The lesion segmentation is still a challenging task due to several problems in skin images such as color and light variations, small affected area, dark corners, low contrast, and artifacts like hair and ruler marks as presented in Figure 1. Moreover, the lack of extensive public datasets impeded the development of computer aided systems for melanoma detection over the past decades. Recently, ISIC (International Skin Imaging Collaboration) has provided public collections of dermoscopic images of skin lesions, and ISBI challenges (IEEE International Symposium on Biomedical Imaging) have been held to improve the skin cancer diagnosis (Gutman, et al., 2016), (Codella, et al., 2018).

Deep convolutional neural networks (CNN) have attained great success in computer vision and machine learning surpassing conventional methods in several challenges (Cireşan, Giusti, Gambardella, & Schmidhuber, 2013), (He, Zhang, Ren, & Sun, 2016). Compared to a shallow network, CNN learns more complex features with more layers but requires a large amount of data for training that is a critical issue in medical imaging. Dataset enlargement by adding other available datasets is generally considered by researchers, whereas preparing labeled data is still time consuming and expensive. Common classical augmentation techniques such as flipping, rotation, and scaling are often applied to produce adequate information to feed the deep neural network (Harangi, 2018), (Hussain, Gimenez, Yi, & Rubin, 2017), (Kwasigroch, Mikołajczyk, & Grochowski, 2017). However, augmentation techniques are specific to dataset and require attention to not lose information or increase irrelevant data. Thus, augmentation techniques have expanded into a field widely used in deep learning particularly medical analysis. It has become an active academic area that various research has been conducted to develop methods to generate data (Zhang, Cisse, Dauphin, & Lopez-Paz, 2017), (Liang, Yang, Zhang, & Yang, 2018), (Frid-Adar, et al., 2018).

In this study, the dataset provided by ISBI challenge is used and a novel CNN based system is proposed for the tasks of lesion segmentation and dermoscopic feature segmentation. The final model is efficient with less computational load compared to models proposed in literature since it is not very deep (15 conv layer in encoder part). Also, preprocessing or excessive data augmentation is not used. Although using a CNN model with more convolutional layer often lead to extract more complex features and higher accuracy but this may not be applicable when dataset is scarce. Instead, we proposed to improve the model by incorporating multi scale and multi direction representations of image from transform domain and the results of this model showed higher performance and lower training time compared to a deeper model. An efficient comprehension of input images is combined to the network and increased the depth of convolution layers that leads to accurate detection on such a small dataset. The only augmentation technique that we used is flipping to increase the dataset size for the task of lesion segmentation and particularly to balance the data for the task of dermoscopic feature segmentation. Another common issue with applying deep architectures on medical data is that training the network from scratch is not feasible while the dataset is scarce. Thus, a pretrained model could be used on which the network is fine-tuned. While there are models pretrained on millions of natural images such as ImageNet (Deng, et al., 2009), a pretrained model on relevant medical images can be hardly found. Although, some research works have shown sufficient results on using a model that is pretrained on a large irrelevant dataset (Bar, Diamant, Wolf, & Greenspan, 2015), (Pour, Seker, & Shao, 2017), issues such as local representation on some layers (Yosinski, Clune, Nguyen, Fuchs, & Lipson, 2015) and the limitation to modify the deep network disparate from pretrained model motivate researchers to not limit the deep architecture to a pretrained model on irrelevant data. Moreover, when dataset of pretrained model is very different to our dataset, only the general features of first layers maybe helpful and rest of the network needs to be retrained. The proposed model in this study is trained from scratch and achieves better results compared to the study that used similar architecture to FCN, used pretrained model and augmented dataset 8times (Pour, Seker, & Shao, 2017). The model outperforms the researches in the literature which also have more complex architecture or composed of very deep network structure. In addition, the proposed model indicates notable performance in images that contain noisy artifacts such as hair, ruler, or bubble. Hence, we did not apply common preprocessing techniques to remove these marks.

The paper is organized into five parts. In section 2, a brief review on recent segmentation techniques is provided. Section 3 details the proposed method followed by experiments and results in part 4. Results are discussed in section 5 and the paper is concluded in section 6.

2. Related Work

Various algorithms of image analysis have been proposed to assist clinicians in early diagnosis of skin cancer. Dermoscopic feature based algorithms such as ABCD rule that includes Asymmetry, Border, Color, and Dermoscopic structure (Nachbar, et al., 1994), and CASH (Color, Architecture, Symmetry, and Homogeneity) (Henning, et al., 2007) are primary methods that have been used for many years. Moreover, common segmentation techniques including edge or region-based methods (Wong, Scharcanski, & Fieguth, 2011), (Tajeddin & Asl, 2018), (Jaisakthi, Mirunalini, & Aravindan, 2018), (Lau, et al., 2018), thresholding techniques (Zortea, Flores, & Scharcanski, 2017), and techniques based on features from transform domain such as wavelet and Fourier (Garnavi, Aldeen, & Bailey, 2012) have been developed for the task of skin lesion segmentation. Recently deep convolutional neural network has surpassed traditional methods in many computer vision tasks. Since CNN originally was designed for classification task, various recent researches focused on proposing models to adopt CNN for segmentation task too (Long, Shelhamer, & Darrell, 2015), (Lin, Milan, Shen, & Reid, 2017), (Jégou, Drozdzal, Vazquez, Romero, & Bengio, 2017). In the following, background on skin lesion segmentation categorized in traditional methods and deep learning-based algorithms is explained.

2.1. Traditional segmentation methods

A method including edge and region-based algorithm suggested in (Jaisakthi, Mirunalini, & Aravindan, 2018). Illumination enhancement and artifact removal such as hair and air bubbles constitute the initial stage as pre-processing phase. Grabcut algorithm that uses both edge and boundaries information has been applied for segmentation followed by further stages including k-means clustering and flood-fill technique to segment the lesion area with enhanced boundaries. Compared to the winners of ISIC 2017 challenge, this method showed lower Jaccard index. Another popular region-based method termed watershed has been developed in various algorithms for medical segmentation tasks. (Masoumi, Behrad , Pourmina, & Roosta, 2012) proposed using watershed algorithm and MLP neural network for feature extraction in an iterative process for liver segmentation. The extracted features from both techniques were compared and the error computed in each iteration used to adjust the required parameters of the algorithm sequentially. In addition, morphological smoothing, Gaussian filtering and morphological gradients used as preprocessing stage but no postprocessing conducted.

Another popular method that has been widely developed for image segmentation is active contour composed of deformable contours that adjust to variety of shapes. The method includes an energy maximization procedure build on region or edge based models and have been employed for segmenting several medical images such as CT, USG and MRI of different organs in the body (Ciecholewski, 2016), (Riaz, Naeem, Nawaz, & Coimbra, 2018). In research (Riaz, Naeem, Nawaz, & Coimbra, 2018), to generate the initial curve, adaptive thresholding was applied and an optimization problem was proposed to maximize the Kullback–Leibler divergence of gray level distribution between the background and the lesion.

A recent research proposed saliency map generated by improved discriminative regional feature integration (mDRFI) (Jahanifar, Tajeddin, & Gooya, 2018). This method also composed of multiple stages including pre-processing such as colour constancy and hair removal, generating an initial mask by thresholding saliency map based on DRFI method and final mask formation using distance regularized level set evolution (DRLSE) framework. They have extended regional property descriptors and proposed a pseudo-background region to improve DRFI method, but the result was still lower than high ranked papers of both ISBI 2016 and 2017 challenge.

(Tajeddin & Asl, 2018) have proposed adding new texture features of peripheral regions for classification and a segmentation technique composed of estimating initial contour and propagating it with an iterative process based on dual-component speed function. Otsu's method following by morphological process conducted to generate the threshold initial mask. They used a level set framework and proposed two component speed function for the image gradient and the color probability distribution of pixels to generate the final mask. General shape features based on common ABCD features, colour-based features and texture related features from luminance channel of La*b* colour space implied for feature extraction phase. They also proposed a textural feature set from peripheral region that is based on masks from segmentation phase. A variety of pre-processing methods have been applied to remove hair, marks and eliminate dark corners, correct image illumination and crop images regarding to the masks. This segmentation method ranked 5th in ISBI challenge 2016.

Techniques based on superpixels also have been extensively used in medical image segmentation. Superpixel is an efficient method to segment images by partitioning the image into groups of connected pixels that have similarities (Nguyen, Benameur, Mignotte, & Lavoie, 2018), (Navarro, Escudero Vinolo, & Bescos, 2018). In the research (Navarro, Escudero Vinolo, & Bescos, 2018), the common SLIC (simple linear iterative clustering) method is improved with focusing on segmenting the ROI precisely instead of segmenting whole image accurately. Firstly, feature points are detected in the image by SIFT operation and then Gaussian distribution applied to place initial centres followed by applying SLIC to these centres. The result showed marginally higher Jaccard index compared to the top results of ISIC 2017 challenge.

2.2. Deep learning-based models for task of segmentation

all traditional methods that mentioned in previous section, are composed of multiple stages such as pre-processing, initialization, edge/region extraction or various techniques for feature extraction while deep learning methods benefit of receiving the input as raw image and generate the output via an end to end learning process. Another drawback of traditional models is that discriminative features play important role in success of these models. Extracting effective features not only is a difficult and time-consuming work but also demands specialist's attempt and experience. Variety of algorithms have been suggested to extract features regarding to the image structure of medical images, but these algorithms mostly deal with particular features of image that may not work for all kinds of images. For instance, low contrast between lesion and the background would not contribute to an accurate thresholding method, weak or noisy edges deter the performance of edge-based segmentation models and active contours build upon an initial contour that may limit the efficiency of model. However, the deep convolution network has advantage of learning the features automatically from general features such as edges and lines extracted in first convolutional layers to high level features like shapes extracted in higher layers. In this paper, we used a deep convolutional neural network and recent CNN based algorithms for task of segmentation are briefly reviewed in the following.

Deep learning algorithms have shown remarkable progress in various computer vision tasks. Notably, Convolutional Neural Network (CNN) has outperformed conventional methods in several pattern recognition and machine learning domains. CNN is introduced as a deep neural network architecture composed of more layers in comparison to shallow conventional neural networks. More convolution layers enable the network to learn more complex features. CNN was originally proposed for the task of classification and extensive research was conducted to design efficient deep architectures. Recently, several studies modified the deep networks that are designed for classification problems to be applied to object detection and segmentation tasks. Early proposed CNN models for segmentation were based on classifying superpixels or region surrounding a pixel (Farabet, Couprie, Najman, & LeCun, 2013), (Ciresan, Giusti, Gambardella, & Schmidhuber, 2012) .

In 2015, fully convolutional neural networks (FCN) have been designed to adapt the classification model to perform segmentation. An end-to-end pixelwise learning architecture was proposed in which fully connected layers are transformed to convolution layers so the network have spatial output maps (Long, Shelhamer, & Darrell, 2015). In FCN, the fractionally strided convolution is introduced as upsampling, also called deconvolution. Moreover, the idea of skip connections was introduced in (Long, Shelhamer, & Darrell, 2015) that fuses coarse, high layer information from downsampling path with fine, low layer information in the corresponding upsampling path. Various subsequent researches were conducted to improve FCN for different segmentation problems. Unet is a popular model with a symmetric encoder-decoder architecture that includes deconvolution layers with larger number of feature channels (compared to FCN) each followed by concatenation of feature maps from the corresponding layer in contraction path (Ronneberger, Fischer, & Brox, 2015).

Most recent proposed CNN based models to deal with segmentation are encoder-decoder architectures in which the encoder part is based on a further improved classification model and decoder part that generates a classified pixel wise high-resolution image from the low-resolution output of the encoder. A Residual Network (ResNet) is proposed to efficiently increase the depth of a convolution network by introducing shortcut connections of identity mapping that connects the output of each layer to a higher layer (He, Zhang, Ren, & Sun, 2016). The encoder part of RefineNet (Lin, Milan, Shen, & Reid, 2017) as a segmentation model, is based on this model and the decoder contains multi-level RefineNet blocks that fuse the features received from the encoder as well as the features from the previous RefineNet block. In (Jégou, Drozdzal, Vazquez, Romero, & Bengio, 2017), a segmentation model termed Fully Convolutional DenseNets was designed, it was built on Dense Convolutional Network which is a CNN based model composed of blocks with densely connected layers. Deconvolution layer, skip connections, and dense blocks constituted the upsampling path.

Recently, deep architectures have been applied to skin lesion analysis (Yu, Chen, Dou, Qin, & Heng, 2017), (Nasr-Esfahani, et al., 2017), (Bi , et al., 2017), (Al-Masni, Al-antari, Choi, Han, & Kim, 2018). A very deep convolutional network composed of fifty layers used in (Yu, Chen, Dou, Qin, & Heng, 2017) and residual learning applied to deal with overfitting. They proposed a fully convolutional residual network (FCRN) for task of segmentation and their experiments ranked second in segmentation task of ISBI 2016 challenge. (Bi , et al., 2017) designed a model based on fully convolution network that contains FCNs in multi stage structure (mFCN). In each stage, the FCN receives inputs including the original input image and the estimated output of previous stage. They also integrate the segmentation results of all stages in a parallel way and their result slightly outperformed the best results of the challenge. The final method proposed in our paper composed of 15 convolution layers that is significantly less complicated compared to FCRN with 50 layers or mFCN that contains FCN architecture in each stage. We investigated to improve the FCN model for task of lesion segmentation but not designing a very deep network due to overfitting problem. In many previous researches on skin lesion segmentation, considering data scarcity, massive data augmentation techniques were applied, or the models were fine-tuned on a model which was pretrained with irrelevant data i.e. natural images not medical images. Alternatively, we proposed a CNN based model for tasks of lesion segmentation and dermoscopic attribute detection that does not employ any preprocessing technique or excessive data augmentation. We also investigated training the deep network from scratch albeit the size of dataset is small.

3. Method

An overview of the proposed segmentation architecture for task of lesion segmentation is depicted in Figure 2. Details of the method including pre-processing, architecture of segmentation model, and post processing are explained in the following.

3.1. Pre/Post Processing

Images from surface of the skin mostly involve dark corners, hairs, ruler marks, variation in color, and uneven illumination. The idea of removing noisy artifacts has extensively been investigated as an essential preprocessing task to enhance the quality of images (Oliveira, et al., 2016). In this study, we do not apply any preprocessing procedure for task of lesion segmentation, but we slightly increase dataset size by flipping images. The effect of dark corners on generated masks is compensated in post processing and the model performs very efficiently on images with hairs and other noisy artifacts. For task of lesion dermoscopic feature segmentation, the images are firstly cropped by applying bounding box and using masks from previous task, then flipping is conducted to balance the dataset. In post processing phase, the masks are firstly resized to the original size, then thresholding and morphological dilation is used to extract the objects in a predicted mask, choose the object closer to center of image, and finally remove unwanted components such as corner's effect, and cover the small holes (Gonzalez Rafael C.and Woods, 2007).

3.2. Segmentation Architecture

Considering the lack of enough data for task of lesion segmentation, most studies based on deep neural network have used transfer learning or they have conducted various augmentation methods to increase the dataset size. In this research, we trained the network from scratch and improved the performance by injecting features from transform domain to the network. In this way, the network will learn not only from raw images but also from image representations from transform domain.

3.2.1. Contourlet Transformation domain

Contourlet transform has proved excellent performance in computer vision problems (Do & Vetterli, 2005). In this research, contourlet transform is used to provide multiscale and multidirectional image representations to be inserted to the convolutional neural network which is trained with limited data. In contourlet transform, multiscale and directional decomposition is achieved by applying a combination of a Laplacian Pyramid (LP) and a Directional Filter Bank (DFB) (Do & Vetterli, 2005). The architecture is illustrated in Figure 3. Laplacian pyramid generates a down sampled low pass sketch of an image and the difference between origin and the prediction that produces band pass image. Directional filter bank receives the band pass image as input and produces final directional decomposition. Consecutively, Laplacian pyramid applies to the low pass image and the process is repeated to reach the desired level of decomposition. As demonstrated in Figure 2, the multidirectional representations generated by directional filter bank in each scale is considered to concatenate with equivalent pooling layer.

3.2.2. Convolutional Neural Network

Various layers including convolution, pooling, and fully connected layers stack to build a typical convolutional neural network. Compared to conventional neural network, convolutional neural network has the advantage of receiving input layer in a shape of 2D information layer and a neuron within any layer is connected to a small region of the previous layer called receptive field where the filter applies, and convolution is computed. Convolution layer is composed of a set of learnable filters that slide over the image and generate feature maps. The output of convolution operation for position (i,j) in a feature map is:

$$\sum_{m=0}^{k-1}\sum_{n=0}^{k-1}x_{si+m,sj+n}K_{m,n}+b, \quad (1)$$

where x is the location in preceding layer, K represents the kernel and k the size of kernel. The stride is the number of pixels that the filter skips while sliding over the image. Rectifier Linear Unit (ReLU) is usually used as activation function and Pooling layer reduces the spatial size of representation and thus parameters. The model we used is based on fully convolutional network (FCN) that is composed of an encoder and a decoder path (Long, Shelhamer, & Darrell, 2015). FCN has been designed to use the convolutional neural network of the classification task as supervised pretraining, and to fine tune the fully convolution network to perform the task of segmentation. in (Long, Shelhamer, & Darrell, 2015), they proposed converting fully connected layers to convolution layers besides adding the feature map from lower layers in encoder path to the corresponding layer in decoder path that termed skip connections. Unet (Ronneberger, Fischer, & Brox, 2015) modified the FCN by using the multitude of feature channels in expanding path compared to FCN that the number of kernels in upsampling path is limited to number of classes. Moreover, learnable filters are utilized and skip connections are concatenated to corresponding upsampling layer instead of fusion in FCN.

3.2.3. Lesion Segmentation

A basic model inspired by Unet and FCN is considered and gradually improved by injecting features from the transform domain and adding CIElab color model of input images. The initial model is composed of encoder and decoder parts. The encoder part of this model includes a series of convolution layers followed by max pooling layer. ReLUs are applied after convolution layers to accelerate the training (Krizhevsky, Sutskever, & Hinton, 2012). The number of feature channels start from 16 in the first convolution layer and duplicate in each subsequent convolution layer. To deal with overfitting, drop out layer is applied for the last two convolutional layers (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). In decoder side, a series of deconvolution layers operate as learnable upsampling layers. The number of kernels in decoder part is considered equal to the number of classes that is two. This model is trained from scratch and the learnable weights in both convolutional layers and deconvolutional layers are initialized with Xavier filters. As recommended in FCN, feature maps from pooling layers in encoder path are connected to later deconvolution layers in decoder path but we used concatenation like Unet. The number of deconvolution layers are also different from Unet and FCN, we have 6 upsampling layers to finally get the output with same resolution of input image. Moreover, we started from 16 as the number of filters for first convolution layers compared to 64 in FCN and Unet, to decrease the depth and number of training parameters since our data set is small. To improve the performance, representations from contourlet transform in four levels are concatenated to pooling layers that makes model2. Discrete contourlet transform with 4 decomposition levels and 4 directions in each level are applied to three color channels of input images, that totally provided 12 representations in each level. In another experiment, we investigated making model 1 a deeper model (increasing number of convolution layers from 7 in encoder part to 15) instead of applying image representations of transform domain and this model named model 3. We improved the architecture of model 2 by using a deeper model termed model 4 (has 15 convolution layers in encoder part instead of 7 convolution layers) and due to high correlation between the red, green and blue colors in RGB, we additionally applied CIELAB color channels as input of final model (model 4 with CIELAB). CIELAB contains lightness component (L), and two-color components (A and B) and has the advantage of being device independent compared to RGB. The 4 models are shortly defined as:

-Model1: the model with 7 convolution layers in encoder part

-Model2: the model with 7 convolution layers in encoder part and integrated with representations of transform domain

-Model3: the model with 15 convolution layers in encoder part

-Model4: the model with 15 convolution layers in encoder part and integrated with representations of transform domain

3.2.4. Lesion Dermoscopic Feature Segmentation

Segmentation of dermoscopic features including globules and streaks helps clinicians to diagnose melanoma from benign skin lesions. The goal of this task is to automatically generate two masks for each lesion that reveal the location of streaks and globules. For this task, the segmentation model contains two parts including encoder and decoder. The encoder part is the same as task of lesion border segmentation, thus, we considered transfer learning from task1. Encoder path consists of two parts for globules and streaks localization that each concludes two convolution layers and four learnable upsampling layers, and so two loss are added up to generate the final loss of the network. The main issue with training the model for this task is the unbalanced data. Nearly half of images in the dataset do not contain any dermoscopic feature and detection of empty masks can improve the performance (Chen, et al., 2018). So, we added classification to the network and the corresponding loss added to the segmentation loss as presented in (Chen, et al., 2018).

Moreover, among those which hold dermoscopic feature, the number of pixels that belong to the classes of globules or streaks are far fewer than background pixels and each skin image does not necessarily contain both streaks and globules features. Almost 42 percent of images contain pixels of globules while the number of images which involve streaks are limited to less than 8 percent of dataset. In (Pour, Seker, & Shao, 2017) a bounding box was applied to separate the lesion region by using ground truth images. Thus, the network will look to a larger region of interest as input and the number of background pixels reduces that lessen class imbalance. Images are cropped by a factor of 1.1 because we found out that in a few cases the streak and globule pixels are located slightly outside of the border of the lesion from the mask provided in section 3.2.3. Accordingly, we used the output masks generated by our lesion segmentation from task1 to crop the test images before entering the network. To deal with the imbalanced data, flipping over vertical axes horizontal axes and both vertical/horizontal are conducted on images which contain streaks. So, the number of images with streaks increased to 23 percent and images with globules increased to 48 percent as most images with streaks contain globules too.

4. Experimental Results

To validate the model, we applied our method to two databases provided by Skin Lesion Analysis towards Melanoma Detection challenges (ISIC 2016 and 2017) (Gutman, et al., 2016), (Codella, et al., 2018). The performance of the proposed method is compared to the results achieved by the winners of both challenges. Further information on datasets, implementation details and results are presented in the following.

4.1. Data Preparation

The challenges of "Skin Lesion Analysis towards Melanoma Detection" provided a dataset of annotated skin lesion images from the ISIC Archive. Two publicly available datasets from ISIC 2016 and ISIC 2017 challenges are utilized in this research for task of lesion segmentation. ISIC 2016 provides dataset for task of Lesion dermoscopic feature segmentation that is practiced in this study as well. Both datasets contain dermoscopic lesion images in JPEG format and corresponding masks in PNG format. Each pixel in the mask is either 0 or 255. 0 represents the background of the image, or areas outside the lesion, and 255 represents the foreground of the image, or areas inside the lesion.

ISIC 2016 dataset is randomly partitioned into both a training and test set, with 900 colored images in the training set and 379 images in the test set for task of lesion segmentation. The number of images provided as training data and test data are 2000 and 600 in ISIC 2017 for this task. Moreover, 807 lesion images, each paired with two binary masks that present locations of the globules and streaks dermoscopic features are provided in ISIC 2016 for task of lesion dermoscopic feature segmentation. Also, 335 images along masks are supplied as test data.

The proposed model takes advantage of not using common methods of preprocessing such as removing hair or artifacts in the images. Instead, raw images are the input of networks in all experiments of this research. In terms of data augmentation, we only used flipping. Specifically, the images and corresponding ground truths are flipped vertically to expand training data for task of lesion segmentation and flipping vertically and horizontally to balance the dataset for task of dermoscopic feature segmentation.

4.2. Implementation

Initially, a basic architecture composed of 7 convolution layers followed by 6 deconvolution layers, named Model1, is considered in this study, and a series of comparative experiments is conducted to improve the results. The training is performed using stochastic gradient descent (SGD), such that weight decay and momentum are set at 0.0005 and 0.99, respectively. An initial learning rate of 0.001 is considered which is then reduced manually by a factor of 10 when the error reaches to plateau. In experiments where the convolutional neural network is trained with Adam optimization method, parameters are determined to $\alpha = 0.001$, $\beta1 = 0.9$ and $\beta2 = 0.999$ as recommended in (Kingma & Ba, 2014) for being good default settings for machine learning problems. To ensure a fair comparison, similar values for parameters such as filter size, stride and learning rate are considered in all models. The hyperparameters of the proposed network are provided in Table 1. For the number of epochs, we set the maximum to 10000 but we used early stopping method to deal with overfitting. The Caffe framework with GPU GeForce GTX TITANX has been used to implement the deep architectures. Caffe is a deep learning framework that is developed by the Berkeley Vision and Learning Center (BVLC) and community contributors. It was released under the BSD 2Clauselicense (Jia, et al., 2014).

4.3. Evaluation Metrics

For the sake of consistency in the literature, the following five different metrics have been used to compare the performance of the methods studied in the paper;

$$SE = \frac{TP}{(TP+FN)} \quad (2)$$

$$AC = \frac{TP+TN}{(TP+FP+TN+FN)} \quad (3)$$

$$SP = \frac{TN}{(TN+FP)} \quad (4)$$

$$DI = \frac{2\times TP}{2\times TP+FN+FP} \quad (5)$$

$$JA = \frac{TP}{(TP+FN+FP)} \quad (6)$$

where SE, AC, SP, DI and JA are sensitivity, accuracy, specificity, dice and Jaccard index respectively. Also, TP, TN, FP, FN, stand for true positive, true negative, false positive, and false negatives respectively and performance metrics are computed at the level of single pixels.

4.4. Results for Task of Lesion segmentation

Since the database is scarce and many parameters in a deep neural network stand on heuristics, we embarked on such a simple model to prevent early overfitting and find appropriate weight initialization and optimization algorithm to train the network from scratch. The model is gradually improved, and the results reported show the compromise made between computational complexity and accuracy. Model 1 includes a series of convolution layers followed by pooling layer that takes input image with size of 3*698*698 in first layer ended to feature representation of size 512*16*16 in last convolution layer. We empirically found that the network converges more slowly when filters are initialized with a Gaussian distribution rather than Xavier. Therefore, all convolution layers are initialized with a Xavier distribution (Glorot & Bengio, 2010), and learned from scratch in all subsequent experiments. Rectified Linear Unit is used after each convolution layer along with normalization layer. Moreover, drop out layers are applied after both latest layers in decoder i.e. conv6 and conv7 (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). Inspired from (Long, Shelhamer, & Darrell, 2015), skip connections are used to concatenate information of primary layers to later deconvolution layers. In addition, we applied early stopping. The results derived from this model are provided in Table 2.

We investigated further improvement by considering two options. The first option consists of combining transform domain representations of input images into convolutional layers, we refer to this as model 2. The rationale here is that integrating these proper features might lead convolutional layers to understand the input better. An alternative common option is to make the network deeper to learn more complex representations. In model 2, the contourlet transform in four levels and four directions is applied to different color channels of images which provides 12 images for each level. These representations have the same size of the outputs of pooling layers with which they are concatenated in various levels and increased the depth. The evaluation metrics for this architecture in table 2 demonstrate improvement for all metrics with a significant raise of 12 percent for Jaccard index compared to model 1. Figure 1 shows a sample image with relevant groundtruth and representations of the image derived from contourlet transform that is injected to the network. The consequence of making the network deeper instead of applying contourlet coefficients explored in model 3. In this model, we extend convolution layers in encoder part from 7 to 15 layers by adding one convolution layer after conv1 and conv2, and adding 2 conv layer after conv3 to conv5. The performance metrics show 6 percent increase in Jaccard index compared to model 1, but still lower performance than model 2.

Moreover, to investigate the system performance in terms of training time, forward and backward execution time averaged over 50 iterations per image are reported in Table 3. The model to which transform domain features are added (model 2) shows higher performance besides less inference time compared to model 3 that is made by more convolution layers. Figure 5 demonstrates the training error curves. The network converges faster if either feature from contourlet transform are added, or the number of convolution layers increased. As extending convolution layers also yields higher performance, we proposed the final model (Model 4) by increasing convolution layers in the model 2. In this architecture 15 convolution and 6 deconvolution layers are applied. Furthermore, performing optimization with Adam led to significant reduction in convergence time compared to SGD in the deeper architecture apart from slightly improving the results. Training error curves in Figure 5 confirms that Model 4 with Adam optimization converges two times faster.

As the deeper model gets more complex, we also increased the training data by flipping to deal with overfitting. At this point, transformed images from RGB to CIELAB color space also added to the network that concatenates with the input. The results are also compared to the model in (Pour, Seker, & Shao, 2017) which has a similar architecture but trained using transfer learning and the dataset is also expanded eight times by augmentation techniques like cropping images to two and flipping horizontally, vertically, and both. The Six segmented lesion cases generated by deep convolutional network in model 3 and similar architecture improved by transform domain features in model 4 are compared to corresponding masks in Figure 6. For these instances, masks produced by a deep convolutional network which is trained using a pre-trained model are presented as well. We compared against a pretrained model to emphasize the advantages of gradually improving the performance by training from scratch that is not easily feasible when limiting the model to be tuned from a pretrained model. In the work presented in (Pour, Seker, & Shao, 2017), fine tuning the network using a pretrained model that is trained on natural images are explored. The results of this study outperform the former particularly for noisy images which contains artifact or hair. For further validation, our model is also evaluated on ISIC 2017 dataset for the task of lesion segmentation. The performance metrics in Table 4 indicate that the proposed model outperforms the challenge results by 7% improvement in Sensitivity, 1.1% improvement in Accuracy and 2.2 % improvement in Jaccard index. We also evaluated the Jaccard index without post processing that was 0.778 and still higher than winner of the challenge and other models in table 4. A histogram of Jaccard Index values is shown in Figure 7. Although the number of images with Jaccard index higher than 0.9 in our model is lower than top challenge result, we achieved more images with Jaccard index between 0.75 to 0.9, besides fewer segmented images with Jaccard under 0.05.

4.5. Results for Task of Lesion Dermoscopic Feature segmentation

Two binary masks are used to identify the position of dermoscopic features (globules and streaks) in lesions. Evaluation metrics are the same as task1 and the aim is to automatically generate two masks (globules and streaks) for each test image. Transfer learning is applied using the model trained in previous part that helps the network converge fast. Encoder includes the similar convolutional layers as pretrained model and this architecture is followed by two parts, each contains two convolutional layers and four deconvolution layers to predict masks for both streaks and globules features.

The earlier layers (encoder part) retained freeze for the first 40 epochs then the whole network retrained with decreased learning rate by factor of ten. The segmentation metrics are calculated over entire test data set and the results are demonstrated in Table 5. Compared to the best result of challenge, 17 % improvement in Jaccard index is observed and samples of images from test dataset with predicted groundtruth are presented in Figure 2.

5. Discussion

Deep convolutional neural networks have widely improved various kinds of tasks solved by classical algorithms in machine learning over the past few years. When it comes to medical analysis, lack of appropriate sized dataset is a major dilemma. Although going deeper led to higher performance, it is more prone to overfitting. A solution for this issue is applying transfer learning. However, a model trained on medical data can be barely found to be used as a pretrained model. Moreover, the model can be hardly improved by modifying the architecture as we limited to use the similar architecture as pretrained model and be cautious to fix a layer due to local distribution representations may found in some layers as in (Yosinski, Clune, Nguyen, Fuchs, & Lipson, 2015) discussed. In (Pour, Seker, & Shao, 2017) using a pretrained model from semantic image dataset for the task of skin lesion segmentation is investigated. In this research, we explored training the network from scratch and improving the model by inserting appropriate features to the network for the task of skin lesion segmentation and dermoscopic attribute detection. A simple model based on convolutional neural network considered and improved gradually by appending appropriate features and optimization technique. we did not apply excessive data augmentation techniques to increase the dataset, instead multiscale and multidirectional representation of input images from transform domain are added to convolutional network that led to considerable increment of 12% in Jaccard index in model 2 compared to 6% raise in Jaccard index in model 3 that is modified by making the network deeper by increasing convolution layers from 7 to 15. We also compared the training time of these models in table 3 that shows the training time increases 128.83ms by increasing the number of layers from 7 to 15, compared to just 34.71ms for model with 7 convolution layer and integrated with the transform domain features. Figure 9 compares the output of 4th, 8th, 9th, and 11th convolution layer in models 3 and 4. Model 4 that includes image representations of contourlet transform is learning the pattern more effectively while model 3 that is a deep model without features of transform domain shows noisier patterns. The proposed model with incorporated representations (model4) shows 10 percent improvement in Jaccard index compared to similar model without adding features (model 3) that confirms the idea of inserting extra features to network particularly when dataset is scarce and going deeper can hardly improve the results due to overfitting problem. In comparison with the model in (Pour, Seker, & Shao, 2017) that is fine-tuned on pretrained model on natural images and data augmentation that is conducted to increase the data 8 times, the average Jaccard index has improved 3% and the proposed model indicates significantly higher performance in noisy images such as images that contains hair or artifacts.

The advantages of the proposed model, in short, contain not using preprocessing and excessive data augmentation, improving the performance of a not very deep and complex CNN based model by integrating transform domain representations, high performance of the model on small dataset and significant improvements in segmenting noisy images. The drawback of proposed model could be training of convolution neural network that is longer and more complicated than traditional methods.

6. Conclusion

In this work, a segmentation model based on convolutional neural network is proposed for tasks of lesion segmentation and dermoscopic feature segmentation. While adding more layers and increasing the depth are common ways to improve the accuracy of a convolutional neural network, may not be applicable to medical data as the network requires more training data that is a major issue in medical domain. To deal with this issue, many researches works use excessive augmentation algorithms that may add irrelevant data as well. Also, providing labelled data in medical domain is expensive and requires expert beside privacy issues of medical records that limits data access in this area. A further solution to use deep architectures for scarce data is transfer learning, this is also limited as a pretrained model on medical dataset can be hardly found. We investigated training the network from scratch and increasing depth of input to convolutional layers by concatenating efficient feature maps from transform domain and using CIELAB colour space in addition to RGB colour channels instead of excessive augmentation or using a pretrained model on natural images.

We supposed a basic convolutional neural network model that progressively improved, and the results are compared to the common techniques such as adding more layers to the network or transfer learning with data augmentation. In the first stage, incorporating multiscale image representations from transform domain improves the Jaccard index 12% while adding layers to the network increases 6% compared to our basic model. The model improved by combining these two models, boosting with CIELAB colour model and flipping as a trivial augmentation that outperformed the winner of 2017 challenge by 2.2% improvement in Jaccard index and 7% in sensitivity. A summary of the achievements of the proposed model includes:

- A convolutional neural network is designed to do automatic learning from training data with a deep architecture that applied to extract low level to high level features in various layers.

- The relevant feature maps concatenated to the network by inserting image representations from transform domain that provides a superior understanding of the input to the model.
- CIELAB colour space is applied in addition to RGB colour channels that provides more information for the network.
- This architecture benefits of not applying pre-processing methods as well as not excessive data augmentation techniques.
- Despite of small dataset, the proposed architecture is trained from scratch and improved the results particularly for noisy images compared to the model that is fine tuned on pretrained image on natural images.
- improves the accuracy 12% while adding layers to the network increases 6% compared to our basic model.
- The model with integrated transform domain features (model2) shows less inference time compared to model 3 that is made by more convolution layers.
- The final proposed model outperforms the results of both 2017 and 2016 challenges with 2% and 7% improvement in Jaccard index and Sensitivity for 2017 and an increase of 1%in Jaccard index with 6%in sensitivity for 2016.

The results demonstrate the efficiency of the proposed model, however there is potential of improvement by dealing with selected efficient relevant features and various deep architectures that would be considered as future work. Moreover, due to the robustness and practicality, this proposed framework will become a gold standard approach to the analysis of similar image data sets, in particular, medical and biological domains where there is always small number of samples available. Another task that we intend to explore in the future is extending this work for medical video segmentation. Incorporating the multi transform and multi directional features to the CNN network can help to detect and track the region of interest in video slides as an initial phase and speed up the segmentation process too.

CRediT author statement

Huseyin Seker: Conceptualization, Supervision, Writing- Reviewing and Editing,
Mansoureh Pezhman Pour: Conceptualization, Methodology, Software, Data curation, Investigation, Writing- Original draft preparation.

Declaration-of-competing-interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alamdari, N., MacKinnon, N., Vasefi, F., Fazel-Rezai, R., Alhashim, M., Akhbardeh, A., . . . Tavakolian, K. (2017). Effect of Lesion Segmentation in Melanoma Diagnosis for a Mobile Health Application. *2017 Design of Medical Devices Conference*, (pp. V001T12A005–V001T12A005).

Al-Masni, M., Al-antari, M., Choi, M.-T., Han, S.-M., & Kim, T.-S. (2018). Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Computer methods and programs in biomedicine, 162*, 221-231.

Bar, Y., Diamant, I., Wolf, L., & Greenspan, H. (2015). Deep learning with non-medical training used for chest pathology identification. *Medical Imaging 2015: Computer-Aided Diagnosis*, *9414*, p. 94140V.

Bi , L., Kim, J., Ahn, E., Kumar, A., Fulham, M., & Feng, D. (2017). Dermoscopic Image Segmentation via multistage fully convolutional networks. *IEEE Transactions on Biomedical Engineering, 64*(9), 2065--2074.

Chen, E., Dong, X., Wu, J., Jiang, H., Li, X., & Rong, R. (2018). Lesion Attributes Segmentation for Melanoma Detection with Deep Learning. *bioRxiv*, 381855.

Ciecholewski, M. (2016). An edge-based active contour model using an inflation/deflation force with a damping coefficient. *Expert Systems with Applications, 44*, 22--36.

Ciresan, D., Giusti, A., Gambardella, L., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in neural information processing systems*, (pp. 2843-2851).

Cireşan, D., Giusti, A., Gambardella, L., & Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. *International Conference on Medical Image Computing and Computer-assisted Intervention*, (pp. 411-418).

Codella, N., Gutman, D., Celebi, M., Helba, B., Marchetti, M., Dusza, S., . . . others. (2018). 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). *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, (pp. 168-172).

Deng, J., Socher, R., Fei-Fei, L., Dong, W., Li, K., & Li, L.-J. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, *00*, pp. 248-255.

Do, M., & Vetterli, M. (2005). The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing, 14*(12), 2091-2106.

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence, 35*(8), 1915-1929.

Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing, 321*, 321-331.

Garnavi, R., Aldeen, M., & Bailey, J. (2012). Computer-aided diagnosis of melanoma using border-and wavelet-based texture analysis. *IEEE Transactions on Information Technology in Biomedicine, 16*(6), 1239-1252.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, (pp. 249-256).

Gonzalez Rafael C.and Woods, R. (2007). *Digital image processing.* Pearson.

Gutman, D., Codella, N., Celebi, E., Helba, B., Marchetti, M., Mishra, N., & Halpern, A. (2016). 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*.

Haenssle, H., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., . . . others. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology, 29*(8), 1836-1842.

Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of biomedical informatics, 86*, 25-32.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 770-778).

Henning, J., Dusza, S., Wang, S., Marghoob, A., Rabinovitz, H., Polsky, D., & Kopf, A. (2007). The CASH (color, architecture, symmetry, and homogeneity) algorithm for dermoscopy. *Journal of the American Academy of Dermatology, 56*(1), 45-52.

Hussain, Z., Gimenez, F., Yi, D., & Rubin, D. (2017). Differential data augmentation techniques for medical imaging classification tasks. *AMIA Annual Symposium Proceedings*, *2017*, p. 979.

Jahanifar, M., Tajeddin, N. Z., & Gooya, A. (2018). upervised Saliency Map Driven Segmentation of Lesions in Dermoscopic Images. *IEEE journal of biomedical and health informatics, 23*(2), 509--518.

Jaisakthi, S., Mirunalini, P., & Aravindan, C. (2018). Automated skin lesion segmentation of. *IET Computer Vision, 12*(8), 1088--1095.

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, (pp. 11-19).

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., . . . Darrell, T. (2014). Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:1408.5093*.

Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, (pp. 1097-1105).

Kwasigroch, A., Mikołajczyk, A., & Grochowski, M. (2017). Deep neural networks approach to skin lesions classification—A comparative analysis. *2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR)*, (pp. 1069-1074).

Lau, H., Chang, J., Daut, N., Tahir, A., Samino, E., & Hijazi, M. (2018). Exploring Edge-Based Segmentation Towards Automated Skin Lesion Diagnosis. *Advanced Science Letters, 24*(2), 1095-1099.

Liang, D., Yang, F., Zhang, T., & Yang, P. (2018). Understanding mixup training methods. *IEEE Access, 6*, 58774-58783.

Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 1925-1934).

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 3431-3440).

Masoumi, H., Behrad , A., Pourmina, M., & Roosta, A. (2012). Automatic liver segmentation in MRI images using an iterative watershed algorithm and artificial neural network. *Biomedical signal processing and control, 7*(5), 429--437.

Nachbar, F., Stolz, W., Merkle, T., Cognetta, A., Vogt, T., Landthaler, M., . . . Plewig, G. (1994). The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology, 30*(4), 551-559.

Nasr-Esfahani, E., Rafiei, S., Jafari, M., Karimi, N., Wrobel, J., Soroushmehr, S., . . . Najarian, K. (2017). Dense Fully Convolutional Network for Skin Lesion Segmentation. *arXiv preprint arXiv:1712.10207*.

Navarro, F., Escudero Vinolo, M., & Bescos, J. (2018). Accurate segmentation and registration of skin lesion images to evaluate lesion change. *IEEE journal of biomedical and health informatics, 23*(2), 501--508.

Nguyen, D., Benameur, S., Mignotte, M., & Lavoie, F. (2018). Superpixel and multi-atlas based fusion entropic model for the segmentation of X-ray images. *Medical image analysis, 48*, 58--74.

Oliveira, R., Mercedes Filho, E., Ma, Z., Papa, J., Pereira, A., & Tavares, J. (2016). Computational methods for the image segmentation of pigmented skin lesions: a review. *Computer methods and programs in* *biomedicine, 131*, 127-141.

Pour, M., Seker, H., & Shao, L. (2017). Automated lesion segmentation and dermoscopic feature segmentation for skin cancer analysis. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (pp. 640-643).

Riaz, F., Naeem, S., Nawaz, R., & Coimbra, M. (2018). Active contours based segmentation and lesion periphery analysis for characterization of skin lesions in dermoscopy images. *IEEE journal of biomedical and health informatics, 23*(2), 489--500.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, (pp. 234-241).

Siegel, R., Miller, K., & Jemal, A. (2018). Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians, 68*(1), 7-30.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

Tajeddin, N. Z., & Asl, B. M. (2018). Melanoma recognition in dermoscopy images using lesion's peripheral region information. *Computer methods and programs in biomedicine, 163*, 143--153.

Wong, A., Scharcanski, J., & Fieguth, P. (2011). Automatic skin lesion segmentation via iterative stochastic region merging. *IEEE Transactions on Information Technology in Biomedicine, 15*(6), 929--936.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., & Lipson, H. (2015). Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579*.

Yu, L., Chen, H., Dou, Q., Qin, J., & Heng, P.-A. (2017). Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE transactions on medical imaging, 36*(4), 994-1004.

Zhang, H., Cisse, M., Dauphin, Y., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.

Zortea, M., Flores, E., & Scharcanski, J. (2017). A simple weighted thresholding method for the segmentation of pigmented skin lesions in macroscopic images. *Pattern Recognition, 64*, 92-104.

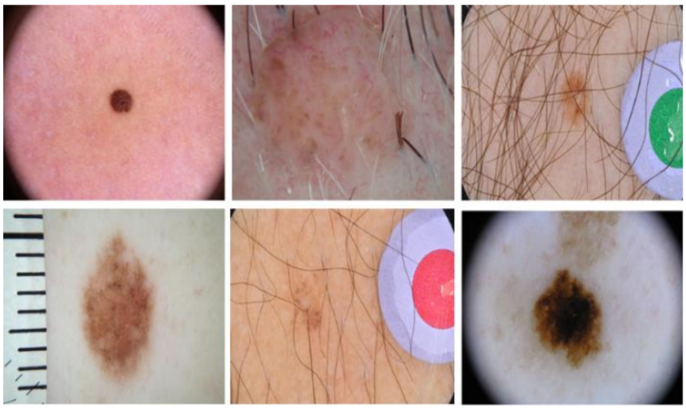


Figure 1: Sample images from ISIC dataset that show various issues such as variety in scales and color, existence of hair and other artifacts, and dark corners on images.

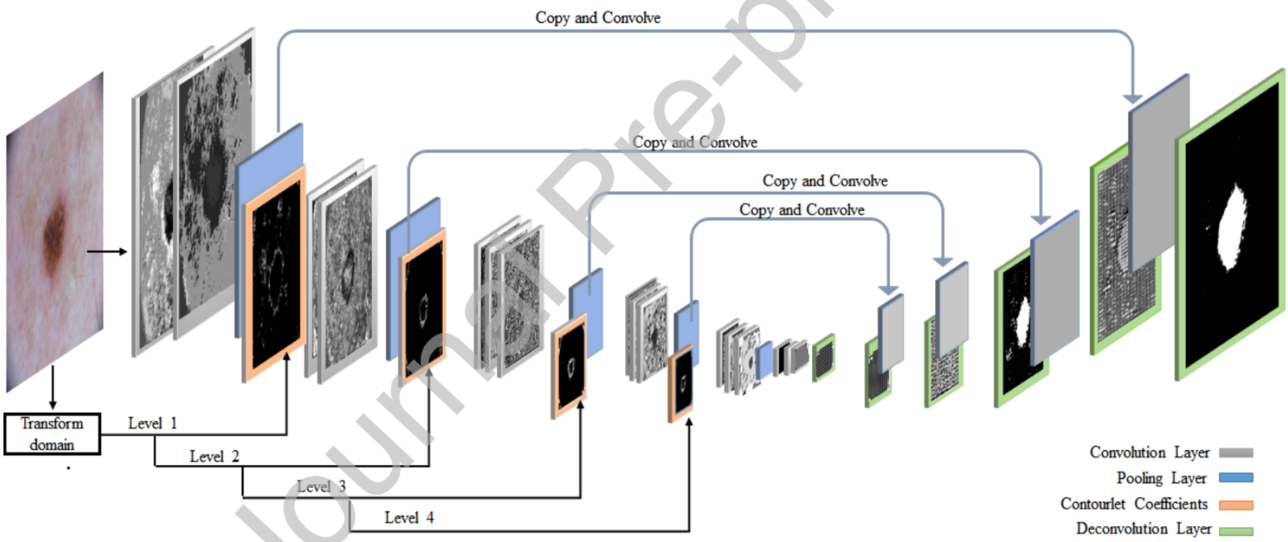


Figure 2: Architecture of proposed model, deep convolutional neural network proposed for lesion segmentation and image representations from various levels of contourlet transform.

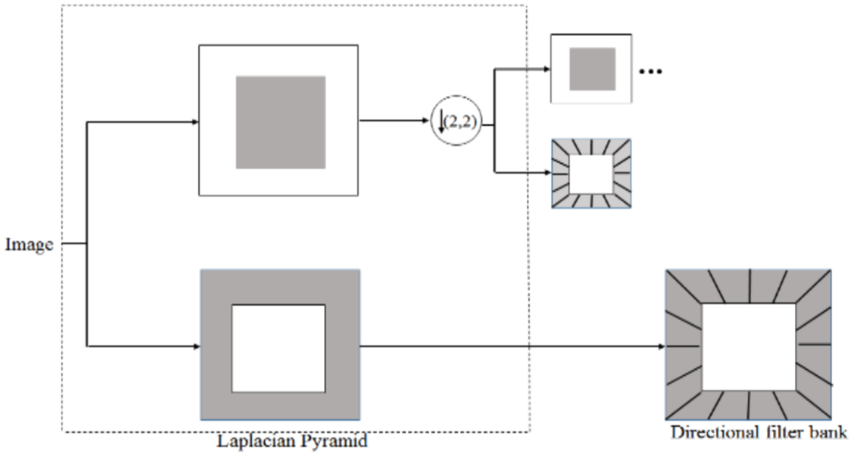


Figure 3: Contourlet transform composed of Laplacian Pyramid and Directional filter bank.

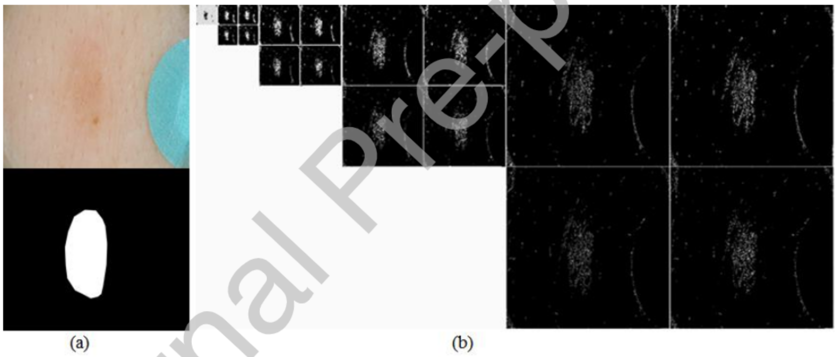


Figure 1: (a) Original image and Mask (b) Multiscale image representations from transform domain.

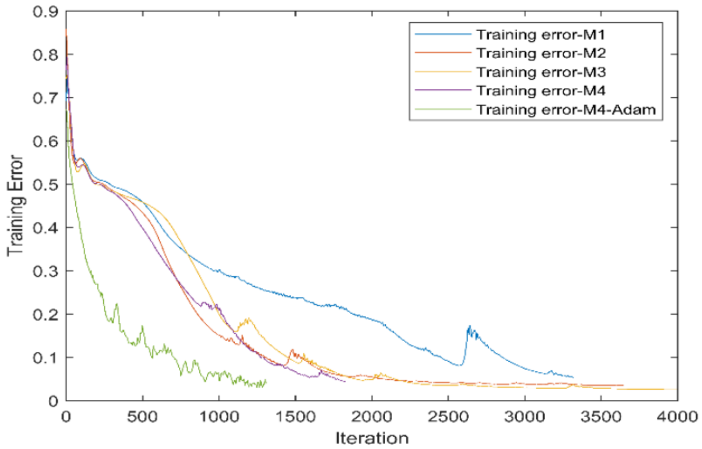


Figure 5: Training curves for models 1-4.

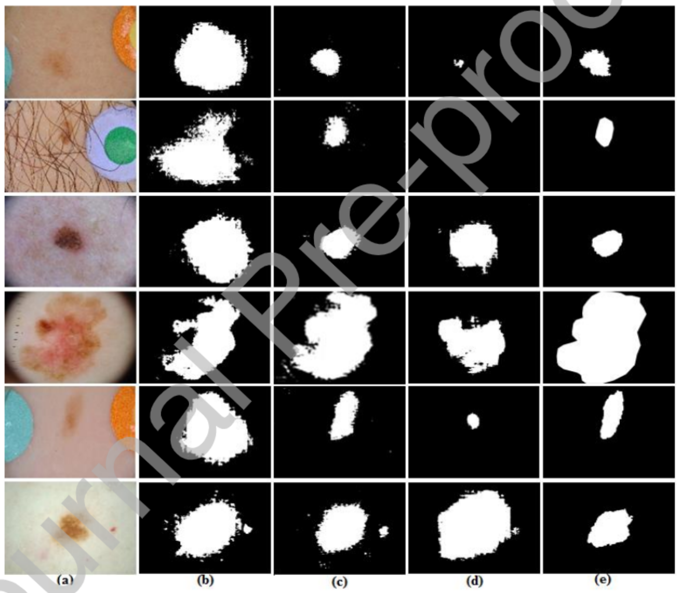


Figure 6: A) Origin image B) Segmented output from model 3 C) Segmented output from model 4 D) Output of model that is fine-tuned on a pretrained model (Pour, Seker, & Shao, 2017) E) Test mask.

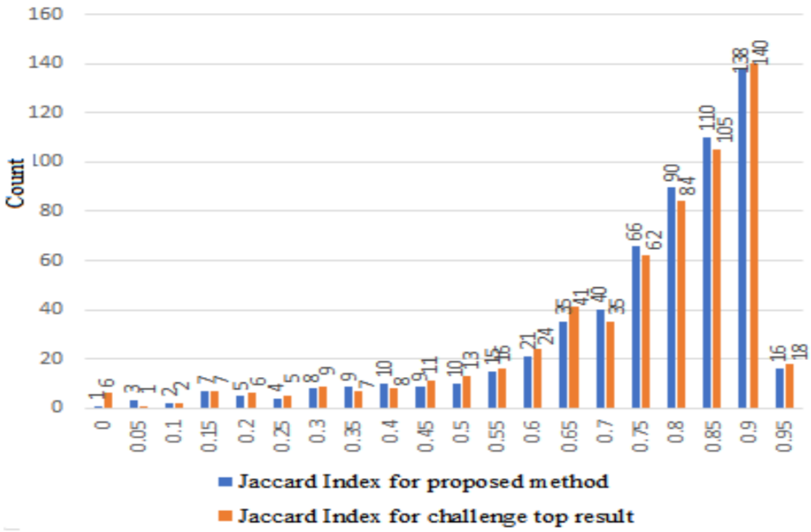


Figure 7: Histogram of Jaccard index values for proposed method compared to top result of the challenge ISIC 2017.

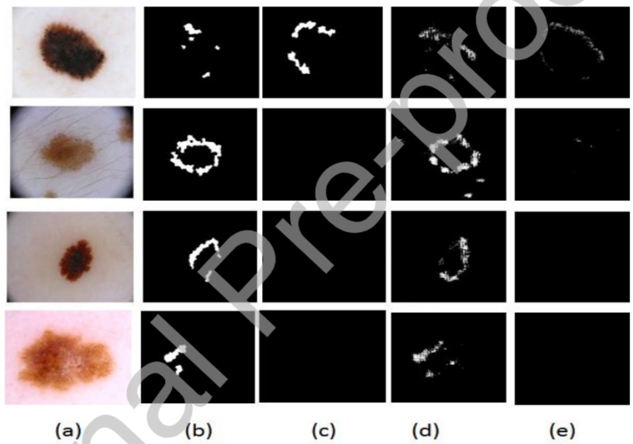


Figure 2: (a), (b), (c), are the origin images with globule and streak groundtruth respectively. (d) is the predicted globule groundtruth and (e) is the predicted streak.

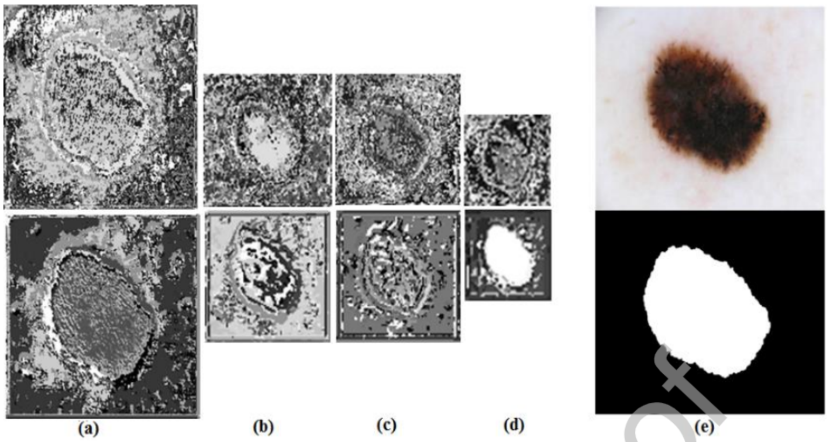


Figure 9: (a), (b), (c), (d) are the outputs of 4th,8th,9th and 11th convolution layer respectively. Images in the first raw are from model 3 and second raw from model 4. (e) Origin image and the groundtruth.

Table 1: Hyperparameters of Convolutional Neural Network- Model 1.

Layer	Filter size	Stride	Number of Filters	Size of Output
Conv1	3×3	1	16	1×16×698×698
Pool1	2×2	2	-	1×16×349×349
Conv2	3×3	1	32	1×32×349×349
Pool2	2×2	2	-	1×32×175×175
Conv3	3×3	1	64	1×64×175×175
Pool3	2×2	2	-	1×64×88×88
Conv4	3×3	1	128	1×128×88×88
Pool4	2×2	2	-	1×128×44×44
Conv5	3×3	1	256	1×256×44×44
Pool5	2×2	2	-	1×256×22×22
Conv6	7×7	1	512	1×512×16×16
Conv7	1×1	1	512	1×512×16×16
Deconv1	7×7	1	2	1×2×22×22
Deconv2	4×4	2	2	1×2×44×44
Deconv3	4×4	2	2	1×2×88×88
Deconv4	3×3	2	2	1×2×175×175
Deconv5	3×3	2	2	1×2×349×349
Deconv6	4×4	2	2	1×2×698×698

Table 2: Evaluation metrics for our different architectures compared to recent researches on skin lesion segmentation. Model 1 refers to the basic architecture composed of 7 convolution layers and 6 deconvolution layers, Model 2 is the model 1 incorporated with representations of transform domain, Model 3 is the model 1 but deeper (15 conv layers), and Model 4 is model 3 with integrated features of transform domain.

Method	SE	SP	ACC	DI	JA
(Bi , et al., 2017)	0.922	0.965	0.955	**0.912**	**0.846**
(Jahanifar, Tajeddin, & Gooya, 2018)	0.901	**0.982**	0.943	0.907	0.838
Best Result of Challenge (Gutman, et al., 2016)	0.910	0.965	0.953	0.910	**0.843**
Second Ranked in challenge, (Yu, Chen, Dou, Qin, & Heng, 2017)	0.911	0.957	0.949	0.897	0.829
Model in (Pour, Seker, & Shao, 2017)	0.911	0.950	0.943	0.893	0.826
Model 1	0.892	0.879	0.885	0.761	0.634
Model 2	0.927	0.913	0.918	0.849	0.752
Model 3	0.934	0.891	0.915	0.817	0.699
Model 4	0.948	0.922	0.939	0.881	0.803
Model4 with added images from CIElab color space model	0.952	0.931	0.947	0.895	0.816
Model 4(Augmented with flipped vertically)	0.974	0.949	**0.961**	**0.921**	**0.852**

Table 3: Training time comparison in different models.

Method	Average Forward pass	Average Backward pass
Model 1 with 7 convolution and 6 deconvolution layers	56.05ms	86.19ms
Model 2 that is Model 1 with contourlet coefficients combined to the network	70.45ms	106.50ms
Model 3 that is Model 1 with more layers (15 convolution and 6 deconvolution layers)	95.61ms	175.46ms

Table 4: Evaluation metrics for ISIC2017 dataset.

Method	SE	SP	ACC	DI	JA
1- (Jahanifar, Tajeddin, & Gooya, 2018)	0.810	0.981	0.930	0.839	0.749
2- (Navarro, Escudero Vinolo, & Bescos, 2018)	-	-	0.955	0.854	0.769
3- Best Result of Challenge (Codella, et al., 2018)	0.825	0.975	0.934	0.849	0.765
4- Proposed method	0.883	0.981	0.945	0.871	0.782
%improvement (Proposed method compared to 3)	7	0.6	1.1	2.5	2.2

Table 5: Evaluated metrics for the task of dermoscopic feature segmentation.

Method	SE	SP	ACC	DI	JA
Best Result of Challenge (Gutman, et al., 2016)	0.396	0.968	0.962	0.128	0.070
Proposed method	0.368	0.979	0.971	0.150	0.082
%improvement	-	1.1	1	17.2	17.1