Deep Convolutional Network-Based Framework for Melanoma Lesion Detection and Segmentation

Adekanmi Adegun and Serestina Viriri(✉)

School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Durban 4000, South Africa
{218082884,viriris}@ukzn.ac.za

Abstract. Analysis of skin lesion images is very crucial in melanoma detection. Melanoma is a form of skin cancer with high mortality rate. Both semi and fully automated systems have been proposed in the recent past for analysis of skin lesions and detection of melanoma. These systems have however been restricted in performance due to the complex visual characteristics of the skin lesions. Skin lesions images are characterised with fuzzy borders, low contrast between lesions and the background, variability in size and resolution and with possible presence of noise and artefacts. In this work, an efficient deep learning framework has been proposed for melanoma lesion detection and segmentation. The proposed method performs pixel-wise classification of skin lesion images to identify melanoma pixels. The framework employs an end-to-end and pixel by pixels learning approach using Deep Convolutional Networks with softmax classifier. A novel framework which learns the complex visual characteristics of skin lesions via an encoder and decoder subnetworks that are connected through a series of skip pathways that brings the semantic level of the encoder feature maps closer to that of the decoder feature maps is hereby designed. This efficiently handles multi-size, multi-resolution and noisy skin lesion images. The proposed system was evaluated on both the ISBI 2018 and PH2 skin lesion datasets.

Keywords: Deep learning · Deep convolution network · Skin lesion · Segmentation · Softmax classifier · Melanoma

1 Introduction

Melanoma skin cancer is a rapidly increasing skin cancer incidence with a very high mortality rate [1]. This type of cancer can be easily treated and cured if detected and diagnosed early [2]. Manual screening of skin lesions for detecting melanoma is cumbersome with inaccurate results. Many Computer Aided Diagnosis (CAD) techniques have been proposed in the recent past for the analysis and segmentation of skin lesions [3]. Both fully and semi-automatic methods have been however limited in performance due to the granulate and complex

J. Blanc-Talon et al. (Eds.): ACIVS 2020, LNCS 12002, pp. 51–62, 2020.
https://doi.org/10.1007/978-3-030-40605-9_5

visual appearance of skin lesions images [3]. Skin lesions possess features such as colour variation, multi-size, multi-resolution, low contrast with background, fuzzy borders and inhomogeneous textures. They are also characterised with artefacts presence such as hair, oil, bubbles, air etc.

Computer vision approaches such as segmentation techniques have been recently utilized for detection and diagnosis of melanoma skin cancer [4]. This work explores some recently developed deep learning techniques for the segmentation and possible detection of melanoma lesions from dermoscopy images. The performance of these techniques also experiences some restrictions due to the [5] distinctive features of the skin lesions. In this work, a deep learning framework that tends to overcome these limitations has been proposed for the detection and segmentation of melanoma lesions. The proposed framework employs an end-to-end training of skin lesion images with melanoma disease labels [6] and then, performs pixel by pixel classification of the images using the Deep Convolutional Network.

This work devises a novel Deep Convolutional Network-based framework which learns the complex visual characteristics of skin lesions via an encoder and decoder sub-networks that are connected through a series of skip pathway for easy transmission. Original UNet model has been redesigned though the introduction of series of skip pathways for connecting the encoder section with the decoder section. Only short skip connection was used in UNET [7]. The system is able to overcome the challenges with complex visual characteristics of the skin lesions through learning general visual characteristics by the encoder networks and the lesion boundaries details through the decoder networks. The system performs pixel-wise classification using the softmax classifier. The proposed network utilizes smaller number of trainable parameters such as employing 3×3 filter size and limiting the encoder-decoder level to five to reduce computational resources and time for processing the lesions images. The system was evaluated on publicly available skin lesions image dataset of ISBI 2018 and PH2 with performance metrics such as Accuracy, Confusion matrix, Dice Coefficient, Sensitivity and Specificity. Each pixel of the skin lesion image can then be presented as either True positives (TP), True negatives (TN), False positives (FP) and False negatives (FN). Our Contributions can be summarized as follows:

- We propose a novel Deep Convolutional Network-based framework which learns the complex visual characteristics of skin lesions via an encoder and decoder sub-networks that are connected through a series of skip pathway for efficient skin lesion segmentation. The encoder sub-network learn the coarse appearance and localization information of the lesion images while the decoder sub-networks learn the lesion boundaries information. The skip pathways brings the semantic level of the encoder feature maps closer to that of the decoder feature maps for easy transmission and improved processing speed.
- We propose pixel-wise classification of skin lesions images for quick and easy melanoma detection via softmax function.

- We compare our results with the existing methods. The results show that our method is effective for segmenting challenging melanoma lesions with fuzzy boundaries and heterogeneous textures.

2 Related Works

Machine learning techniques have been applied for various tasks in the past [8]. Recently, machine learning techniques most especially deep learning methods have been used to perform segmentation and classification of medical images. Specifically deep learning approach have been used in the segmentation and classification of skin lesion images. Bi et al. [9] proposed a method for automated skin lesion segmentation and melanoma detection by introducing a probability-based step-wise integration to combine complementary segmentation results derived from individual class-specific learning models. The system achieved an average Dice coefficient of 85.66% on the ISBI 2017 Skin Lesion Challenge (SLC), 91.77% on the ISBI 2016 SLC and 92.10% on the PH2 datasets with corresponding Jaccard indices of 77.73%, 85.92% and 85.90%, respectively. The complexity level of the system architecture is still higher.

A segmentation methodology through full resolution convolutional networks (FrCN) was proposed by Al-masni et al. [10]. The proposed method directly learns the full resolution features of each individual pixel of the input data without the need for pre or post-processing operations such as artefact removal, low contrast adjustment, or further enhancement of the segmented skin lesion boundaries. The system was evaluated using two publicly available databases, the IEEE International Symposium on Biomedical Imaging (ISBI) 2017 Challenge and PH2 datasets. The proposed FrCN achieved a segmentation accuracy of 95.62% for benign cases, 90.78% for the melanoma cases, and 91.29% for the seborrheic keratosis cases in the ISBI 2017. Bi et al. [11] proposed leveraging fully convolutional networks (FCNs) to automatically segment the skin lesions.

He et al. [12] proposed a skin lesion segmentation network using a very deep dense deconvolution network based on dermoscopic images. The deep dense layer and generic multi-path Deep RefineNet were combined to improve the segmentation performance. The deep representation of all available layers was aggregated to form the global feature maps using skip connection. The dense deconvolution layer is leveraged to capture diverse appearance features via the contextual information and to smooth segmentation maps for final high-resolution output. Bi et al. [13] proposed semi-automated skin lesion segmentation method that incorporates fully convolutional networks (FCNs) with multi-scale integration to overcome problems with over- or under-segmentation with challenging skin lesions such as when a lesion is partially connected to the background or when image contrast is low. Esteva et al. [14] developed CNN architecture using GoogleNet Inception v3 that was pre-trained on approximately 1.28 million images for melanoma detection. The system was pre-trained on a suitable dataset using a set of more than 129,450 high quality skin images. The system was deconstructed down to the pixel level and trained with pattern and its diagnosis [14].

Goyal et al. [15] proposed an end-to-end solution using fully convolutional networks (FCNs) for multi-class semantic segmentation. The system automatically segmented the melanoma into keratoses and benign lesions. Ramachandram et al. [16] proposed an approach based on a Fully Convolutional Neural Network architecture which was trained end to end, from scratch, on a small dataset. The semantic segmentation architecture utilized combined the use of atrous convolutions to increase the effective field of view of the network's receptive field without increasing the number of parameters, network-in-network convolution layers to increase network capacity and super-resolution upsampling of predictions using subpixel. A deep learning framework consisting of two fully-convolutional residual networks (FCRN) was proposed to simultaneously produce the segmentation result and the coarse classification result of skin lesion [17]. A lesion index calculation unit (LICU) was then developed to refine the coarse classification results by calculating the distance heat-map. A straight-forward CNN was also proposed for the dermoscopic feature extraction task between melanoma and non-melanoma lesions. Lastly Ramachandram et al. [18] presented an image segmentation method based on deep hyper-column descriptors for the segmentation of several classes of benign and malignant skin lesions. The system focuses on the task of accurately segmenting benign and malignant skin lesions in dermoscopic images through a means of lesion quantification.

Our proposed system aims at lowering trainable parameters to reduce computational resources and time and make the system feasible for real-time medical diagnosis. Most of the systems discussed above employ larger and more complex deep learning architecture. Our proposed system is able to perform both segmentation and pixel-wise classification of melanoma lesion pixels using a moderate-size deep convolutional network.

3 Methodology

Our methodology employs a pre-processing step that crops, resize and re-samples images to ensure that both the training images and the ground truth conform to the same resolution and size. The processed images are then sent into Deep Convolutional network. The architectural diagram of Deep Convolutional Network and the whole framework of the system is described in Figs. 1 and 2 respectively. Our architecture is a deep-supervised encoder-decoder network where the encoder and decoder sub-networks are connected through a series of skip pathways which are dense convolution block with two number of convolution layers. The dense convolution block brings the semantic level of the encoder feature maps closer to that of the decoder feature maps. This increases the processing speed. The whole process of learning and training is accomplished in end-to-end and pixels by pixels. Each pixel from the training images is assigned with a pixel from the ground truth labels. The framework also ensures less train-able weights through the application of the pixel-wise classification with the reduced number of feature maps in the last convolutional block and limited to five numbers of the encoding-decoding levels. This ensures a medium-size with lower cost in terms of memory and computation time.

3.1 Dataset

Data Pre-processing. The skin lesions images acquired for processing are always characterised with variation in size and multi-scale and multi-resolution nature features of the skin lesion images. The first task in the pre-processing stage is to have all images in the same scale and resolution via cropping, resizing and re-sampling. In this work, we use relatively small image size of 160×224 as inputs as this affects the input feature map size. This size has been chosen to avoid image deformations and for better presentation. The images are then automatically normalized by computing the mean pixel value the standard deviation for data normalization.

Image Augmentation. The system applied elastic deformations through random displacements before augmenting the dataset. Elastic deformation utilizes local distortion and random affine transformation for high-quality output. These transformation takes place with random displacement. In addition, simple and random rotation is adopted in the augmentation process to increase the training dataset and to improve the performance.

3.2 Deep Convolutional Network

The framework is based on Deep Convolutional Network with an Encoder-Decoder sub-network. The network learns the important visual characteristics and the pixels' localization of the skin lesions through the encoder sub-networks. The decoder sub-network learns the spatial information recovery and the lesion boundaries for the melanoma lesions. Training of skin lesions images in this network is performed from end-to-end and pixels-wise using only the pixels from lesion images and corresponding ground truth labels as the system input. The network architecture is described in Fig. 1 below.

In the encoder network, we use five 3×3 convolutions blocks with each block composing 2 convolutional layers, 2×2 max-pooling layers for down-sampling, filter kernel and ReLU activation function. The convolution layers perform feature extraction and generates feature maps from the input image. Features maps are extracted from the input images by the convolutional layers. The ReLU activation function transforms the feature maps to enable training and learning of the image patterns. The transformed output is then sent to the next convolution layer as input. This is illustrated below:

$$F_m = conv2d(ReLU, K_f, I_m) \tag{1}$$

where F_m is the extracted feature map, I_m is the input image into the convolutional layer, K_f is the filter kernel and $ReLU$ is the activation function applied on the feature map for each layer. The down-sampling layers performs size reduction of the extracted feature maps to eliminate and reduce features redundancy and over-fitting and eventually minimises the overall computing

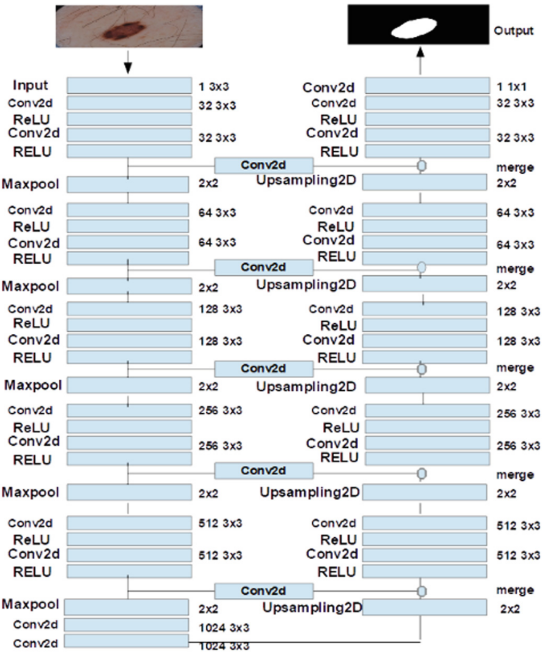

Fig. 1. Network architecture diagram for Deep Convolutional Network

processing time. This however causes reduction of the spatial features resolution of the input image. It utilizes the max-pooling function stated below.

$$layer = maxPooling2dLayer(poolSize) \tag{2}$$

The decoder learns the spatial features recovery information and the restores the original size of the feature map. In the decoder network, we use five 3×3 convolutions blocks with each block composing 2 convolutional layers, 2×2 up-sampling layers, filter kernel and ReLU activation function. The up-sampling layers with size 2×2 performs spatial feature recovery and boundaries localization.

$$y = upsample(x, n) \tag{3}$$

The skip connection in-between the encoder and the decoder has been replaced with series of skip pathway to ensure better segmentation. The encoded features are merged with the decoded features at a given spatial resolution.

3.3 Pixel-Wise Classification

The final decoder output with high dimensional feature representation is sent into a trainable soft-max classifier with dice loss function. This module identifies melanoma lesions through a pixel-wise classification of skin lesions images. The

final feature maps are sent into the softmax module for pixels classification using the mathematical function stated below. The softmax classifier predicts the class for each pixel as melanoma or non-melanoma. In this research, we employ binary classification where n represents two number of classes and the output is a two-channel image of probabilities. The predicted segmentation therefore corresponds to the class with maximum probability at each pixel.

$$P(y = i|x) = \frac{e^{x^T w_i}}{\sum_{n=1} e^{x^T w_n}} \tag{4}$$

where $x^T w$ represents the product of x and w, x is the feature map and w is the kernel operator.

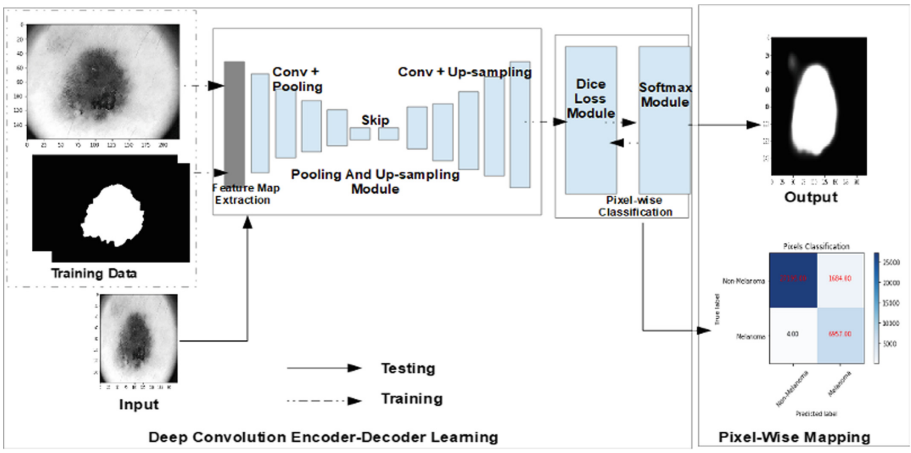

Fig. 2. The framework for the proposed system

4 Experimental Results

4.1 Dataset

The two well-established publicly available datasets used in the evaluation of the proposed segmentation method are from the ISIC challenge in skin lesion segmentation and PH2 data repository. PH2 [19] contains 200 skin lesion images with highest resolution of 765×574 pixels. This dataset was categorized into training and testing image set both comprising of images and ground truth labels respectively. The input dataset are skin lesion images in JPEG format while the ground truth are mask image in PNG format. ISBI 2018 [20] contains 2000 training images with the ground truth provided by expert clinicians. The proposed system is an expert system that has been subjected to test and generally acceptable. The image sizes possess highest resolution of 1022×767.

4.2 Evaluation Metrics

The most common and standard skin lesion segmentation evaluation metrics were used for comparison including: Dice Similarity Coefficient (DSC.), Sensitivity, Specificity, Precision and Accuracy. These metrics were used for evaluation of the model. They are illustrated below:

Dice Similarity Coefficient: It measures the similarity or overlap between the ground truth and the automatic segmentation. It is defined as

$$\textbf{DSC} = \frac{2TP}{FP + 2TP + FN} \tag{5}$$

Sensitivity: It measures the proportion of those with positive values among those who are actually positive. It is also known as True Positive rate (TPR)

$$\textbf{Sensitivity} = \frac{TP}{TP + FN} \tag{6}$$

Precision: This is the proportion of positive predicted values that are truly positive.

$$\textbf{Precision} = \frac{TP}{TP + FP} \tag{7}$$

Accuracy: It measures the proportion of true results (both true positives and true negatives) among the total number of cases examined.

$$\textbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Where FP is the number of false positive pixels, FN is the number of false negative pixels, TP is the number of true positive pixels and TN is the number of true negative pixels.

4.3 Results and Discussion

In the result, Figs. 5 and 6 show the performance of the proposed method on ISBI 2018 and PH2 skin lesion image datasets respectively. The results show improvement over existing methods as displayed in Table 1. The result give us accuracy and dice coefficient of 95% and 93% with the specificity and sensitivity of 95% and 93% respectively as shown in Table 1. This reflects in our predicted segmented output in the third row when compared with the ground truth result in the second row in Fig. 5.

Three images have been analysed pixel-wisely. The results are shown in Fig. 6 with the confusion matrix on the last column. The first image gives pixel classification result of 4060 pixels classified as melanoma correctly, 25135 pixels classified as non-melanoma correctly (this can be the background of the lesion or the healthy tissue) while 7 pixels are classified as non-melanoma in-correctly and 2639 pixels of non-melanoma classified as melanoma. The second image gives

pixel classification result of 6957 pixels classified as melanoma correctly, 27195 pixels as non-melanoma classified correctly, 4 pixels of melanoma classified as non-melanoma and 1684 pixels of non-melanoma classified as melanoma. Lastly, the third image gives 5112 pixels classified as melanoma correctly, 29260 pixels classified as non-melanoma correctly while 153 pixels of melanoma are classified as non-melanoma and 1318 pixels of non-melanoma classified as melanoma.

Figure 3 below shows the dice coefficient and training loss curves. These curves identify that the loss reduces significantly as the dice coefficient increases. It clearly shows the relationship with the dice loss function relationship adopted in the model. The overall performance indicates that the dice coefficient which translates to the similarities between the predicted result output and the ground truth is very high with 93% of dice coefficient and less than 10% loss acquired. It can be inferred that the system performs efficiently.

Figure 4 shows the sensitivity and precision curves. These curves identify that the errors in classifying melanoma pixels as non-melanoma pixels reduces significantly as the sensitivity increase. Our sensitivity and precision shows that the number of melanoma pixels classified correctly as melanoma is high. It clearly shows the relationship between sensitivity and precision curves. The overall performance indicates that the sensitivity which translates to true positive rate is very high with value of 94% and the Precision also known positive predictive value also high with value of 93%.

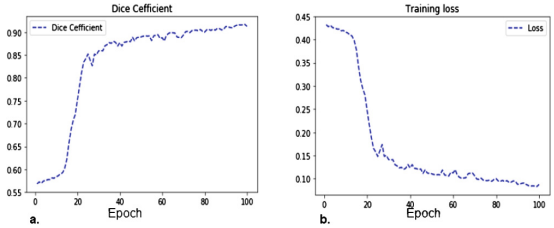

Fig. 3. (a) Dice coefficient curve (b) Training loss curve of the proposed method on the PH2 datasets

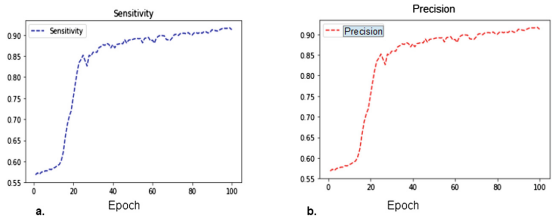

Fig. 4. (a) Sensitivity curve (b) Precision curve of the proposed method on the PH2 datasets

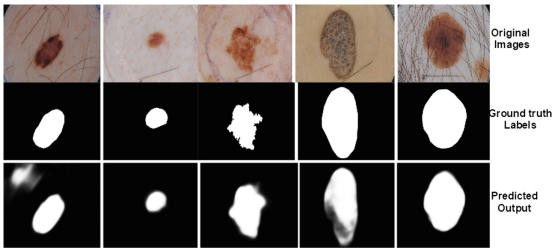

Fig. 5. Segmentation output of the proposed method on ISBI 2018 dataset with the third row showing the predicted results

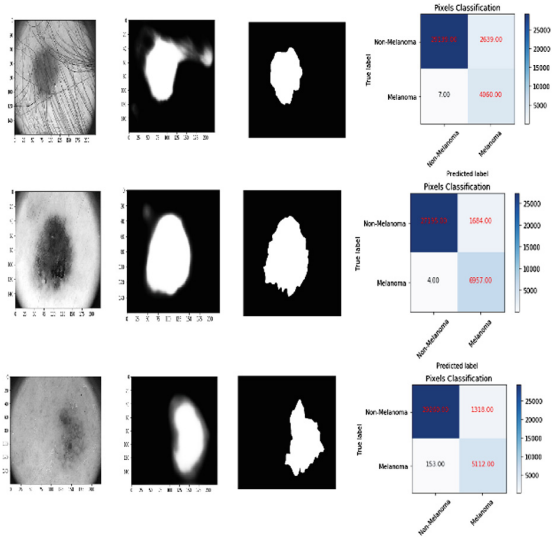

Fig. 6. Performance of the proposed method on PH2 dataset with the second column showing the segmented output and fourth column showing the confusion matrix result on each image. First column is the input image and the third column is the corresponding ground truth label.

Table 1. Segmentation performance (%) of the proposed model on PH2 datatest.

Techniques	Accuracy	Dice score	Sensitivity	Specificity
Proposed model	**95**	**93**	**93**	**95**
FrCN [10]	95	91	95	95
FCN [11]	94	90	95	94
mFCN [11]	96	91	96	95

5 Conclusion

This research explores Deep Convolutional Network-Based Framework for Melanoma Lesion Detection and Segmentation. The proposed system performs segmentation of melanoma lesions. The background of the images used have been given appropriate size for a reliable accuracy results. The system proposes pixel-wise classification of skin lesions for detection of melanoma cancer. The system aims at reducing deep learning architecture complexity in detecting melanoma. It also aims at developing an efficient system that can meet up with real time medical diagnosis task in diagnosing melanoma cancer. The proposed method is feasible for medical practices with the processing time for each dermoscopy image at averagely 5 s. The system was evaluated and tested on two publicly available database of ISBI 2018 and PH2. The results show that it out-performs some existing state-of-the-arts methods. Our results also show that the proposed system only under-performs in less than 10% of the image samples collected which could be caused by missing labels. Post-processing approach has been proposed for further refinements.

References

1. Apalla, Z., Lallas, A., Sotiriou, E., Lazaridou, E., Ioannides, D.: Epidemiological trends in skin cancer. Dermatol. Pract. Concept. **7**(2), 1 (2017)
2. 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)
3. Masood, A., Al-Jumaily, A.A.: Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. Int. J. Biomed. Imaging **2013**, 1–23 (2013)
4. Mengistu, A.D., Alemayehu, D.M.: Computer vision for skin cancer diagnosis and recognition using RBF and SOM. Int. J. Image Process. (IJIP) **9**(6), 311–319 (2015)
5. Salido, J.A.A., Ruiz Jr., C.: Using deep learning to detect melanoma in dermoscopy images. Int. J. Mach. Learn. Comput. **8**(1), 61–68 (2018)
6. Khagi, B., Kwon, G.-R.: Pixel-label-based segmentation of cross-sectional brain MRI using simplified SegNet architecture-based CNN. J. Healthc. Eng. **2018**, 1–9 (2018)
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
8. Adegun, A.A., Akande, N.O., Ogundokun, R.O., Asani, E.O.: Image segmentation and classification of large scale satellite imagery for land use: a review of the state of the arts. Int. J. Civ. Eng. Technol. **9**(11), 1534–1541 (2018)
9. Bi, L., Kim, J., Ahn, E., Kumar, A., Feng, D., Fulham, M.: Step-wise integration of deep class-specific learning for dermoscopic image segmentation. Pattern Recogn. **85**, 78–89 (2019)
10. 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. Comput. Methods Programs Biomed. **162**, 221–231 (2018)

11. Bi, L., Kim, J., Ahn, E., Kumar, A., Fulham, M., Feng, D.: Dermoscopic image segmentation via multistage fully convolutional networks. IEEE Trans. Biomed. Eng. **64**(9), 2065–2074 (2017)
12. He, X., Yu, Z., Wang, T., Lei, B., Shi, Y.: Dense deconvolution net: multi path fusion and dense deconvolution for high resolution skin lesion segmentation. Technol. Health Care **26**(S1), 307–316 (2018)
13. Bi, L., Kim, J., Ahn, E., Feng, D., Fulham, M.: Semi-automatic skin lesion segmentation via fully convolutional networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 561–564. IEEE (2017)
14. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115 (2017)
15. Goyal, M., Yap, M.H.: Multi-class semantic segmentation of skin lesions via fully convolutional networks. arXiv preprint arXiv:1711.10449 (2017)
16. Ramachandram, D., DeVries, T.: LesionSeg: semantic segmentation of skin lesions using deep convolutional neural network. arXiv preprint arXiv:1703.03372 (2017)
17. Li, Y., Shen, L.: Skin lesion analysis towards melanoma detection using deep learning network. Sensors **18**(2), 556 (2018)
18. Ramachandram, D., Taylor, G.W.: Skin lesion segmentation using deep hypercolumn descriptors. J. Comput. Vis. Imaging Syst. **3**(1), 1–3 (2017)
19. Mendonça, T., Ferreira, P.M., Marques, J.S., Marcal, A.R.S., Rozeira, J.: PH 2-A dermoscopic image database for research and benchmarking. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5437–5440. IEEE (2013)
20. Codella, N.C.F., 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. IEEE (2018)