

Supervised Versus Unsupervised Deep Learning Based Methods for Skin Lesion Segmentation in Dermoscopy Images

Abder-Rahman Ali¹, Jingpeng Li¹, and Thomas Trappenberg²

¹ Division of Computer Science and Mathematics, University of Stirling, United Kingdom

abder@cs.stir.ac.uk, jingpeng.Li@stir.ac.uk

² Faculty of Computer Science, Dalhousie University, Canada
tt@dal.cs.ca

Abstract. Image segmentation is considered a crucial step in automatic dermoscopic image analysis as it affects the accuracy of subsequent steps. The huge progress in deep learning has recently revolutionized the image recognition and computer vision domains. In this paper, we compare a supervised deep learning based approach with an unsupervised deep learning based approach for the task of skin lesion segmentation in dermoscopy images. Results show that, by using the default parameter settings and network configurations proposed in the original approaches, although the unsupervised approach could detect fine structures of skin lesions in some occasions, the supervised approach shows much higher accuracy in terms of Dice coefficient and Jaccard index compared to the unsupervised approach, resulting in 77.7% vs. 40% and 67.2% vs. 30.4%, respectively. With a proposed modification to the unsupervised approach, the Dice and Jaccard values improved to 54.3% and 44%, respectively.

Keywords: deep learning · dermoscopy · melanoma.

1 Introduction

Skin cancer is considered the most common cancer worldwide such that one in every three cancers diagnosed is a skin cancer. In Canada more than 80,000 skin cancer cases are diagnosed each year, of which more than 5,000 cases are melanoma, the deadliest form of skin cancer [1]. In the UK around 100,000 cases are diagnosed each year, accounting for 20% of all cancer cases. In recent years, there have been great advances of fully and semi-automated Computer Aided Diagnosis (CAD) methods [2]. These CAD methods can assist dermatologists in different steps of analysis, such as detection of the lesion boundary, quantification of diagnostic features, classification into different lesion types (tumor staging), and visualization [3]. Essentially, the standard pipeline in automatic dermoscopic image analysis consists of three stages: image segmentation, feature extraction and selection, and lesion classification. Image segmentation is considered a crucial

stage since it affects the accuracy of subsequent stages. However, segmentation is difficult due to great variety of lesion shapes, sizes, textures and colors. Other difficulties are related to the presence of dark hair covering the lesions and the existence of specular reflections [4].

Different skin lesion segmentation methods have been proposed in the literature, which can be classified as [5] low-level based image processing methods [6, 7], active contour based methods [8], and supervised learning based methods [9]. The huge progress in machine learning, specifically in the area of deep learning, has recently revolutionized the image recognition and computer vision domains. Numerous studies have subsequently emerged employing machine learning in skin cancer and melanoma detection. The ultimate goal of applying machine learning to medical images is to recognize patterns in a better and quicker way than humans can, and thus to increase the productivity of doctors and the patient healthcare outcomes. One such application is skin lesion *segmentation*.

Some related work is highlighted in Section 2, U-Net is described in Section 3 as a supervised deep learning based method, the unsupervised deep learning based method is described in Section 4, results are depicted and discussed in Section 5, and the paper is concluded in Section 6.

2 Related work

Jafari et al. [10] used a Convolutional Neural Network (CNN) that accepts input images (pre-processed for reducing the artefacts) and combines both local and global contextual information, eventually producing a label for each pixel which results in a segmentation mask that represents the skin lesion. The results have been further refined by some post-processing operations, and comparisons to state-of-the-art algorithms have shown to outperform such methods in terms of segmentation accuracy and sensitivity, which are 98.5% and 95%, respectively.

A hybrid approach that uses convolutional and recurrent neural networks was proposed by Attia et al. [11]. The approach was tested on the ISIC 2016 challenge (International Skin Imaging Collaboration) including 900 training images and 375 test images, without making any pre-processing. This method has achieved a segmentation average accuracy of 98% and Jaccard index of 93%, and has demonstrated to be robust against different artifacts with high sensitivity. Compared to other state-of-the-art methods including those using CNNs only, the approach shows outperformance.

Yuan et al. [12] proposed a fully automatic method that uses a 19-layer convolutional neural network for skin lesion segmentation. Moreover, the authors used a novel loss-function based on Jaccard distance which eliminates the need for sample re-weighting that is utilized when cross entropy is used as the loss-function. The approach is tested on the ISIC 2016 ³ challenge and the PH^2 ⁴ database, and has shown to outperform other state-of-the-art methods.

³ <https://www.isic-archive.com>

⁴ <http://www.fc.up.pt/addi/ph2%20database.html>

Al-Masni et al. [13] proposed a segmentation method that uses Full resolution Convolutional Networks (FrCN) that learn the full resolution features of each individual pixel of the input image to achieve better pixel-wise segmentation of the skin lesions. Such task is performed by eliminating the subsampling layers in the networks and enabling the convolutional layers in order to extract and learn full spatial features of the skin lesions, resulting in finely segmented contours of the lesions. No pre-processing or post-processing operations are used. The approach has been tested on the ISIC 2017 challenge and on the PH^2 database. The proposed method has achieved an average Jaccard index and overall segmentation accuracy of 77.11% and 94.03% on the ISIC 2017 dataset, respectively. On the PH^2 database, the results were 84.79% and 95.08%, respectively. Compared with FCN [15], U-Net [16] and SegNet [17], the proposed method outperforms those networks by 4.94%, 15.47% and 7.48% for Jaccard index and 1.31%, 3.89% and 2.27% for the segmentation accuracy, respectively.

A deep learning architecture based on U-Net that uses multi-scale residual connections (Multiscale Residual UNet) to maximize the learning capability and performance of the network was proposed in [14]. The residual connections preserve the information lost in the encoder stages due to max-pooling. A layer with binary cross-entropy loss function based on Jaccard index is included for the purpose of pixel classification. The method was applied on the ISIC 2017 challenge dataset and showed to have better performance compared to other methods in literature, in addition to producing results where the boundaries of lesion regions and the background are well separated. Many other deep learning based skin lesion segmentation methods have been proposed. Some of those methods can be referred to in [15]-[28].

Unlike most of the studies mentioned above which are considered supervised deep learning based methods, we utilize a commonly used architecture in the medical image domain for the supervised part of the paper (i.e. U-Net). Taking that further, we also employ an unsupervised deep learning based approach for skin lesion segmentation, which could be considered a good approach when a shortage of data and its ground truth is faced, which is a common challenge in the medical image domain. No pre-processing or post-processing operations are carried out.

3 Supervised deep learning based segmentation

In this section we are going to discuss U-Net as a supervised deep learning based network architecture. U-Net is an end-to-end encoder-decoder network for semantic segmentation which was firstly used in medical image segmentation. The architecture is composed of left (down) and right (up) sides. The *down* part, which follows the typical convolutional network architecture, is the encoder part where convolution blocks are applied followed by maxpooling in order to encode the input image into feature representations at multiple levels. The number of feature channels are doubled at each downsampling step. On the other hand, the *up* part consists of upsampling the feature map followed by a convolution oper-

ation that brings the number of feature channels to half. A concatenation with the corresponding cropped feature map from the *down* part occurs, followed by two 3×3 convolutions which are also followed by two ReLU operations and one 2×2 max-pooling operation with stride 2 used for downsampling. The cropping process is essential as border pixels are lost at each convolution. Thus, while up-sampling we concatenate the higher resolution features obtained from the *down* part with the upsampled features in order to localize and learn representations better. The resulting architecture is that the expansive path is symmetric to the concatenating path, yielding a u-shaped architecture (Fig. 1). The final layer of U-Net uses a 1×1 convolution to map each 64 feature vector to the desired number of classes. The network is composed of 23 convolutional layers in total, provided that it does not have any fully connected layers and uses only the valid part of each convolution. For the border region of the image, the pixels are predicted by an *overlap-tile strategy* such that the missing context is extrapolated by mirroring the input image, thus allowing the U-Net network to be applied on large images. The outcome of this stage is the segmented image, where the foreground (in white) represents the skin lesion and the background (in black) represents the skin.

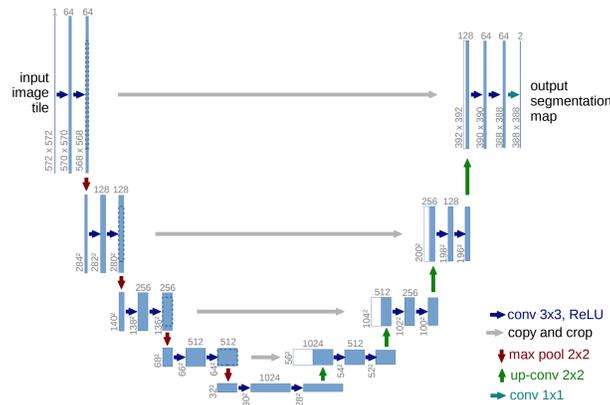


Fig. 1. U-Net architecture: blue boxes correspond to a multi-channel feature map with the number of channels shown on the top of the box. The value on the lower left edge of the box is the x-y-size. The copied feature maps are shown in white boxes, and the arrows refer to different operations [16]

4 Unsupervised deep learning based segmentation

CNNs have recently been used in semantic segmentation [15, 17, 29]. However, as pixel-level annotations are considered difficult to get for image segmentation, weakly supervised learning approaches that utilize object bounding boxes [30,

31] or image-level class labels [32]-[35] are widely used in the training process [36]. In this study, we use a CNN for image segmentation in a fully unsupervised manner as proposed by Kanezaki [36]. No training data and labels thus need to be prepared beforehand. This unsupervised setting could be crucial when it comes to skin lesion images that suffer from shortage in sufficient amount of data with defined ground truth that can be used for training a CNN.

Let $\{i_n \in R^p\}_{n=1}^N$ be a set of p -dimensional feature vectors of image pixels, such that N refers to the number of pixels in the input image. Cluster labels $\{c_n \in Z\}_{n=1}^N$ are assigned to all pixels by $c_n = f(i_n)$, where $f : R^p \rightarrow Z$ is a mapping function. The parameters f and $\{i_n\}$ are trained in a fully unsupervised manner, predicting an unknown $\{c_n\}$.

Three constraints need to be met for the prediction of $\{c_n\}$: (i) feature similarity, (ii) spatial continuity, and (iii) number of unique cluster labels. In the *feature similarity* constraint, pixels with similar features are assigned the same label. A p -dimensional feature map $\{i_n\}$ (feature extraction) is computed from the input image through M convolutional components that are composed of a 2D convolution, p filters of size 3×3 , ReLU activation function, and a batch normalization function (the batch is the N pixels in the input image). A linear classifier is applied to obtain a response map $\{y_n = W_c i_n + b_c\}_{n=1}^N$, where $W_c \in R^{q \times p}$ and $b_c \in R^q$, which is eventually normalized to $\{y'_n\}_{n=1}^N$ with zero mean and unit variance. The cluster label c_n for each pixel is obtained by selecting the dimension that has the maximum value in $\{y'_n\}$ (argmax classification). Thus, the process clusters the feature vectors into q clusters and each pixel is assigned to the closest point amongst the q representative points, grouping similar pixels into clusters.

Based on the second constraint, the clusters of image pixels have to be *spatially continuous* (cluster labels of neighboring pixels are identical). Here, K superpixels (a group of connected pixels with similar colors) $\{S_k \in R^n\}_{k=1}^K$ are extracted from the input image using SLIC [37], where S_k refers to the set of pixel indices that belong to the k^{th} superpixel, and each pixel belonging to some superpixel will have the same cluster label. The most frequent cluster label c_{max} for all $c_n \in \{1, \dots, q\}$ is then selected.

While the first two criteria aid in the grouping of pixels, the *third* criterion poses a constraint on the number of unique cluster labels, avoiding by that any undersegmentation (i.e. only one cluster) as this criterion gives preference to a large number of clusters.

The network starts by predicting the cluster labels with fixed network parameters (forward process of a network followed by superpixel refinement), and then training the network parameters with the fixed predicted cluster labels (backward process of a network based on gradient descent). Similar to supervised learning, the softmax loss is calculated between the responses $\{y'_n\}$ and the refined cluster labels $\{c'_n\}$, and error signals are backpropagated in order to update the parameters of both the classifiers and convolutional filters. This

forward-backward process is iterated T times to obtain the final prediction of cluster labels $\{c_n\}$. Fig.2 illustrates the unsupervised deep learning approach.

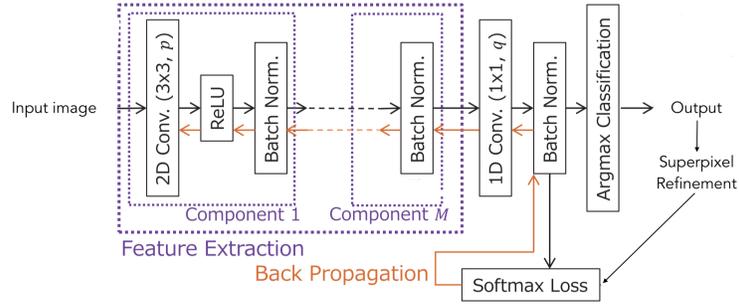


Fig. 2. Unsupervised deep learning based approach [38]: the network starts by predicting the cluster labels with fixed parameters in the forward process, such that in feature extraction a p -dimensional feature map is computed through M convolutional components. A linear classifier (argmax classification) is then applied to obtain a response map, eventually resulting in a cluster label for each pixel. The superpixel refinement step makes sure that the cluster labels of neighboring pixels are identical. The network parameters are then trained with the fixed predicted cluster labels, and the softmax loss is calculated between the response and refined cluster labels (resembles supervised learning). The process is repeated T times to obtain the final cluster labels predictions

5 Results and discussion

The U-Net architecture is trained on 2344 dermoscopy images along with their corresponding ground truth response masks from the "ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection" grand challenge datasets [38, 39]. Fig. 3 depicts some examples on the training dataset and the corresponding ground truth.

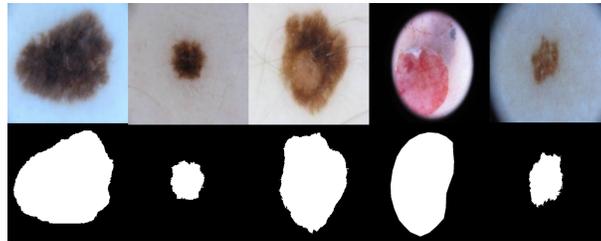


Fig. 3. Training data examples

To make the most out of the training data, augmentation using some transformations has been applied (such as rotation, shifting horizontally and vertically,

shearing, zooming, horizontal flip, and filling in newly created pixels which can appear after a rotation or a horizontal/vertical shift). This avoids the model from over-fitting and leads to better generalization. Images used have been resized to 512×512 . The U-Net model was trained for 20 epochs on a Tesla P100 GPU. Fig. 4 shows how accuracy has improved across the epochs.

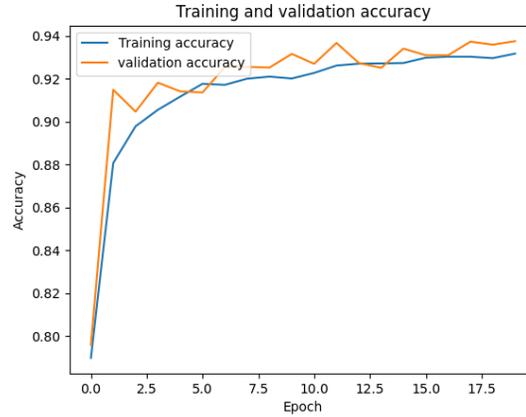


Fig. 4. Accuracy improvement across different epochs

Different parameters have been set to the unsupervised deep learning approach (number of channels 100, iterations 1000, clusters 2, learning rate 0.1, convolutional layers 2, super-pixels 10000, and compactness of superpixels 100), and the approach has also been run on a Tesla P100 GPU.

The approaches were tested on 250 images for segmentation. Fig. 5 shows some examples on the tested images, ground truth, and the results of both approaches. Results show that the unsupervised approach shows less quality results than those of the U-Net. However, this is not the case with some other test images where the unsupervised approach could be better in detecting the fine structure of the skin lesion. Fig. 6 highlights some examples on such cases.

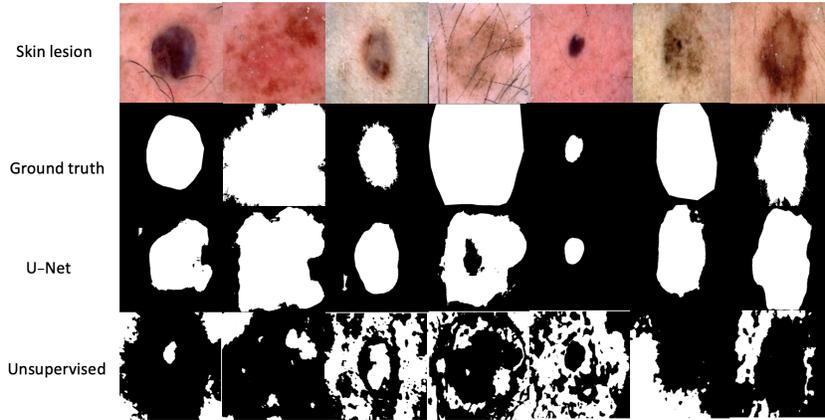


Fig. 5. The unsupervised deep learning based approach shows less quality results than U-Net

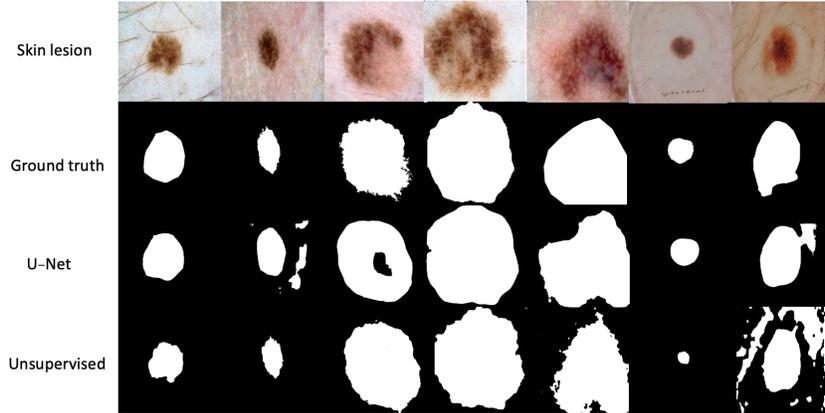


Fig. 6. The unsupervised deep learning based approach can be better in detecting fine structures than U-Net in some cases

To quantitatively evaluate the two approaches against the ground truth, we use two metrics (Dice coefficient [40] and Jaccard Index [41]) to measure the similarity between two images. The Dice coefficient is a measure that finds the spatial overlap between two binary images, resulting in a value that lies between 0 (no overlap) and 1 (agree perfectly). The Dice coefficient can be defined as:

$$D = \frac{2|A \cap G|}{|A + G|} \times 100\% \quad (1)$$

where A is the algorithm output and G is the ground truth.

Jaccard index is popular and frequently used as a similarity index. The area of overlap J is calculated between the segmented binary image A and its ground truth G as shown in the following equation [42]:

$$J = \frac{|A \cap G|}{|A \cup G|} \times 100\% \quad (2)$$

Table 1 shows the Dice coefficient values for the U-Net and unsupervised results shown in Fig. 5, where image numbers reflect the order of the images from left to right. This applies to Table 2 that depicts the Jaccard index values of Fig. 5. Results of Fig. 6 are shown in Table 3 and Table 4, respectively.

Table 1. Dice coefficient <https://www.overleaf.com/project/5c17bd05cc3926770b513e37> values for the U-Net and unsupervised results shown in Fig. 5. The numbers represent the ordering of the images as read from left to right

Method	Image-1	Image-2	Image-3	Image-4	Image-5	Image-6	Image-7
U-Net	83.2%	86.2%	86.7%	77.5%	90.7%	78.5%	88.4%
Unsupervised	7.3%	13.1%	19.6%	33%	0%	18.3%	0.3%

Table 2. Jaccard index values for the U-Net and unsupervised results shown in Fig. 5

Method	Image-1	Image-2	Image-3	Image-4	Image-5	Image-6	Image-7
U-Net	71.2%	75.7%	76.5%	63.3%	83%	64.6%	79.2%
Unsupervised	3.8%	7%	10.8%	19.7%	0%	10%	0.2%

Table 3. Dice coefficient values for the U-Net and unsupervised results shown in Fig. 6

Method	Image-1	Image-2	Image-3	Image-4	Image-5	Image-6	Image-7
U-Net	94.7%	68.6%	89.8%	96.9%	91.4%	92.4%	85%
Unsupervised	75.5%	85.4%	94.4%	97%	89.8%	37.4%	52.4%

Table 4. Jaccard index values for the U-Net and unsupervised results shown in Fig. 6

Method	Image-1	Image-2	Image-3	Image-4	Image-5	Image-6	Image-7
U-Net	89.9%	52.2%	81.4%	94%	84.2%	85.8%	74%
Unsupervised	60.6%	74.5%	89.4%	94.3%	81.5%	23%	35.5%

The average Dice coefficient and Jaccard index values for the two approaches when applied on 250 test images are summarized in Table 5, which shows that U-Net performs much better than the unsupervised approach in skin lesion segmentation.

Table 5. Average Dice and Jaccard values for U-Net and the unsupervised deep learning based approach

Method	Dice	Jaccard
U-Net	77.7%	67.2%
Unsupervised	40%	30.4%

We have tried different variations of the two approaches, for which the evaluations are depicted in Table 6. For the *unsupervised based approach*, two variations have been made. The first variation consists of: (i) sigmoid is used instead of ReLU as an activation function in the convolutional components; (ii) quick shift clustering [43] has been used instead of SLIC; (iii) adam has been used as an optimization function instead of the stochastic gradient descent. The second variation consists of: (i) SELU (Scaled Exponential Linear Unit) [45] is used instead of ReLU; (ii) using Felzenszwalb’s efficient graph based image segmentation method [44] instead of SLIC.

For *U-Net*, two variations have been made. The first variation is using tanh instead of ReLU as the activation function in the convolutional layers on both sides of the architecture (i.e. down and up). The second variation is using SELU the an activation functions instead of ReLU on both sides of the architecture. Results show that the default setting of U-Net performs better on the test images (in Tables 1-4), and that the unsupervised approach could be improved using some variations.

Table 6. Average Dice and Jaccard values for diiferent variations of U-Net and the unsupervised deep learning based approach

Method	Dice	Jaccard
Unsupervised (sigmoid-quickshift-adam)	52.3%	41.8%
Unsupervised (selu-felzenszwalb-sgd)	54.3%	44%
U-Net(tanh)	68%	57.4%
U-Net(selu)	67.2%	55%

6 Conclusion

In this paper we compare a supervised deep learning based approach (U-Net) with an unsupervised deep learning based approach for skin lesion segmentation in dermoscopy images. The unsupervised approach is able to detect fine structures in skin lesions better than U-Net in some test samples. However, U-Net shows to provide much better accuracy in terms of Dice coefficient and Jaccard index (77.7% vs. 40% and 67.2% vs. 30.4%, respectively). A proposed modification to the unsupervised approach causes the Dice and Jaccard values to improve to 54.3% and 44%, respectively. As a future work, we plan to combine both supervised and unsupervised deep learning based approaches (ensemble) for skin

lesion segmentation, such that the output of one approach could be used as an input to the other approach.

References

1. About Skin Cancer [cited 2018 Dec 17]. Available from: <http://www.canadianskincancerfoundation.com/about-skin-cancer.html>
2. M. Burroni, R. Corona, G. DellEva, F. Sera, R. Bono, P. Puddu, R. Perotti, F. Nobile, L. Andreassi, P. Rubegni.: Melanoma computer-aided diagnosis: reliability and feasibility study. *Clin Cancer Res*, 10(6) (2004)
3. Schmid-Saugeona P., Guillod J., and Thirana J.: Towards a computer-aided diagnosis system for pigmented skin lesions. *Comput Med Imaging Graph*, 27(1):6578, (2003)
4. Silveira M. and Nascimento J.: Comparison of segmentation methods for melanoma diagnosis in dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing*, 3(1), (2009)
5. Zeng, G., Zheng, G.: Multi-scale fully convolutional denseNets for automated skin lesion segmentation in dermoscopy images. *International Conference Image Analysis and Recognition*, pp.513-521 (2018)
6. Peruch, F., et al.: Simpler, faster, more accurate melanocytic lesion segmentation through MEDS. *IEEE Trans. Biomed. Eng.* 61(2), 557565 (2014)
7. Abbas, Q., et al.: Unified approach for lesion border detection based on mixture modeling and local entropy thresholding. *Skin Res. Technol.* 19(3), 314319 (2013)
8. Mete, M., Sirakov, N.M.: Lesion detection in dermoscopy images with novel density-based and active contour approaches. *BMC Bioinform.* 11(Suppl. 6), s23 (2010)
9. Sadri, A.R., et al.: Segmentation of dermoscopy images using wavelet networks. *IEEE Trans. Biomed. Eng.* 60(4), 11341141 (2013)
10. Jafari1, M.H., Karimi1, N., Nasr-Esfahani1, E., Samavi1, S., Soroushmehr, S.M.R., Ward, K., Najarian, K.: Skin lesion segmentation in clinical images using deep learning. *23rd International Conference on Pattern Recognition (ICPR)* (2016)
11. Attia, M., Hossny, M., Nahavandi, S., Yazdabadi, A.: Skin melanoma segmentation using recurrent and convolutional neural networks. In *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 292296 (2017)
12. Yuan, Y., Chao, M., Lo, Y.C.: Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. *IEEE Trans. Med. Imaging* (2017)
13. Al-Masni, M., Al-Antari, M., Choi, M., Han, S., Kim, T.: Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Computer Methods and Programs in Biomedicine* 162 221231 (2018)
14. Gurum Munirathnam, V., Yarlapati Ganesh, N., Little, S., O'Connor, N.: A deep residual architecture for skin lesion segmentation. In: *ISIC Skin Image Analysis Workshop and Challenge at MICCAI* (2018)
15. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. *CoRR*, abs/1411.4038 (2014)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In *MICCAI* (2015)
17. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *arXiv:1511.00561v2 [cs.CV]* (2015)

18. Bozorgtabar, B., Ge, Z., Chakravorty, R., Abedini, M., Demyanov, S., Garnavi, R.: Investigating deep side layers for skin lesion segmentation. In 14th International Symposium on Biomedical Imaging, pages 256-260 (2017)
19. He, X., Yu, Z., Wang, T., Lei, B.: Skin lesion segmentation via deep RefineNet. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer; p. 303-311 (2017)
20. Yuan, Y., Lo, Y.C.: Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks. *IEEE J. Biomed. Health Inform* (2017)
21. Deng, Z., Fan, H., Xie, F., Cui, Y., Liu, J.: Segmentation of dermoscopy images based on fully convolutional neural network. In Proc. IEEE Int. Conf. Image Process. (ICIP), pp.1732-1736 (2017)
22. Bi, L., Kim, J., Ahn, E., Kumar, A., Feng, D.: Dermoscopic image segmentation via multistage fully convolutional networks. *IEEE Trans. Biomed. Eng.*, vol.64, no.9, pp.2065-2074 (2017)
23. Mishra, R., Daescu, O: Deep learning for skin lesion segmentation. In IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pages 11891194 (2017)
24. Burdick, J., et al.: Rethinking skin lesion segmentation in a convolutional classifier. *Journal of digital imaging*. pp.1-6 (2017)
25. Li, X., Yu, L., Chen, H., Fu, C.-W., Heng, P.-A.: Semisupervised skin lesion segmentation via transformation consistent self-ensembling model. *BMVC* (2018)
26. Vesal, S., Patil, S., Ravikumar, N., Maier, A.K.: A multi-task framework for skin lesion detection and segmentation. Workshop on skin image analysis, MICCAI (2018)
27. Yang, T., Peng, S., Hu, P., Huang, L.: Pigmented skin lesion segmentation based on random forest and full convolutional neural networks. *Proc. SPIE* (2018)
28. Izadi, S., Mirikharaji, Z., Kawahara, J., Hamarneh, G.: Generative adversarial networks to segment skin lesions. In IEEE 15th International Symposium on Biomedical Imaging, pages 881884 (2018)
29. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.: Semantic image segmentation with deep convolutional nets and fully connected crfs. In ICLR (2015)
30. Zhu, J., Mao, J., Yuille, A.: Learning from weakly supervised data by the expectation loss svm (e-svm) algorithm. In NIPS (2014)
31. Chang, F., Lin, Y., Hsu, K.: Multiple structured-instance learning for semantic segmentation with uncertain training data. In CVPR (2014)
32. Pathak, D., Krahenbuhl, P., Darrell, T.: Constrained convolutional neural networks for weakly supervised segmentation. In ICCV (2015)
33. Pourian, N., Karthikeyan, S., Manjunath, B.: Weakly supervised graph based semantic segmentation by learning communities of image-parts. In ICCV (2015)
34. Shi, Z., Yang, Y., Hospedales, T., Xiang, T.: Weakly-supervised image annotation and segmentation with objects and attributes. *IEEE Trans. on Pattern Analysis and Machine Intelligence* (2016)
35. Shimoda, W., Yanai, K.: Distinct class-specific saliency maps for weakly supervised semantic segmentation. In ECCV (2016)
36. Kanezaki, A.: Unsupervised image segmentation by backpropagation. <http://sigport.org/2710> IEEE SigPort, 2018. Last accessed 18 Dec 2018
37. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Ssstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.34, no.11 (2012)
38. Codella, N., Gutman, D., Emre Celebi, M., Helba, B., Marchetti, M., Dusza, S., Kallou, A., Liopyris, K., Mishra, N., Kittler, H., Halpern, A.: 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), arXiv:1710.05006 (2017)
39. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* 5, 180161 doi:10.1038/sdata.2018.161 (2018)
 40. Dice, R.: Measures of the amount of ecologic association between species. *Ecology*. 26 (3): 297-302. doi:10.2307/1932409. JSTOR 1932409 (1945)
 41. Jaccard, P.: Etude comparative de la distribution orale dans une portion des alpes et des jura. *Bulletin de la Societ Vaudoise des Sciences Naturelles*, 37, 547579 (1901)
 42. Narayana, C., Sreenivasa Reddy, E., Seetharama Prasad, M.: Automatic image segmentation using ultrafuzziness. *International Journal of Computers and Applications*, 49(12), 613. doi:10.5120/7677-0977 (2012)
 43. Vedaldi A., Soatto, S.: Quick shift and kernel methods for mode seeking. In *Proc. ECCV* (2008)
 44. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. *International Journal of Computer Vision (IJCV)*, 59(2):167181 (2004)
 45. Klambauer, G., Unterthiner, T., Mayr, A, Hochreiter, S.: Self-normalizing neural networks. arXiv preprint arXiv:1706.02515 (2017)