Skin lesion segmentation with deep learning

Jane Lameski, Andrej Jovanov, Eftim Zdravevski, Petre Lameski, Sonja Gievska
Faculty of Computer Science and Engineering
Ss.Cyril and Methodius University, Skopje, Macedonia
Email: jane.lameski96@gmail, andrej.jovanov@students.finki.ukim.mk
eftim.zdravevski@finki.ukim.mk, petre.lameski@finki.ukim.mk, sonja.gievska@finki.ukim.mk

***Abstract*—Skin lesion segmentation is an important process in skin diagnostics because it improves manual and computer-aided diagnostics by focusing the medical personnel on specific parts of the skin. Image segmentation is a common task in computer vision that partitions a digital image into multiple segments, for which deep neural networks have been proven to be reliable. In this paper, we investigate the applicability of deep learning methods for skin lesion segmentation evaluating three architectures: a pre-trained VGG16 encoder combined with SegNet decoder, TernausNet, and DeepLabV3+. The data set consists of images with RGB skin lesions and the ground truth of their segmentation. All the image sizes vary from hundreds to thousands of pixels per dimension. We evaluated the approaches with the Jaccard index and the computational efficiency of the training. The results show that the three deep neural network architectures achieve Jaccard Index scores of above 0.82, while the DeeplabV3+ outperforms the other approaches with a score of 0.876. The results are encouraging and can lead to fully-fledged automated approaches for skin lesion segmentation.**

***Keywords*—Deep Learning, Neural Networks, Segmentation, Skin Lesion, Melanoma**

I. Introduction

Automatic and accurate segmentation of lesions is essential for studying the progress of a particular disease, diagnosing patients, monitoring treatments and assessing clinical trials for new treatments. Lesion segmentation is crucial for many tasks about medical diagnosis, from the detection of skin lesions [1], [2] to identification of brain lesions in 2D to identification of mass lesions in mammogram images. With semantic segmentation, the objects within an image are segmented and classified based on a concept. It can be performed in 2D images, videos or even 3D data and is still a challenging task. Semantic segmentation is of great importance because it is a critical pre-processing phase for other tasks such as object detection, scene understanding, and parsing. Deep learning has been used for semantic segmentation of objects across many domains, from agriculture [3], [4] to medicine.

Recently, convolutional neural networks (CNNs) have achieved remarkable results in computer vision tasks. Many approaches for semantic segmentation are based on CNNs in which each pixel is labeled with the class of its enclosing object or region.

This paper addresses the use of different deep learning models, their design and their segmentation results when applied to the problem of lesion segmentation. We apply the state-of-the-art semantic segmentation approaches in this domain and present the obtained results. Additionally, several suitable pre- and post-processing techniques have been applied. The research studies most related to ours are discussed in the next section, followed by a description of our approach and a discussion of the results obtained on [2].

II. Related work

Traditionally, image segmentation was performed using image processing techniques. The authors in [5] have presented some image processing approaches such as Region-based segmentation, covering both threshold segmentation and region growth segmentation, and segmentation based on clustering. One of the most used threshold segmentation algorithms is the largest inter-class variance method (Otsu). Otsu finds the optimal threshold by maximizing the variance between classes. Region growth method is serial region segmentation algorithm, which selects pixels having similar properties and combines them into one region. It is a sound segmentation approach because it separates the connected regions with the same characteristics and provides useful boundary information and segmentation results. Segmentation based on clustering most commonly uses the K-means algorithm to gather samples into different clusters according to the distance.

Authors in [6] review state-of-the-art CNNs. The pioneer of deep CNNs, AlexNet, won ILSVCR-TOP-5 with test accuracy of 84.6%, compared with 73.8% from the runner-up using traditional techniques.

Visual Geometry Group (VGG) [7] is a model introduced by Visual Geometry Group from the University of Oxford. VGG16, a CNN composed of 16 layers, achieved 92.7% accuracy on ILSVCR Caltech-101 dataset which contains 9K images labeled into 102 classes (101 object categories and a background class) with TOP-5 test accuracy.

Microsoft's ResNet [8] is an exceptional CNN that won ILSVRC-2016 with 96.4% accuracy. The network is well-known due to the depth (152 layers) and the introduction to residual blocks. The idea in this approach is to ensure that the next layer learns new weights different from what the input already encoded. The connections that we encounter in this model overcome the vanishing gradient problem. Two approaches for image segmentation of breast cancer images have been presented in [9]. The first one segments the images with an automated region growing using threshold by a trained artificial neural network (ANN). The second approach utilizes a cellular neural network (CNN), whose parameters

978-1-5386-9301-8/19/$31.00

are obtained by a genetic algorithm (GA). The approach was applied on a breast cancer images dataset.

The architectures proposed and evaluated in our study have been inspired by previous work of other research in the filed. In particular, the outstanding results of an 11-layer CNN architecture evaluated on a task related to ours, namely segmentation of 3D MRI images of brain lesions [10], have inspired one of our models.

Another promising CNN 3D segmentation model V-Net, trained end to end on MRI volumes representing prostate, proposed by [11] have shown good performances on demanding dataset such as [12]. This model using Dice-based loss has received a score of 82.39, which are comparable to [13] who were the winners of that competition with a score of 85.72.

III. Methods

A. Dataset

The models for melanoma detection proposed in this paper have been evaluated on the ISIC 2018 competition: "Skin Lesion Analysis Towards Melanoma Detection" [2], Task 1 dataset for "Lesion Boundary Segmentation", which contains 2594 images and their corresponding ground truth masks. We split this training set into 70% and 30% subsets for training and testing, respectively, using the same splits for evaluating each approach. Several preprocessing techniques were needed to face the challenge of having images of varying size and type.

B. Pre-processing

Border replication has been used as a padding technique to standardize images to the same size, as shown in Figure 1. We should note that alternative padding techniques were not explored, although the selection of the most suitable method for the task at hand might be a direction for future research.

Related research, such as [14], suggests that data augmentation could lead to performance gains when training deep learning models on a limited number of data. To mitigate the problem of the small dataset, we performed online augmentation, in line with the recommendations described in [15]. We used various types of transformations of the original images, such as rotation, translation, flipping, etc., resulting in an increased number of images corresponding to the training iterations.

C. Models

The success of TernausNet model [16] at the Kaggle's "Carvana Image Masking Challenge", a competition of models for removing the photo studio background in vehicle images [17], inspired **Model 1**. We utilized a modified version of the original model, which combines VGG16 with U-net. Instead of a pre-trained VGG11, we used a pre-trained VGG16 with ImageNet weights as an encoder, we trained the decoder layers from U-net [18] and froze the pre-trained weights from VGG16.

Model 2 was created similarly, yet, instead of U-net, we used the state-of-the-art deep CNN architecture, SegNet [3]. It serves as a decoder architecture, although without the pooling indices of the initially proposed network in [3].

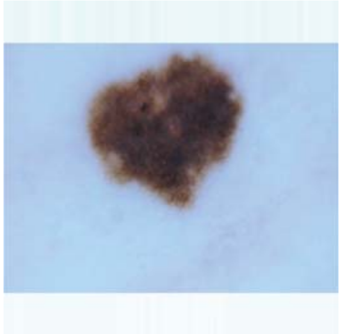


Fig. 1: Exemplary image with replicate border padding to equalize width and height

Model 3 is based on DeeplabV3 with backbone weights from MobileNetV2. DeeplabV3 is a state-of-the-art segmentation tool [19], which is a deep CNN such as VGG16, but changed in fully convolutional fashion, using atrous convolution or dilated convolution, introducing another parameter in the convolutional layers called the dilation rate. The rate is a spacing between the values in the kernel. A 5v5 kernel with dilation rate of 2 in an image is deleting every second column and row. *A bilinear interpolation stage enlarges the feature maps to the original image resolution*, as stated in [20]. After that, a fully connected Conditional random fields (CRFs) is applied to refine the segmentation results.

The idea for this approach is inspired by recent research by Google [21], which is an upgraded version of DeeplabV3. It uses the DeeplabV3 model, on top of a backbone CNN architecture [22]. Instead of MobileNet backbone architecture, we used the updated version MobileNetV2 [23].

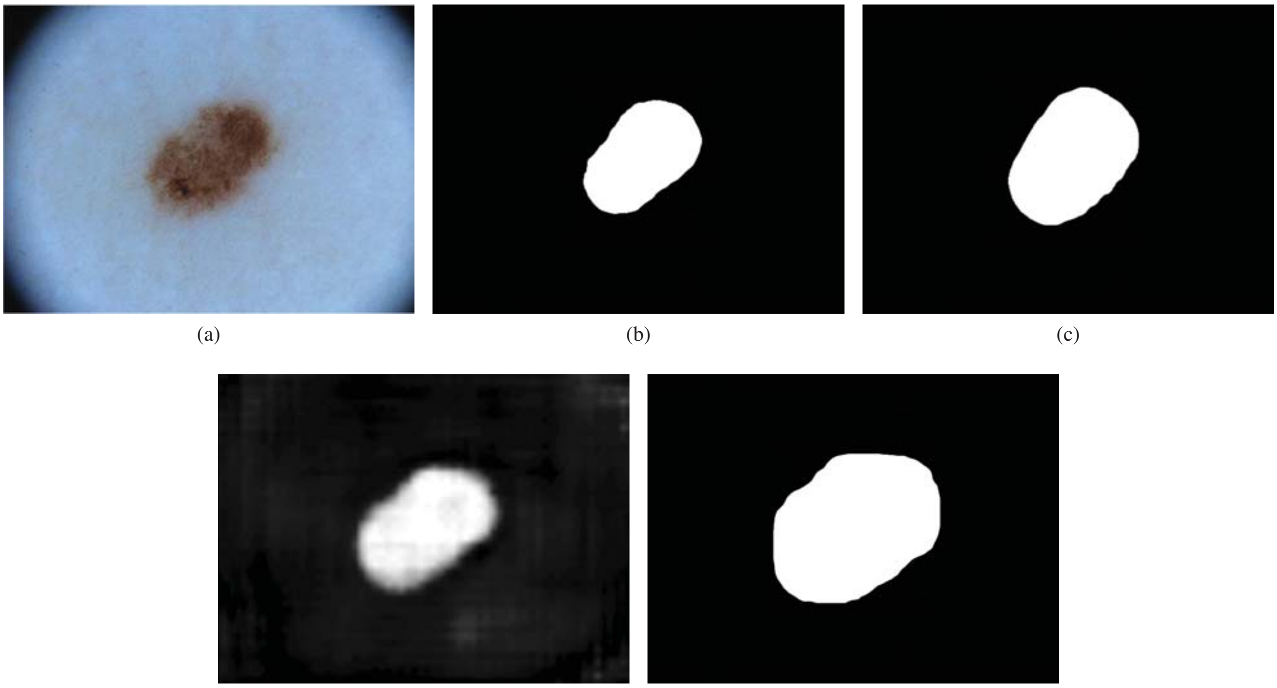
D. Post-processing of Model 3

After the segmentation was performed, the first two approaches yielded smooth masks. However, Model 3 introduced some background noise. To remove the noise, we used threshold in combination with multiple morphological dilatations (for the results shown in the Results section, we used 120 (0-255 pixel values) as a threshold value and 5 as times dilation). Figures 2(d) and 2(e) show the noisy segmentation mask, and the one obtained with the post-processing, respectively.

E. Evaluation metric

The three deep learning models were evaluated in a manner recommended by the ISIC 2018 competition: "Skin Lesion Analysis Towards Melanoma Detection". Namely, the final score for an image is the Jaccard index if it is above 0.65; otherwise, the score is 0. The Jaccard index, i.e., intersection over union (IoU), is defined with (1), where α is the ground truth, and β is the obtained segmented mask in the output image.

$$J(\alpha,\beta)=\frac{\mid \alpha\cap\beta \mid}{\mid \alpha\cup\beta \mid} \tag{1}$$



(a) (b) (c) (d) (e)

Fig. 2: Lesion image with segmentation results:
(a) Input Image (b) Model 2 (c) Model 1
(d) Model 3 before processing (e) Model 3 after processing

After the resulting masks are produced from the deep learning models, the size is restored to their original so that we can test the results with the ground truth using the Jaccard index.

IV. Results

TABLE I: Training time and Jaccard scored of the used deep learning models

Model name	Training time	Steps and Epochs	Jaccard Index score
Model 1	**5540 sec**	60 Steps 100 Epochs	0.855
Model 2	5663 sec	60 Steps 100 Epochs	0.821
Model 3	5550 sec	60 Steps 100 Epochs	**0.876**

The number of epochs and the computational time needed for training of each of the three models are presented in Table 1. Although there can be an improvement for some of the models if we change the number of epochs, we kept the same, so that we can have an efficient comparison.

The accuracy and the loss across the training epoch yielded by the three deep learning models are shown in Figure 3.

Our current efforts are on improving the results by experimenting with different post-processing and backbone weights for Model 3. We are also considering using pooling indices as a recommended approach to Model 2.

V. Conclusion

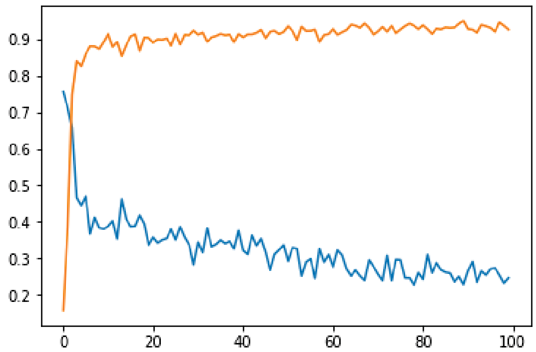
Automatic lesion segmentation is of great importance, because of its greater accuracy and speed, when compared to manual analysis. Recently, deep neural networks have offered the best results for analyzing medical images, compared to traditional machine learning based approaches. In this paper, we evaluated three approaches based on deep learning. Out of these three, Model 3 has shown the best accuracy and Jaccard Index score.

The evaluated approaches are promising and could be used in practical applications, especially as more labeled images with skin lesions are collected. A key challenge is to provide correctly and consistently labeled datasets.

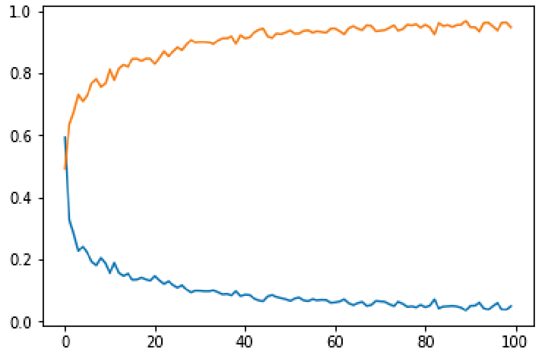
For improvement of Model 1, instead of pre-trained ImageNet weights in VGG16, one could explore other weights, which might improve the current score. For Model 2, we can include the pooling indices as stated in [3]. The Model 3 could be improved by experimenting with the post-processing and the backbone weights, by trying a different number of dilation cycles and threshold values for the post-processing, and by experimenting with other backbone architectures.

Acknowledgments

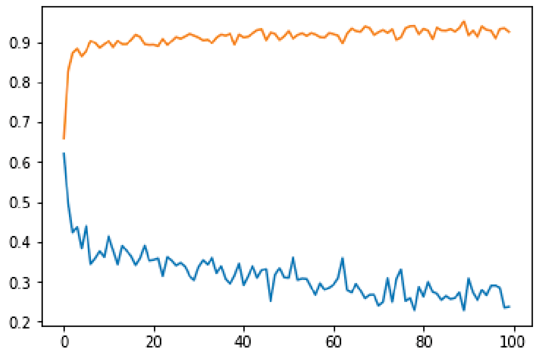
This work was partially financed by the Faculty of Computer Science and Engineering at the Ss. Cyril and Methodius University, Skopje, Macedonia. We also gratefully acknowledge the support of Microsoft Azure for Research and the NVIDIA



(a)



(b)



(c)

Fig. 3: Training performance plots. The X axis shows the number of epochs, and the Y axis is the loss/accuracy (loss is the darker and accuracy the lighter curve)

(a) Model 1 (b) Model 3 (c) Model 2

Corporation through grants providing GPU resources for this work.

REFERENCES

[1] P. Lameski, E. Zdravevski, S. Kalajdziski, V. Trajkova, and E. Hadzieva, “Computer-aided detection of melanoma, a case study,” in *Proceedigns of ETAI 2018*, ETAI Society of Macedonia, 2018.

[2] “Isic 2018: Skin lesion analysis towards melanoma detection.” https://challenge2018.isic-archive.com/.

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, p. 24812495, 2017.

[4] P. Lameski, E. Zdravevski, V. Trajkovik, and A. Kulakov, “Weed detection dataset with rgb images taken under variable light conditions,” in *International Conference on ICT Innovations*, pp. 112–119, Springer, 2017.

[5] S. Yuheng and Y. Hao, “Image segmentation algorithms overview,” *CoRR*, vol. abs/1707.02051, 2017.

[6] A. Garcia-Garcia, S. Orts, S. Oprea, V. Villena-Martinez, and J. G. Rodríguez, “A review on deep learning techniques applied to semantic segmentation,” *CoRR*, vol. abs/1704.06857, 2017.

[7] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv*, 09 2014.

[8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[9] R. Rouhi, M. Jafari, S. Kasaei, and P. Keshavarzian, “Benign and malignant breast tumors classification based on region growing and cnn segmentation,” *Expert Systems with Applications*, vol. 42, no. 3, pp. 990 – 1002, 2015.

[10] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, “Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation,” *Medical Image Analysis*, vol. 36, pp. 61 – 78, 2017.

[11] F. Milletari, N. Navab, and S. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 Fourth International Conference on 3D Vision (3DV)*, Oct 2016.

[12] G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken, G. Vincent, G. Guillard, N. Birbeck, J. Zhang, *et al.*, “Evaluation of prostate segmentation algorithms for mri: the promise12 challenge,” *Medical image analysis*, vol. 18, no. 2, pp. 359–373, 2014.

[13] G. Vincent, G. Guillard, and M. Bowes, “Fully automatic segmentation of the prostate using active appearance models,” *MICCAI Grand Challenge: Prostate MR Image Segmentation*, vol. 2012, 2012.

[14] L. Perez and J. Wang, “The effectiveness of data augmentation in image classification using deep learning,” *arXiv preprint arXiv:1712.04621*, 2017.

[15] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *International conference on medical image computing and computer-assisted intervention*, pp. 424–432, Springer, 2016.

[16] V. Iglovikov and A. Shvets, “Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation,” *ArXiv e-prints*, 2018.

[17] “Carvana image masking challenge automatically identify the boundaries of the car in an image.” https://www.kaggle.com/c/carvana-image-masking-challenge/.

[18] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, p. 234241, 2015.

[19] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017.

[20] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 834–848, April 2018.

[21] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," *Lecture Notes in Computer Science*, p. 833851, 2018.

[22] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.