



OPEN

Fully automatic wound segmentation with deep convolutional neural networks

Chuanbo Wang¹, D. M. Anisuzzaman¹, Victor Williamson¹, Mrinal Kanti Dhar¹, Behrouz Rostami², Jeffrey Niezgodá³, Sandeep Gopalakrishnan⁴✉ & Zeyun Yu^{1,5}✉

Acute and chronic wounds have varying etiologies and are an economic burden to healthcare systems around the world. The advanced wound care market is expected to exceed \$22 billion by 2024. Wound care professionals rely heavily on images and image documentation for proper diagnosis and treatment. Unfortunately lack of expertise can lead to improper diagnosis of wound etiology and inaccurate wound management and documentation. Fully automatic segmentation of wound areas in natural images is an important part of the diagnosis and care protocol since it is crucial to measure the area of the wound and provide quantitative parameters in the treatment. Various deep learning models have gained success in image analysis including semantic segmentation. This manuscript proposes a novel convolutional framework based on MobileNetV2 and connected component labelling to segment wound regions from natural images. The advantage of this model is its lightweight and less compute-intensive architecture. The performance is not compromised and is comparable to deeper neural networks. We build an annotated wound image dataset consisting of 1109 foot ulcer images from 889 patients to train and test the deep learning models. We demonstrate the effectiveness and mobility of our method by conducting comprehensive experiments and analyses on various segmentation neural networks. The full implementation is available at <https://github.com/uwm-bigdata/wound-segmentation>.

Acute and chronic nonhealing wounds represent a heavy burden to healthcare systems, affecting millions of patients around the world¹. In the United States, medicare cost projections for all wounds are estimated to be between \$28.1B and \$96.8B². Unlike acute wounds, chronic wounds fail to predictably progress through the phases of healing in an orderly and timely fashion, thus require hospitalization and additional treatment adding billions in cost for health services annually³. The shortage of well-trained wound care clinicians in primary and rural healthcare settings decreases the access and quality of care to millions of Americans. Accurate measurement of the wound area is critical to the evaluation and management of chronic wounds to monitor the wound healing trajectory and to determine future interventions. However, manual measurement is time-consuming and often inaccurate which can cause a negative impact on patients. Wound segmentation from images is a popular solution to these problems that not only automates the measurement of the wound area but also allows efficient data entry into the electronic medical record to enhance patient care.

Related studies on wound segmentation can be roughly categorized into two groups: traditional computer vision methods and deep learning methods. Studies in the first group focus on combining computer vision techniques and traditional machine learning approaches. These studies apply manually-designed feature extraction to build a dataset that is later used to support machine learning algorithms. Song et al. described 49 features that are extracted from a wound image using K-means clustering, edge detection, thresholding, and region growing in both grayscale and RGB⁴. These features are filtered and prepared into a feature vector that is used to train a Multi-Layer Perceptron (MLP) and a Radial Basis Function (RBF) neural network to identify the region of a chronic wound. Ahmad et al. proposed generating a Red-Yellow-Black-White (RYKW) probability map of an input image with a modified hue-saturation-value (HSV) model⁵. This map then guides the segmentation process

¹Big Data Analytics and Visualization Laboratory, Department of Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI, USA. ²Department of Electrical Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI, USA. ³Advancing the Zenith of Healthcare (AZH) Wound and Vascular Center, Milwaukee, WI, USA. ⁴College of Nursing, University of Wisconsin Milwaukee, Milwaukee, WI 53211, USA. ⁵Department of Biomedical Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI, USA. ✉email: sandeep@uwm.edu; yuz@uwm.edu

using either optimal thresholding or region growing. Hettiarachchi et al. demonstrated an energy minimizing discrete dynamic contour algorithm applied on the saturation plane of the image in its HSV color model⁶. The wound area is then calculated from a flood fill inside the enclosed contour. Hani et al. proposed applying an Independent Component Analysis (ICA) algorithm to the pre-processed RGB images to generate hemoglobin-based images, which are used as input of K-means clustering to segment the granulation tissue from the wound images⁷. These segmented areas are utilized as an assessment of the early stages of ulcer healing by detecting the growth of granulation tissue on ulcer bed. Wantanajittikul et al. proposed a similar system to segment the burn wound from images⁸. Cr-Transformation and Luv-Transformation are applied to the input images to remove the background and highlight the wound region. The transformed images are segmented with a pixel-wise Fuzzy C-mean Clustering (FCM) algorithm. These methods suffer from at least one of the following limitations: (1) as in many computer vision systems, the hand-crafted features are affected by skin pigmentation, illumination, and image resolution, (2) they depend on manually tuned parameters and empirically handcrafted features which does not guarantee an optimal result. Additionally, they are not immune to severe pathologies and rare cases, which are very impractical from a clinical perspective, and (3) the performance is evaluated on a small biased dataset.

Since the successes AlexNet⁹ achieved in the 2012 Imagenet large scale visual recognition challenge¹⁰, the application of deep learning¹¹ in the domain of computer vision sparked interests in semantic segmentation¹² using deep convolutional neural networks (CNN)¹³. Typically, traditional machine learning and computer vision methods make decisions based on feature extraction. To segment the region of interest, one must guess a set of important features and then handcraft sophisticated algorithms that capture these features¹⁴. However, a CNN integrates feature extraction and decision making. The convolutional kernels of CNN extract the features and their importance is determined during the training of the network. In a typical CNN architecture, the input are processed by a sequence of convolutional layers and the output is generated by a fully connected layer that requires fixed-sized input. One successful variant of CNN is fully convolutional neural networks (FCN)¹⁵. A FCN is composed of convolutional layers without a fully connected layer as the output layer. This allows arbitrary input sizes and prevents the loss of spatial information caused by the fully connected layers in CNNs. Several FCN-based methods have been proposed to solve the wound segmentation problem. For example, Wang et al. estimated the wound area by segmenting wounds¹⁶ with the vanilla FCN architecture¹⁵. With time-series data consisting of the estimated wound areas and corresponding images, wound healing progress is predicted using a Gaussian process regression function model. However, the mean Dice accuracy of the segmentation is only evaluated to be 64.2%. Goyal et al. proposed to employ the FCN-16 architecture on the wound images in a pixel-wise manner that each pixel of an image is predicted to which class it belongs¹⁷. The segmentation result is simply derived from the pixels classified as a wound. By testing different FCN architectures they are able to achieve a Dice coefficient of 79.4% on their dataset. However, the network's segmentation accuracy is limited in distinguishing small wounds and wounds with irregular borders as the tendency is to draw smooth contours. Liu et al. proposed a new FCN architecture that replaces the decoder of the vanilla FCN with a skip-layer concatenation upsampled with bilinear interpolation¹⁸. A pixel-wise softmax layer is appended to the end of the network to produce a probability map, which is post-processed to be the final segmentation. A dice accuracy of 91.6% is achieved on their dataset with 950 images taken under an uncontrolled lighting environment with a complex background. However, images in their dataset are semi-automatically annotated using a watershed algorithm. This means that the deep learning model is learning how the watershed algorithm labels wounds as opposed to human specialists.

To better explore the capacity of deep learning on the wound segmentation problem, we propose an efficient and accurate framework to automatically segment wound regions. The segmentation network of this framework is built above MobileNetsV2¹⁹. This network is light-weight and computationally efficient since significantly fewer parameters are used during the training process.

Our contributions can be summarized as follows:

1. We build a large dataset of wound images with segmentation annotations done by wound specialists. This is by far the largest dataset focused on wound segmentation (to the best of our knowledge).
2. We propose a fully automatic wound segmentation framework based on MobileNetsV2 that balances computational efficiency and accuracy.
3. Our proposed framework shows high efficiency and accuracy in wound image segmentation.

Dataset

Dataset construction. There is currently no public dataset large enough for training deep-learning-based models for wound segmentation. To explore the effectiveness of wound segmentation using deep learning models, we collaborated with the Advancing the Zenith of Healthcare (AZH) Wound and Vascular Center, Milwaukee, WI. Our chronic wound dataset was collected over 2 years at the center and includes 1109 foot ulcer images taken from 889 patients during multiple clinical visits. The raw images were taken by Canon SX 620 HS digital camera and iPad Pro under uncontrolled illumination conditions, with various backgrounds. Figure 1 shows some sample images in our dataset.

The raw images collected are of various sizes and cannot be fed into our deep learning model directly since our model requires fixed-size input images. To unify the size of images in our dataset, we first localize the wound by placing bounding boxes around the wound using an object localization model we trained de novo, YOLOv3²⁰. Our localization dataset contains 1010 images, which are also collected from the AZH Wound and Vascular Center. We augmented the images and built a training set containing 3645 images and a testing set containing 405 images. For training our model we have used LabelImg²¹ to manually label all the data (both for training

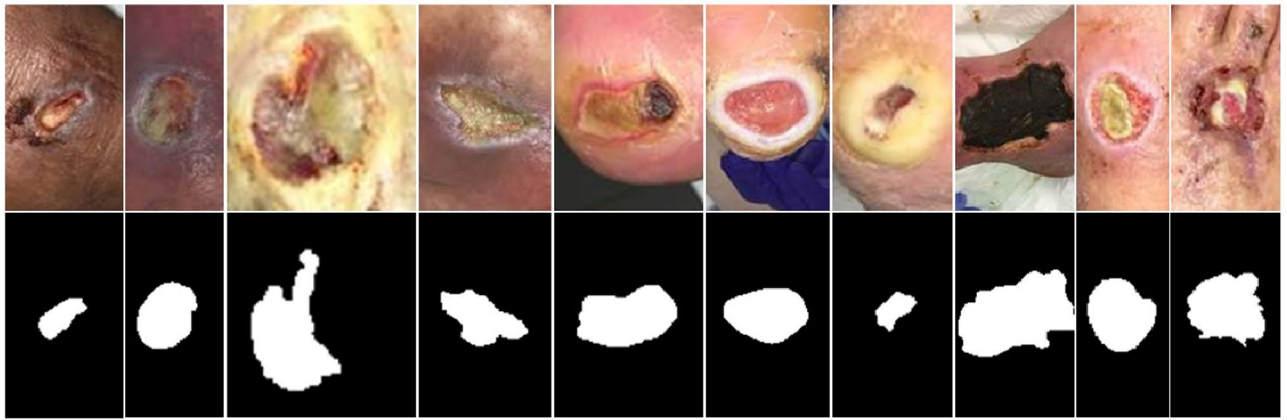


Figure 1. An illustration of images in our dataset. The first row contains the raw images collected. The second row consists of segmentation mask annotations we create with the AZH wound and vascular center.

and testing). The YOLO format has been used for image labelling. The model has been trained with a batch size of 8 for 273 epochs. With an intersection over union (IoU) rate of 0.5 and non-maximum suppression of 1.00, we get the mean Average Precision (mAP) value of 0.939. In the next step, image patches are cropped based on the bounding boxes result from the localization model. We unify the image size (224 pixels by 224 pixels) by applying zero-padding to these images, which are regarded in our dataset data points. We confirm that the data collected was de-identified and in accordance to relevant guidelines and regulations and the patient's informed consent is waived by the institutional review board of University of Wisconsin-Milwaukee.

Data annotation. During training, a deep learning model is learning the annotations of the training dataset. Thus, the quality of annotations is essential. Automatic annotation generated with computer vision algorithms is not ideal when deep learning models are trained to learn how human experts recognize the wound region. In our dataset, the images were manually annotated with segmentation masks that were further reviewed and verified by wound care specialists from the collaborating wound clinic. Initially only foot ulcer images were annotated and included in the dataset as these wounds tend to be smaller than other types of chronic wounds, which makes it easier and less time-consuming to manually annotate the pixel-wise segmentation masks. In the future we plan to create larger image libraries to include all types of chronic wounds, such as venous leg ulcers, pressure ulcers, and surgery wounds as well as non-wound reference images. The AZH Wound and Vascular Center, Milwaukee, WI, had consented to make our dataset publicly available.

Methods

In this section we describe our method with the architecture of the deep learning model for wound segmentation. The transfer learning used during the training of our model and the post-processing methods including hole filling and removal of small noises are also described. We confirm that the research is approved by the institutional review board of University of Wisconsin-Milwaukee.

Pre-processing. Besides cropping and zero-padding discussed in the dataset construction section, standard data augmentation techniques are applied to our dataset before being fed into the deep learning model. These image transformations include arbitrary rotations in the range of +25 to -25 degrees, random left-right and top-down flippings with a probability of 0.5, and random zooming within 80% of the original image area. Random zooming is performed as the only non-rigid transformation because we suspect that other non-rigid transformations like shearings do not represent common wound shape variations. Eventually, the training dataset is augmented to around 5000 images. We keep the validation dataset unaugmented to generate convincing evaluation outcomes.

Model architecture overview. A convolutional neural network (CNN), MobileNetV2¹⁹, is adopted to segment the wound from the images. Compared with conventional CNNs, this network substitutes the fundamental convolutional layers with depth-wise separable convolutional layers²² where each layer can be separated into a depth-wise convolution layer and a point-wise convolution layer. A depth-wise convolution performs lightweight filtering by applying a convolutional filter per input channel. A point-wise convolution is a 1×1 convolution responsible for building new features through linear combinations of the input channels. This substitution reduces the computational cost compared to traditional convolution layers by almost a factor of k^2 where k is the convolutional kernel size. Thus, depth-wise separable convolutions are much more computationally efficient than conventional convolutions suitable for mobile or embedded applications where computing resource is limited. For example, the mobility of MobileNetV2 could benefit medical professionals and patients by allowing instant wound segmentation and wound area measurement immediately after the photo is taken

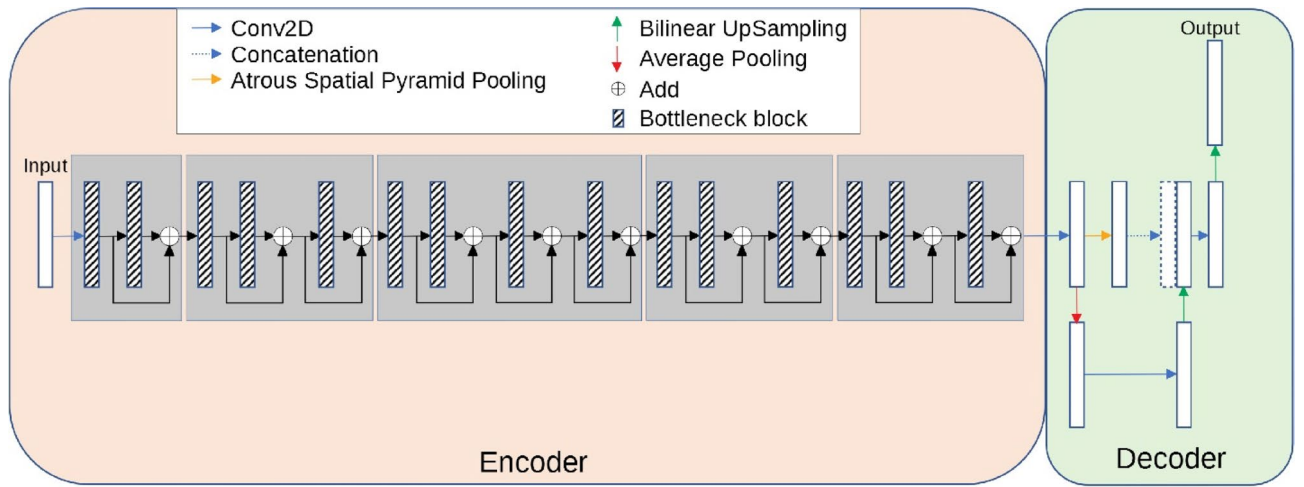


Figure 2. The encoder–decoder architecture of MobilenetV2¹⁹.

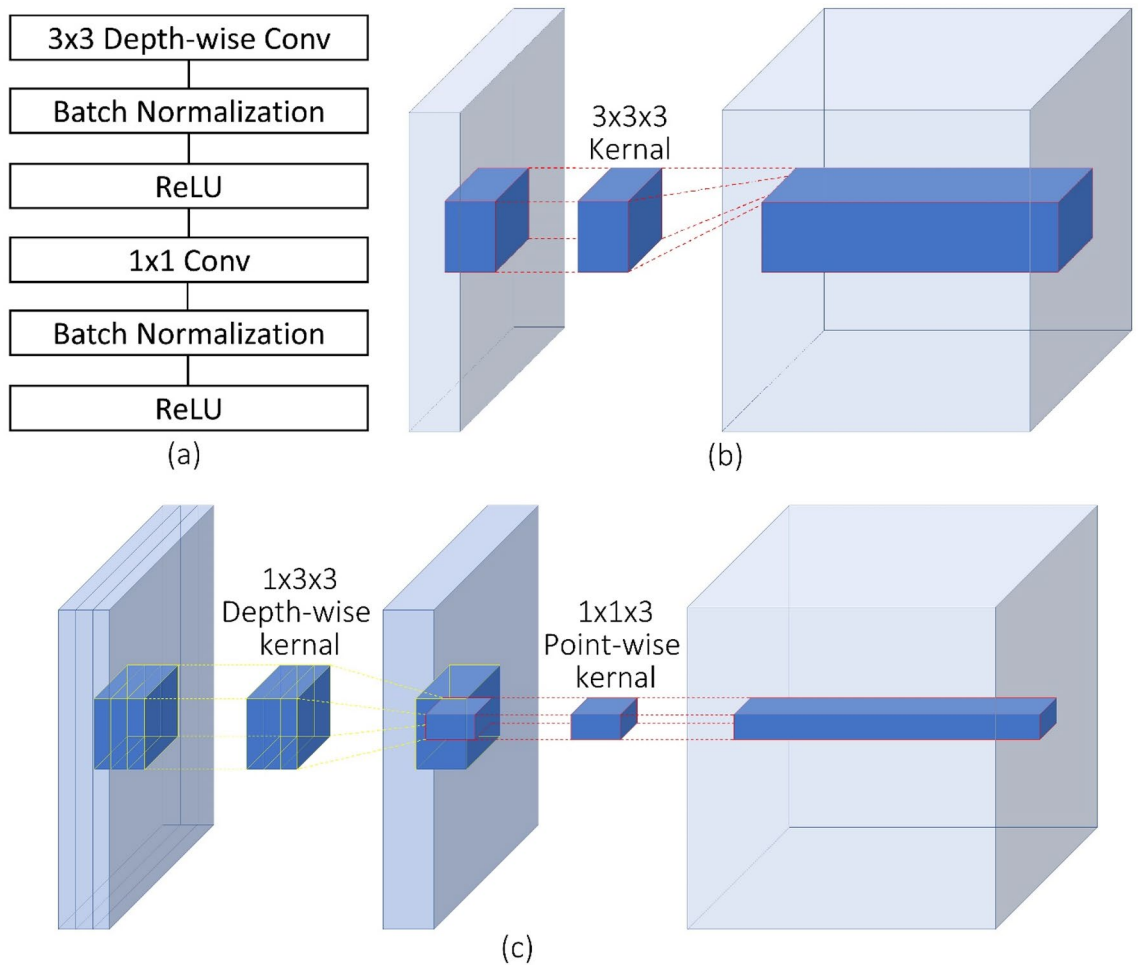


Figure 3. (a) A depth-separable convolution block. The block contains a 3×3 depth-wise convolutional layer and a 1×1 point-wise convolution layer. Each convolutional layer is followed by batch normalization and Relu6 activation. (b) An example of a convolution layer with a $3 \times 3 \times 3$ kernel. (c) An example of a depth-wise separable convolution layer equivalent to (b).

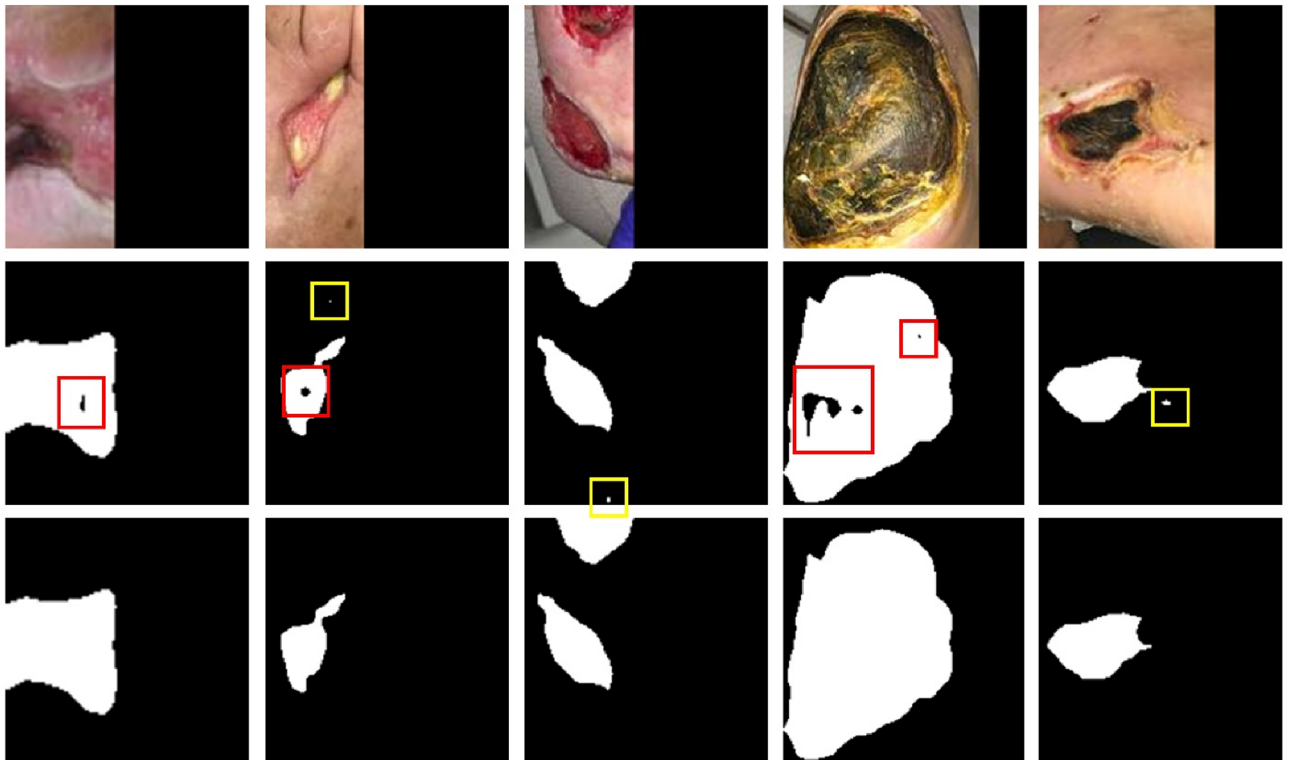


Figure 4. An illustration of the segmentation result and the post processing method. The first row illustrates images in the testing dataset. The second row shows the segmentation results predicted by our model without any post processing. The holes are marked with red boxes and the noises are marked with yellow boxes. The third row shows the final segmentation masks generated by the post processing method.

using mobile devices like smartphones and tablets. An example of a depth-wise separable convolution layer is shown in Fig. 3c, compared to a traditional convolutional layer shown in Fig. 3b.

The model has an encoder-decoder architecture as shown in Fig. 2. The encoder is built by repeatedly applying the depth-separable convolution block (marked with diagonal lines in Fig. 2). Each block, illustrated in Fig. 3a, consists of six layers: a 3×3 depth-wise convolutional layer followed by batch normalization and ReLU activation²³, and a 1×1 point-wise convolution layer followed again by batch normalization and ReLU. To be more specific, ReLU6²⁴ was used as the activation function. In the decoder, shown in Fig. 2, the encoded features are captured in multiscale with a spatial pyramid pooling block, and then concatenated with higher-level features generated from a pooling layer and a bilinear up-sampling layer. After the concatenation, we apply a few 3×3 convolutions to refine the features followed by another simple bilinear up-sampling by a factor of 4 to generate the final output. A batch normalization layer is inserted into each bottleneck block and a dropout layer is inserted right before the output layer. In MobileNetV2, a width multiplier α is introduced to deal with various dimensions of input images. we let $\alpha = 1$ thus the input image size is set to 224 pixels \times 224 pixels in our model.

Transfer learning. To make the training more efficient, we used transfer learning for our deep learning model. Instead of randomly initializing the weights in our model, the MobileNetV2 model, pre-trained on the Pascal VOC segmentation dataset²⁵ was loaded before training. Transfer learning with the pre-trained model is beneficial to the training process in the sense that the weights converge faster and better.

Post-processing. The raw segmentation masks predicted by our model are grayscale images with pixel intensities that range from 0 to 255. In the post processing step, binary segmentation masks are first generated from thresholding with a fixed threshold of 127, which is half the max intensity. The binary masks are further processed by hole filling and removal of small regions to generate the final segmentation masks as shown in Fig. 4. We notice that abnormal tissue like fibrinous tissue within chronic wounds could be identified as non-wound and cause holes in the segmented wound regions. Such holes are detected by finding small connected components in the segmentation results and filled to improve the true positive rate using connected component labelling (CCL)²⁶. The small false-positive noises are removed in the same way. The images in the dataset are cropped from the raw image for each wound. So, we simply remove noises in the segmentation results by removing the connected component small enough based on adaptive thresholds. To be more specific, a connected region is removed when the number of pixels within the region is less than a threshold, which is adaptively calculated based on the total number of pixels segmented as wound pixels in the image.

Model	VGG16 (%)	SegNet (%)	U-Net (%)	Mask-RCNN (%)	MobileNetV2 (%)	MobileNetV2 + CCL (%)
Precision	83.91	83.66	89.04	94.30	90.86	91.01
Recall	78.35	86.49	91.29	86.40	89.76	89.97
Dice	81.03	85.05	90.15	90.20	90.30	90.47

Table 1. The precision, recall, and dice score evaluated using various models on our dataset. Bold values indicate the best performance among the present models

Results

We describe the evaluation metrics and compare the segmentation performance of our method with several popular and state-of-the-art methods. Our deep learning model is trained with data augmentation and preprocessing. Extensive experiments were conducted to investigate the effectiveness of our network. FCN-VGG-16 is a popular network architecture for wound image segmentation^{17,27}. Thus, we trained this network on our dataset as the baseline model. For fairness of comparison, we used the same training strategies and data augmentation strategies throughout the experiments.

Evaluation metrics. To evaluate the segmentation performance, Precision, Recall, and the Dice coefficient are adopted as the evaluation metrics²⁸.

Precision. Precision shows the accuracy of segmentation. More specifically, Precision measures the percentage of correctly segmented pixels in the segmentation and is computed by:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}}$$

Recall. Recall also shows the accuracy of segmentation. More specifically, it measures the percentage of correctly segmented pixels in the ground truth and is computed by:

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

Dice coefficient (Dice). Dice shows the similarity between the segmentation and the ground truth. Dice is also called F1 score as a measurement balancing Precision and Recall. More specifically, Dice is computed by the harmonic mean of Precision and Recall:

$$\text{Dice} = \frac{2 \times \text{True positives}}{2 \times \text{True positives} + \text{False negatives} + \text{False positives}}$$

Experiment setup. The deep learning model in the presented work was implemented in python with Keras²⁹ and Tensorflow³⁰ backend. To speed up the training, the models were trained on a 64-bit Ubuntu PC with an 8-core 3.4 GHz CPU and a single NVIDIA RTX 2080Ti GPU. For updating the parameters in the network, we employed the Adam optimization algorithm³¹, which has been popularized in the field of stochastic optimization due to its fast convergence compared to other optimization functions. Binary cross entropy was used as the loss function and we also monitored Precision, Recall, and the Dice score as the evaluation matrices. The initial learning rate was set to 0.0001 and each minibatch contained only 2 images for balancing the training accuracy and efficiency. The convolutional kernels of our network were initialized with HE initialization³² to speed up the training process and the training time of a single epoch took about 77 s. We used early stopping to terminate the training so that the best result was saved when there was no improvement for more than 100 epochs in terms of Dice score. Eventually, our deep learning model was trained for around 1000 epochs before overfitting.

To evaluate the performance of the proposed method, we compared the segmentation results achieved by our methods with those by FCN-VGG-16^{17,27}, SegNet¹⁶, and Mask-RCNN^{33,34}. We also added 2D U-Net³⁵ to the comparison due to its outstanding segmentation performance on biomedical images with a relatively small training dataset. The segmentation results predicted by our model are demonstrated in Fig. 4 along with the illustration of our post processing method. Quantitative results evaluated with the different networks are presented in Table 1 where bold numbers indicate the best results among all the models.

Comparing our method to the others. In the performance measures, the Recall of our model was evaluated to be the second highest among all models, at 89.97%. This was 1.32% behind the highest Recall, 91.29%, which was achieved by U-Net. Our model also achieved the second highest Precision of 91.01%. Overall, the results show that our model achieves the highest accuracy with a mean Dice score of 90.47%. The VGG16 was shown to have the worst performance among all the other CNN architectures. Mask-RCNN achieved the highest Precision of 94.30%, which indicates that the segmentation predicted by Mask-RCNN contains the highest per-

Model	VGG16 (%)	SegNet (%)	U-Net (%)	Mask-RCNN (%)	MobileNetV2 (%)	MobileNetV2 + CCL (%)
Precision	77.84	72.03	86.84	98.40	93.69	93.84
Recall	80.69	73.87	81.33	88.60	94.06	94.27
Dice	79.24	72.94	84.01	93.20	93.88	94.05

Table 2. The precision, recall, and dice score evaluated using various models on the Medetec dataset. Bold values indicate the best performance among the present models

Model name	FCN-VGG16	SegNet	U-Net	Mask-RCNN	MobileNetV2
Number of parameters	134,264,641	902,561	4,834,839	63,621,918	2,141,505

Table 3. Comparison of total numbers of trainable parameters.

centage of true positive pixels. However, the Recall is only evaluated to 86.40%, meaning that more false negative pixels are undetected compared to U-Net and MobileNetV2. Our accuracy was slightly higher than U-Net and Mask-RCNN, and significantly higher than SegNet and VGG16.

Comparison within the Medetec Dataset. Apart from our dataset, we also conducted experiments on the Medetec Wound Dataset³⁶ and compared the segmentation performance of these methods. The results are shown in Table 2. We annotated the dataset in the same way that our dataset was annotated and trained the networks with the same experimental setup. The highest Dice score is evaluated to 94.05% using MobileNetV2 + CCL. The performance evaluation agrees with the conclusion drawn from our dataset where our method outperforms the others regardless of which chronic wound segmentation dataset is used, thereby demonstrating that our model is robust and unbiased.

Discussion

Comparing our method to VGG16, the Dice score is boosted from 81.03 to 90.47% tested on our dataset. Based on the appearance of chronic wounds, we know that wound segmentation is complicated by various shapes, colors, and the presence of different types of tissue. The patient images captured in clinic settings also suffer from various lighting conditions and perspectives. In MobileNetV2, the deeper architecture has more convolutional layers than VGG16, which makes MobileNetV2 more capable to understand and solve these variables. MobileNetV2 utilizes residual blocks with skip connections instead of the sequential convolution layers in VGG16 to build a deeper network. These skip connections bridging the beginning and the end of a convolutional block allows the network to access earlier activations that weren't modified in the convolutional block and enhance the capacity of the network.

Another comparison between U-Net and SegNet indicates that the former model is significantly better in terms of mean Dice score. Similar to the previous comparison, U-Net also introduces skip connections between convolutional layers to replace the pooling indices operation in the architecture of SegNet. These skip connections concatenate the output of the transposed convolution layers with the feature maps from the encoder at the same level. Thus, the expansion section which consists of a large number of feature channels allows the network to propagate localization combined with contextual information from the contraction section to higher resolution layers. Intuitively, in the expansion section or “decoder” of the U-Net architecture, the segmentation results are reconstructed with the structural features that are learned in the contraction section or the “decoder”. This allows the U-Net to make predictions at more precise locations. These comparisons have illustrated the effectiveness of skip connections for improving the accuracy of wound segmentation.

Besides the performance, our method is also efficient and lightweight. As shown in Table 3, the total number of trainable parameters in the adopted MobileNetV2 was only a fraction of the numbers in U-Net, VGG16, and Mask-RCNN. Thus, the network took less time during training and could be applied to mobile devices with less memory and limited computational power. Alternatively, higher-resolution input images could be fed into MobileNetV2 with less memory size and computational power comparing to the other models.

Conclusions

We attempted to solve the automated segmentation problem of chronic foot ulcers in a dataset we built on our own using deep learning. We conducted comprehensive experiments and analyses on SegNet, VGG16, U-Net, Mask-RCNN, and our model based on MobileNetV2 and CCL to evaluate the performance of chronic wound segmentation. In the comparison of various neural networks, our method has demonstrated its effectiveness and mobility in the field of image segmentation due to its fully convolutional architecture consisting of depth-wise separable convolutional layers. We demonstrated the robustness of our model by testing it on the foot ulcer images in the publicly available Medetec Wound Dataset where our model still achieves the highest Dice score. In the future, we plan to improve our work by a novel multi-stream neural network architecture that extracts the shape features separately from the pixel-wise convolution in our deep learning model. A sketch of this idea is demonstrated in Fig. 5. With the advance of hardware and mobile computing, larger deep learning models

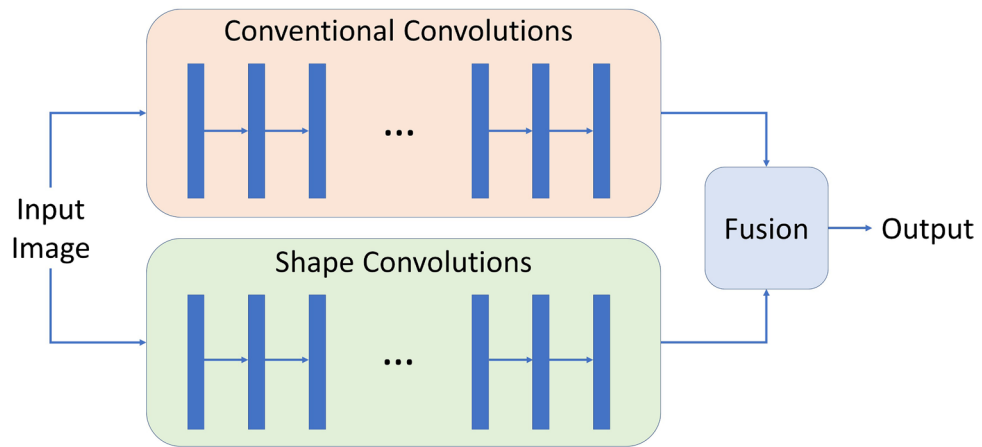


Figure 5. An illustration of the idea of multi-stream model that recognize pixels intensities and shape information separately. To predict segmentation masks, the output tensors from both streams can be fused using carefully designed concatenation layers or multi-scale pooling layers.

will be runnable on mobile devices. Another future research is testing deeper neural networks on our dataset. Also, we will include more data in the dataset to improve the robustness and prediction accuracy of our method.

Data availability

The dataset generated and analysed during the current study is available in the repository, *The Foot Ulcer Dataset* (https://github.com/uwm-bigdata/wound-segmentation/tree/master/data/wound_dataset).

Received: 7 May 2020; Accepted: 30 November 2020

Published online: 14 December 2020

References

1. Frykberg, R. G. Challenges in the treatment of chronic wounds. *Adv. Wound Care* **4**, 560–582 (2015).
2. Sen, C. K. Human wounds and its burden: an updated compendium of estimates. *Adv. Wound Care* **8**, 39–48 (2019).
3. Branski, L. K., Gauglitz, G. G., Herndon, D. N. & Jeschke, M. G. A review of gene and stem cell therapy in cutaneous wound healing. *Burns* **35**, 171–180 (2009).
4. Song, B. & Sacan, A. Automated wound identification system based on image segmentation and artificial neural networks. In *2012 IEEE International Conference on Bioinformatics and Biomedicine*, 1–4 (2012).
5. Fauzi, M. F. A. *et al.* Computerized segmentation and measurement of chronic wound images. *Comput. Biol. Med.* **60**, 74–85 (2015).
6. Hettiarachchi, N. D. J., Mahindaratne, R. B. H., Mendis, G. D. C., Nanayakkara, H. T. & Nanayakkara, N. D. Mobile based wound measurement. In *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*, 298–301 (2013).
7. Hani, A. F. M., Arshad, L., Malik, A. S., Jamil, A. & Bin, F. Y. B. Haemoglobin distribution in ulcers for healing assessment. In *2012 4th International Conference on Intelligent and Advanced Systems (ICIAS2012)*, 362–367 (2012).
8. Wantanajittikul, K., Auephanwiriyakul, S., Theera-Umpon, N. & Koanantakool, T. Automatic segmentation and degree identification in burn colour images. In *The 4th 2011 Biomedical Engineering International Conference*, 169–173 (2012).
9. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 1097–1105 (2012).
10. Russakovsky, O. *et al.* Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
11. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017).
12. Garcia-Garcia, A., S. Orts-Escobedo, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. arXiv preprint <http://arXiv.org/1704.06857> (2017).
13. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
14. Wang, C., Guo, Y., Chen, W. & Yu, Z. Fully automatic intervertebral disc segmentation using multimodal 3d u-net. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, 730–739 (2019).
15. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440 (2015).
16. Wang, C. *et al.* A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2415–2418 (2015).
17. Goyal, M., Yap, M. H., Reeves, N. D., Rajbhandari, S. & Spragg, J. Fully convolutional networks for diabetic foot ulcer segmentation. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 618–623 (2017).
18. Liu, X. *et al.* A framework of wound segmentation based on deep convolutional networks. In *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 1–7 (2017).
19. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. Mobilenetv2: inverted residuals and linear bottlenecks. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 4510–4520 (2018).
20. Redmon, J. & Farhadi, A. Yolov3: an incremental improvement. Preprint at <http://arXiv.org/1804.02767> (2018).
21. Tzatalin. *LabelImg*. *Git Code* (2015). <https://github.com/tzatalin/labelimg>.
22. Chollet, F. Xception: deep learning with depthwise separable convolutions. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1251–1258 (2017).
23. Nair, V. & Hinton, G. E. Rectified linear units improve restricted boltzmann machines. In *Proc. 27th International Conference on Machine Learning (ICML-10)*, 807–814 (2010).
24. Krizhevsky, A., and G. Hinton. Convolutional deep belief networks on cifar-10. *Unpublished manuscript* **40**, 1–9 (2010).

25. Everingham, M. *et al.* The pascal visual object classes challenge: a retrospective. *Int. J. Comput. Vis.* **111**(1), 98–136 (2015).
26. Pearce, D. J. *An Improved Algorithm for Finding the Strongly Connected Components of a Directed Graph* (Victoria University, Wellington, 2005).
27. Li, F., Wang, C., Liu, X., Peng, Y. & Jin, S. A composite model of wound segmentation based on traditional methods and deep neural networks. *Comput. Intell. Neurosci.* **2018**, 1–12 (2018).
28. Zou, K. H. *et al.* Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Acad. Radiol.* **11**, 178–189 (2004).
29. Chollet, F. *et al.* Keras, chollet2015keras. <https://keras.io>.
30. Girija, S. S. Tensorflow: large-scale machine learning on heterogeneous distributed systems. (2016).
31. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at <http://arXiv.org/1412.6980> (2014).
32. He, K., Zhang, X., Ren, S. & Sun, J.. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In *Proc. IEEE International Conference on Computer Vision*, 1026–1034 (2015).
33. He, K. *et al.* Mask r-cnn. In *Proc. IEEE International Conference on Computer Vision* (2017).
34. Abdulla, W. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. *Git code* (2017). https://github.com/matterport/Mask_RCNN.
35. Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241 (2015).
36. Thomas, S. Stock pictures of wounds. *Medetec Wound Database* (2020). <http://www.medetec.co.uk/files/medetec-image-databases.html>.

Author contributions

C.W. wrote the main manuscript text. D.M.A. implemented the image localization module in the pre-processing program and wrote the corresponding paragraph. V.W. and B.R. helped collecting and preparing the images in the dataset. C.W. and M.K.D. implemented the main program. J.N. provided and consented the images to be used in this research. Z.Y. and S.G. led and guided the research. All authors reviewed the manuscript.

Funding

This work is partially supported by the Discovery and Innovation Grant (DIG) award and the Catalyst Grant Program at The University of Wisconsin-Milwaukee.

Competing interests

The authors declare no competing interests. Additional information

Correspondence and requests for materials should be addressed to S.G. or Z.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020