Detect-and-segment: A deep learning approach to automate wound image segmentation☆

Gaetano Scebba [a],*, Jia Zhang [a], Sabrina Catanzaro [b], Carina Mihai [c], Oliver Distler [c], Martin Berli [d], Walter Karlen [a]

[a] *Mobile Health Systems Lab, Institute of Robotics and Intelligent Systems, Department of Health Sciences and Technology, ETH Zurich, Switzerland*
[b] *Unit for Clinical and Applied Research, Balgrist University Hospital, Zurich, Switzerland*
[c] *Department of Rheumatology, University Hospital Zurich, University of Zurich, Switzerland*
[d] *Division of Technical Orthopedics, Department of Orthopedic Surgery, Balgrist University Hospital, Zurich, Switzerland*

ARTICLE INFO

ABSTRACT

Chronic wounds significantly impact quality of life. They can rapidly deteriorate and require close monitoring of healing progress. Image-based wound analysis is a way of objectively assessing the wound status by quantifying important features that are related to healing. However, high heterogeneity of the wound types and imaging conditions challenge the robust segmentation of wound images. We present Detect-and-Segment (DS), a deep learning approach to produce wound segmentation maps with high generalization capabilities. In our approach, dedicated deep neural networks detected the wound position, isolated the wound from the perturbing background, and computed a wound segmentation map. We tested this approach on a diabetic foot ulcers data set and compared it to a segmentation method based on the full image. To evaluate its generalizability on out-of-distribution data, we measured the performance of the DS approach on 4 additional independent data sets, with larger variety of wound types from different body locations. The Matthews' correlation coefficient (MCC) improved from 0.29 (full image) to 0.85 (DS) on the diabetic foot ulcer data set. When the DS was tested on the independent data sets, the mean MCC increased from 0.17 to 0.85 . Furthermore, the DS enabled the training of segmentation models with up to 90% less training data without impacting the segmentation performance. The proposed DS approach is a step towards automating wound analysis and reducing efforts to manage chronic wounds.

1. Introduction

Chronic wounds cause a great reduction of quality of life in patients [1]. If not treated, they can lead to severe complications such as limb amputations and death [2,3]. In fact, periodic examination and thorough treatment are essential to prevent deterioration [4]. Wound management requires to regularly assess, document, and treat the wound by medical professionals [5]. Trained nurses annotate the relevant information in assessment reports that are then used to track the healing progression and plan the most appropriate treatment strategy. The assessment is typically performed by visual inspection of wound features, such as area and depth measurements, and the annotation of growth tissue [6]. Despite the importance of objective and accurate wound documentation, the assessment reports are often inconsistent and sparse [7].

Wound imaging through cameras and smartphones has been adopted to support wound documentation with a more objective tool [8–10]. Wound images are taken at each examination and medical professionals obtain insights on the wound healing progression by visually comparing with images taken at past appointments. However, the use of wound images does not entirely eliminate the subjectivity of wound documentation and might even introduce an additional source of variability to the whole documentation procedure, due to the non-quantitative examination of past images.

Image analysis techniques could potentially be used to obtain all the necessary information for the wound documentation in a more objective manner. In this process, a challenging task is the semantic segmentation. In this task, each pixel of an image is classified to either belonging to the foreground or background class. In the case of wound
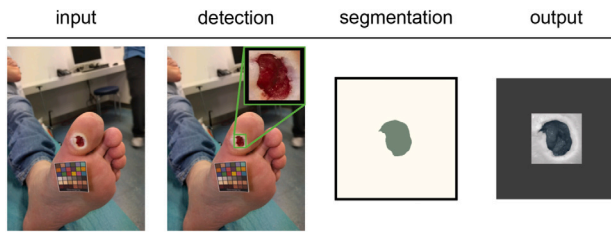
Fig. 1. Proposed approach for robust segmentation of wound images. Dedicated deep neural networks detect the wound position, isolate the wound bed, and produce the wound segmentation map.

Fig. 2. Detect-and-segment (DS) approach (gray box). The raw wound image (a) is processed with a wound detector neural network (b) to localize the wound within the image. The pre-processing module (c) performs cropping, zero-padding and resizing of the image to exclude uninformative background pixels. The segmentation model (d) produces the final segmentation mask (e).

images, the accurate distinction between wound bed and background information is important because the wound bed area is considered as one of the most relevant biomarkers for the prediction of wound healing [11] and excluding uninformative pixels can improve further analysis.

However, generalizing segmentation performance across wound images is challenging. Heterogeneity of wound types, colors, shapes, body position, background composition, image capturing conditions, and capturing devices are all factors that lead to highly diverse wound images [12]. Clinical wound photography established a set of guidelines to standardize the working procedures and improve objectivity of the image capturing [13]. While these steps would provide wound nurses with additional workload, they do not consistently solve the generalization difficulties.

The goal of this work was to simplify the wound imaging process with digital cameras by developing an image segmentation approach that is robust to the heterogeneity of wound images. We focused on the development of a deep learning approach to accomplish wound segmentation that was capable of high generalization towards out-of-distribution wound images. Our approach enabled a wound image segmentation without user interaction and consequently alleviated human efforts for image standardization during wound images collection (Fig. 1).

1.1. Deep learning in object detection and segmentation for medicine

Deep learning has been widely applied for biomedical image processing in pathology, radiology or dermatology. Numerous approaches proposed deep neural networks (DNN) to localize objects of interest such as lesions [9,14] or anatomical structures [15,16]. Similarly, DNNs have also been adopted to segment single cells from electron microscopy sections [17,18] or from histopathological images [19], brain and cardiac anatomical substructures from magnetic resonance imaging (MRI) [20] or computer tomography (CT) scans [21], tissue lesions from breast MRI [22], lungs and liver CT scans [23,24], or skin lesions such as melanoma [25] and wounds [8] from smartphone images. However, the segmentation of medical images is often challenged by the highly variable background of many image domains. In some cases, pre-processing the images by removing uninformative background pixels, either manually or automatically, is a necessary step [16,26]. In this work, we build on these advances to develop a novel approach to robust wound image segmentation.

1.2. Segmenting wound images

State of the art methods for wound image segmentation are based on several machine learning approaches, ranging from the use of classification from manually engineered image features, to the use of end-to-end DNNs. Initially, researchers have proposed to use ad-hoc image features, which can represent the complexity and variability of wound shapes and colors, in conjunction with supervised and unsupervised classifiers. These works have used color and texture features with
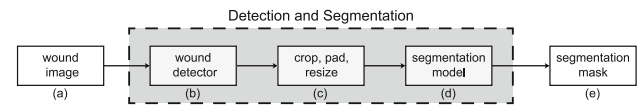
region growing algorithm and optimal thresholding [27], with support vector machines [28,29], with neural networks [30,31], and with k-means clustering algorithm [32,33]. However, coping with the high variability of wounds is challenging and features that are designed and evaluated for a limited set of images might not be able to generalize and be effective to segment images with unseen type of wounds or different lighting conditions.

To replace the need for manually designing features for wound image segmentation, research interest has also investigated the use of deep learning. Convolutional neural networks (CNN) have been used to perform image segmentation because they can autonomously learn to extract the most meaningful set of features for the segmentation task [34]. Wang et al. proposed to use the ConvNet [35], which is based on an encoder–decoder CNN architecture, and tested it on a data set containing diabetic foot ulcers that were manually pre-processed with a modified version of the GrabCut algorithm [36] to remove background. Similarly, other CNN architectures have been used, such as the U-Net [17,37] or the fully convolutional network (FCN) [38,39], all in combination with pre-processing steps that remove uninformative background either by the manual interaction of the user [37] or by manual features engineering to detect background pixels [15,39]. Standardizing the background in advance, before taking the picture, has also been a technical choice proposed in several works to keep the background consistent over different wound images. These approaches used capturing box to capture wound feet pictures [8] or paper barriers to cover the background [40,41].

In contrast to previous work that focused on the segmentation of diabetic foot images and developed semi-automatic methods which are dependent on user interaction, we present an automatic deep learning approach for the robust image segmentation of wounds. In this manuscript, we validate this approach and demonstrate robustness using 4 independent data sets with images from diverse wound types and different body locations. Our deep learning approach aims to address the challenge of high heterogeneity of wound images and enables the reduction of pixel-level training data with high preservation of segmentation performance.

2. Detect-and-segment approach

We have developed the detect-and-segment (DS), a deep learning approach for robust segmentation of wound images collected through either smartphones or digital cameras. Our approach consisted of three main steps (Fig. 2). First, the wound detector localized the wound bed region. Second, we processed the image using the localized wound bed to automatically remove uninformative background pixels. Third, a binary classification map was obtained by segmenting the resulting image into wound bed and background pixels.

2.1. Wound bed detection

To localize the position of the wound bed region within the image, we trained an object detector, with an architecture specifically designed for object detection. The model output was a set of bounding boxes that correspond to all the detected wounds in the image. Each box consisted of a set of 4 coordinates, which represented the top left and

the bottom right corner of a box enclosing the wound bed. Additionally, the object detector did output a confidence score that represented the probability for the detected object to be a wound. To select the best object detector for the DS approach, we conducted a systematic model selection. We optimized the object detection RetinaNet model [42] with three different feature extractors that were pre-trained on the Microsoft COCO data set [43]: resnet50 [44], Efficient Det [45] with backbone D0 and D1, and the MobileNet [46]. The training loss was set to the sum of the focal loss and the weighted L1 loss used for box regression [42]. We used the Adam optimizer [47] with 100 epochs training and an exponential learning rate decay strategy. Data of the diabetic foot ulcers grand challenge 2020 were used for the wound detector model selection, with a split of 90% of all wound images into training and 10% into held-out test sets [9,48]. We additionally split the training set using a 5-fold cross validation. The performance of each wound detection model was evaluated in terms of mean average precision at an intersection over union of 50%. We selected the MobileNet architecture as the wound detector of the DS approach, as it showed highest wound detection accuracy. Information regarding training data, data set split, hyperparameters optimization, and performance metrics is provided as supplementary material (Section 1, supplementary material).

2.2. Detection post-processing

In the post-processing stage, we selected only those bounding boxes with a confidence score greater than 0.3, which was obtained empirically. We noticed that the bounding box produced by the wound detector had the tendency to tightly enclose the wound, resulting in unbalanced pixel count between the wound and the background class. Therefore, the area of the selected boxes was then increased by 50% to include part of the surrounding skin and reduce class imbalance between wound and background classes. Thus, we cropped the wound image based on the coordinates of the post-processed box to exclude uninformative background pixels. To ensure standardized input size to the wound segmentation models, the obtained cropped image was zero padded and resized to 512×512 pixels, using the nearest neighbor interpolation algorithm provided by the Python Image Library [49].

2.3. Wound bed segmentation

We considered the segmentation of the wound image as a binary segmentation task. Our goal was to produce a segmentation mask in which each value represented whether the correspondent pixel on the wound image belonged to the wound or background class. We evaluated performance of the DS approach by training four different segmentation architectures: U-Net [17], ConvNet [35], DeepLabV3 [50] with a ResNet-101 [44] backbone, and a fully convolutional neural network (FCN) [41] with a VGG16 [51] backbone. To account for the class imbalance between wound bed and background pixels of the wound images, we used the pixel-wise weighted binary cross entropy loss, where the weighting term was computed as the ratio between the total number of wound bed and background pixels of each training set fold. We used the Adam optimizer [47] for up to 100 epochs with an early stopping of 12 epochs on the validation set loss and an exponential learning rate decay strategy. To additionally combat overfitting risk, we augmented the training images by applying vertical and horizontal flip, brightness variation with maximum delta of 0.1, and scale and crop with a scale factor ranging from 0.1 to 2.

3. Experiments

We evaluated the performance of the DS approach using multiple independent data sets, including images from different body locations and with different wound types. To evaluate how the DS approach improves generalization, we extended the testing stage to out-of-distribution images from data sets not used for training. In addition, we conducted a series of experiments to test whether (1) the automatic detection subcomponent of the DS approach improves the wound segmentation performance compared to manual detection and absence of detection, (2) the combination of automatic detection and segmentation enhances the generalization capabilities of the segmentation models when tested on out-of-distribution wound images, and (3) the DS approach reduces the amount of segmentation masks needed for training stage.

3.1. Data sets

In this work, we relied on wound images obtained from six independent data sets. The SwissWOU data set was built by the authors for this manuscript and included two subsets of images with diabetic foot and systemic sclerosis digital ulcers. The diabetic foot ulcers grand challenge 2020 (DFUC) [9,48], Medtec [52], second intention healing (SIH) [53], and foot ulcer segmentation challenge (FUSC) [34] data sets are publicly available data sets that included images with a variety of wound types. The data of the DFUC data set were used for the selection of the wound detector (Section 1, supplementary material). The data of the SwissWOU subset with diabetic foot wounds were used during training and testing stages of the segmentation models. The data of the SwissWOU subset with systemic sclerosis digital ulcers, Medtec, SIH, and FUSC data sets were only used during testing.

SwissWOU — diabetic foot ulcers (SW-DFU). We collected wound images from March 2018 to January 2020 from 76 consenting patients (23 females and 53 males, mean age: 67 ± 11 y) at the wound consultation unit of the Balgrist University Hospital, Zurich, Switzerland. The data set included a total of 1096 high-resolution wound images from 756 wounds that were acquired with smartphone cameras. The images were collected at around 50 cm camera-wound distance, with the smartphone oriented parallel to the wound plane. Wound nurses and research assistants were responsible for images collection. At each wound consultation appointment, we collected one image in ambient light and one image using an external LED ring flashlight attached to the smartphone.

SwissWOU — systemic sclerosis digital ulcers (SW-SSD). We collected images from May 2019 to May 2020 from 15 consenting patients that conducted a regular, self-guided wound assessment at home as part of a study that investigated the wound image quality [10]. The patients were provided with a smartphone and a holder specifically designed to standardize the camera-wound distance. Wound images were automatically uploaded to a remote database using RedCap [54]. Due to the small sizes of the digital compared to diabetic foot ulcers, in the analysis of this work we included only those images with a wound coverage greater than 0.1%, which is defined as the ratio between the wound and image area. As a result, we included 7 patients (6 females and 1 male, mean age: 47 ± 11 y) with a total of 63 digital ulcers images.

Image sizes of the SwissWOU data sets ranged from 1080×1920 to 3000×4000 pixels. The studies conducted to collect the images of the SW-DFU and SW-SSD data sets followed the ethical principles for research involving humans according the Declaration of Helsinki and were approved by the institutional ethics committee (ETH Zurich EK 2017-N-07 and 2019-N-22). The smartphone app used in both studies featured a camera module that disabled all the image processing techniques that might have altered image colors other than auto-exposure. The app was installed on a Nexus 5x (LG Electronics, Seoul, South Korea) and a Xiaomi Mi A1 (Xiaomi Corporation, Beijing, China).

Medtec. Medtec is a publicly available wound image database [52] that includes a large variety of wound types. We used a subset of 53 randomly selected wound images containing wounds caused by pressure, surgery, diabetes, and meningitis, which were digitalized from 35 mm analog films. No further information regarding data collection was available.

Table 1
Data distribution of training folds and test set from the SwissWOU data set containing diabetic foot ulcers (SW-DFU). Splits are divided into the experiment without (*None*) and with manual (*Manual*) detection step, as well as training (train) and validation (val).

Fold (#)	Patients train set (#)	Images/patient train set (#)	Mean coverage train set (%)	Patients val set (#)	Image/patient val set (#)	Mean coverage val set (%)
			No detection step - *None*			
1	54	8	2.28	13	10	1.47
2	53	9	2.09	14	6	2.56
3	54	8	2.23	13	9	1.66
4	53	8	2.20	14	10	1.84
5	54	9	1.89	13	7	3.07
Test set	–	–	–	7	8	3.10
			Manual detection step - *Manual*			
1	54	9	13.15	13	9	10.29
2	54	9	12.92	13	8	11.23
3	53	9	12.09	14	6	14.49
4	53	8	12.45	14	9	13.17
5	54	8	12.35	13	10	13.61
Test set	–	–	–	7	8	10.77

Secondary intention healing. The SIH data set was released by Yang et al. [53] and included wound images of patients undergoing secondary intention healing at the Severance Hospital, Seoul, South Korea. We randomly selected a subset of the data set consisting of 58 wound images from 12 patients, with diverse wound types from different body positions. The images of the data set were collected at 30–50 cm distance, with different digital single-lens reflex (DSLR) cameras and included wounds from different body locations such as face, hands, back, and feet.

Foot ulcer segmentation challenge. The FUSC data set included images of diabetic foot ulcers [34]. We randomly selected a subset 60 wound images. The data were collected by nurses and clinicians of the Advancing the Zenith of Healthcare Wound and Vascular Center, Milwaukee, Wisconsin, United States. The images were captured using a Canon SX 620 HS camera (Canon Inc., Tokyo, Japan) and an iPad Pro (Apple Inc., Cupertino, California, United States).

3.1.1. Annotation

The medical annotations of the SW-DFU, SW-SSD, Medtec, SIH, and FUSC data sets consisted of segmentation masks with pixel-level labels for each image, in which 1 corresponded to the wound bed and 0 to the background. The images of the FUSC data set included pixel-level labels of the wound bed produced by wound care specialists. Image labeling for the SW-DFU, SW-SSD, Medtec, and SIH data sets was performed by a research assistant and the first author, who were extensively trained by professional wound nurses to recognize the wound bed area.

3.1.2. Data set split

The SW-DFU data set was used during training and testing stages of the segmentation models and was split considering the number of patients. The SW-DFU data set included a deidentified patient id for each wound image that allowed a subject dependent split of the data set. We split the SW-DFU data set into training set (90% of all patients), which was used for training and hyperparameter tuning, and held-out test set (10% of all patients), which was only used for final performance reporting. The training set was additionally split according to the 5-fold cross validation scheme and each split was conducted at random stratification by number of images per patient and mean wound coverage (Table 1). This way, we balanced all the folds and ensured that multiple wound images from the same patient were only included in one fold.

3.2. Hyperparameters tuning

To select the best set of hyperparameters, we followed a systematic approach with random search strategy, with hyperparameters randomly chosen from predefined ranges (Table S.2, Supplementary material). We performed 50 random search iterations to increase the probability of finding a set of hyperparameters closer to the optimum and to further evaluate the performance of the analyzed models against random sets of hyperparameters. The performance of each set of hyperparameters was evaluated against the validation set following a 5-fold cross validation. The validation performances computed against each of the 5 validation sets were averaged to obtain one performance value per optimization run. We selected the best set of hyperparameters across the 50 optimization runs by choosing the segmentation model with highest Matthews' correlation coefficient. The best set of hyperparameters was then used to train each model on the entire training set (merging training and validation set) and the final performance evaluated on the test set.

3.3. Evaluation

3.3.1. Segmentation performance

To quantify the segmentation performance of the DS approach, we compared the performance of the UNet, ConvNet, DeepLab and FCN models, which were trained and tested only on the images of the SW-DFU that differed for the wound detection step:

- **Manual:** the images of train, validation, and test set were pre-cropped by manually selecting a rectangle that included the wound.
- **None:** the images of train, validation, and test set were used with their full dimensions without any wound detection and cropping.
- **Automatic:** the images of train and validation set were manually pre-cropped and those of the test set were pre-cropped using the selected wound detection model.

3.3.2. Generalizability

To evaluate the generalization capability of the DS approach, we computed its segmentation performance against out-of-distribution wound images. We compared the performance of all the segmentation models, which were trained only on the images of the SW-DFU data set and tested on the SW-SSD, Medtec, SIH, and FUSC data sets with and without Manual and Automatic wound detection step. These data sets not only included various wound types from different body locations, but also diverse background–foreground compositions with large differences in mean skin and wound coverage (Fig. 3).
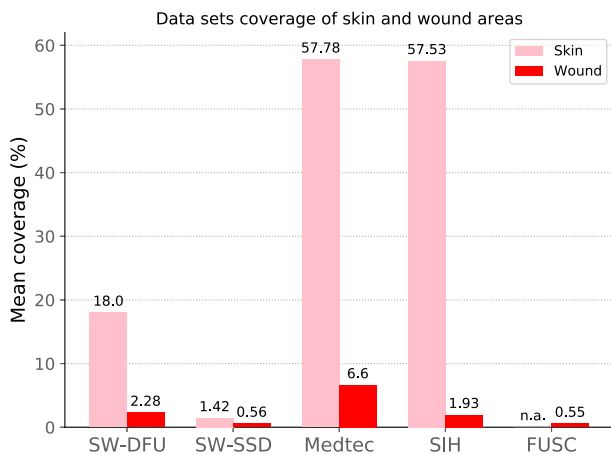
Data sets coverage of skin and wound areas

Fig. 3. Data set mean coverage. Comparison of the skin and wound mean coverage percentage of the SwissWOU with diabetic foot ulcers (SW-DFU) and systemic sclerosis digital ulcers (SW-SSD), Medtec, SIH, and FUSC data sets. Mean skin coverage of the FUSC data set is not available as skin segmentation maps are not provided by the data set authors.

3.3.3. Impact of the data set size

To determine how the DS approach preserved the segmentation performance with reduced amount of training data, we re-trained all the segmentation models on the SW-DFU data set, by selecting their optimal hyperparameters and randomly removing from 10% up to 90% of the training set images.

3.4. Performance metrics

We evaluated the segmentation performance of the DS approach in terms of Matthews' correlation coefficient (MCC), dice score and intersection over union (IoU, also known as Jaccard index). The MCC is widely used to handle data imbalance and can be seen as the discretization of the Pearson's correlation coefficient for binary variables [55] and ranges from −1 to 1. Dice score and IoU are the most commonly used metrics in medical image segmentation because they adequately reflect the agreement for object segmentation [56]. The IoU is defined as the ratio between the intersection portion of the detected and ground truth boxes divided by their union and ranges from 0 to 1.

Statistical comparisons were performed at a significance level of 95% ($p < 0.05$). We first tested the normality of the analyzed MCC and IoU distributions using the Shapiro–Wilk test and compared them with the Mann–Whitney U test. Bonferroni correction was applied in case of multiple comparisons.

4. Results

4.1. Segmentation performance

The DS approach achieved significantly higher MCC and IoU than all the models trained on full dimension images without the detection step ($p < 0.05$). Additionally, the DS approach significantly improved the segmentation performance of the U-Net and ConvNet compared to the manual detection step (Table 2).

4.2. Generalizability

The DS approach did not show significantly different performance compared to the manual detection step when tested on the images of the Medtec and SIH data sets for all the analyzed segmentation models (Table 3). The manual detection step showed significantly higher MCC and IoU than the DS approach when applied to the images of the SW-SSD subset for all the analyzed segmentation models (Table 3).

Additionally, we found that the DS approach significantly outperformed the no-detection step for the U-Net and Convnet models when tested on the images of the SW-SSD, Medtec, SIH, and FUSC data sets (Table 3).

4.3. Impact of the data set size

In terms of segmentation performance as measured by the MCC, applying the DS approach allowed all the segmentation models to perform significantly better than the no-detection across a different amount of training data (Fig. 4). In particular, when the DS approach was applied with U-Net and FCN models, they preserved the segmentation performance within the ranges of 6.4% (MCC = [0.84 – 0.78] @ 90%–10% of training set size) and 0.3% (MCC = [0.770–0.768] @ 90%–10% of training set size) across a reduction of 25 to 90% the training set size.

5. Discussion

We presented and evaluated a novel deep learning approach for automatic segmentation of chronic wounds. Our approach featured two dedicated deep learning architectures that detect and segment the wound region without user interaction, in images collected from smartphones and DSLR cameras. To the best of our knowledge, this is the first study to extensively analyze the effects of the combination of the two processing steps, which have been only considered separately in past wound imaging literature. We evaluated the segmentation performance of the DS approach not only on the test set that was isolated from the same distribution of the training set, but also on wound images from independent data sets with different types of wounds. Our evaluation demonstrated that the DS approach enables the segmentation of wound images to generalize well to images from out-of-distribution data, preserve performance with reduced training set size, and to be applied fully automatically.

5.1. Segmentation performance

The DS approach combined the tasks of detection and segmentation. The key element of the proposed approach was the automatic exclusion of uninformative background pixels given by the wound detection model. Pre-cropping the wound image was important for the training of segmentation models to focus only on learning to distinguish between the wound region and the surrounding skin.

We demonstrated that the DS approach improved the performance of all the analyzed segmentation models compared to the use of the full size images without pre-cropping. Furthermore, our results indicated that the automatic wound detection model of the DS approach enabled the segmentation models to perform similarly or even better than the use of manual wound detection. A possible explanation for the superior segmentation performance of the DS approach on the images of the SW-DFU data set could be the smaller area of the automatically detected bounding boxes compared to those obtained by manual cropping (Table 2). With shrunk bounding box areas, the receptive fields of the segmentation convolutional layers processed a reduced number of background pixels at the border region between the wound bed and the skin, with a potential improvement of segmentation performance. This confirmed the importance of the wound detection task previously investigated [9,35] and expanded the work of Goyal et al. [41], who successfully apply convolutional deep neural networks to wound images with normalized background. Together, these findings will enable new applications where standardizing the background is either not trivial (e.g. in a telemedicine application) or time consuming (e.g. in a clinical environment) and lead to more effective and accurate algorithms.

Table 2

Segmentation performance. Comparison of UNet, ConvNet, DeepLab and FCN models trained on wound images pre-processed with automatic (DS approach), manual and no wound detector in terms of Matthews' correlation coefficient (MCC) and intersection over union (IoU) for segmenting wound images from the SwissWOU subsets with diabetic feet (SW-DFU). In parentheses are the standard deviations.

	Detection	UNet		ConvNet		DeepLab		FCN	
		MCC	IoU	MCC	IoU	MCC	IoU	MCC	IoU
SW-DFU	Manual	*0.80 (0.04)	*0.67 (0.06)	*0.77 (0.05)	*0.63 (0.07)	0.76 (0.06)	0.61 (0.08)	*0.81 (0.03)	*0.68 (0.04)
	None	*0.29 (0.12)	*0.10 (0.07)	*0.35 (0.13)	*0.15 (0.10)	*0.60 (0.10)	*0.43 (0.08)	*0.59 (0.11)	*0.42 (0.08)
	Automatic	0.85 (0.04)	0.75 (0.07)	0.82 (0.05)	0.70 (0.08)	0.78 (0.05)	0.63 (0.07)	0.77 (0.07)	0.63 (0.09)

*Significant at $p < 0.05/3$ to performance of models trained on images pre-processed with the Automatic wound detector (DS approach).

Table 3

Quantitative generalization performance. Comparison of UNet, ConvNet, DeepLab and FCN models trained on wound images pre-processed with automatic (DS approach) and manual wound detector in terms of Matthews' correlation coefficient (MCC) and intersection over union (IoU) for segmenting wound images from the SwissWOU subsets with systemic sclerosis digital ulcers images (SW-SSD), Medtec, secondary intention healing (SIH), and foot ulcer segmentation challenge (FUSC) data sets. In parentheses are the standard deviations.

	Detection	UNet		ConvNet		DeepLab		FCN	
		MCC	IoU	MCC	IoU	MCC	IoU	MCC	IoU
SW-SSD	Manual	*0.83 (0.02)	*0.71 (0.03)	*0.80 (0.03)	*0.68 (0.05)	*0.78 (0.08)	*0.65 (0.11)	*0.80 (0.09)	*0.68 (0.10)
	None	*0.01 (0.01)	*0.00 (0.00)	*0.05 (0.08)	*0.01 (0.02)	*0.68 (0.09)	*0.51 (0.07)	*0.59 (0.05)	*0.41 (0.05)
	Automatic	0.74 (0.07)	0.60 (0.09)	0.74 (0.05)	0.59 (0.06)	0.75 (0.05)	0.59 (0.07)	0.67 (0.15)	0.53 (0.13)
Medtec	Manual	0.90 (0.01)	0.83 (0.02)	*0.89 (0.01)	*0.80 (0.02)	0.88 (0.02)	0.78 (0.04)	0.85 (0.02)	0.74 (0.03)
	None	*0.11 (0.04)	*0.02 (0.02)	*0.44 (0.07)	*0.23 (0.05)	*0.83 (0.04)	*0.73 (0.06)	*0.78 (0.02)	*0.66 (0.02)
	Automatic	0.90 (0.01)	0.83 (0.02)	0.88 (0.01)	0.79 (0.02)	0.89 (0.02)	0.80 (0.03)	0.85 (0.01)	0.74 (0.02)
SIH	Manual	0.85 (0.02)	0.78 (0.02)	*0.87 (0.02)	*0.79 (0.03)	0.87 (0.02)	0.80 (0.03)	0.83 (0.02)	0.75 (0.04)
	None	*0.25 (0.06)	*0.07 (0.02)	*0.53 (0.06)	*0.29 (0.06)	*0.89 (0.02)	*0.81 (0.04)	*0.85 (0.03)	0.75 (0.05)
	Automatic	0.87 (0.04)	0.81 (0.06)	0.86 (0.04)	0.79 (0.07)	0.86 (0.03)	0.80 (0.05)	0.83 (0.04)	0.76 (0.06)
FUSC	Manual	0.89 (0.01)	0.80 (0.03)	0.87 (0.02)	0.77 (0.03)	*0.81 (0.02)	*0.67 (0.04)	*0.80 (0.02)	*0.66 (0.03)
	None	*0.22 (0.04)	*0.05 (0.02)	*0.34 (0.04)	*0.12 (0.03)	*0.75 (0.05)	*0.61 (0.08)	*0.80 (0.02)	*0.67 (0.02)
	Automatic	0.89 (0.02)	0.79 (0.03)	0.88 (0.01)	0.78 (0.02)	0.83 (0.03)	0.69 (0.04)	0.82 (0.02)	0.69 (0.03)

*Significant at $p < 0.05$ to performance of models trained on images pre-processed with the Automatic wound detector (DS approach).
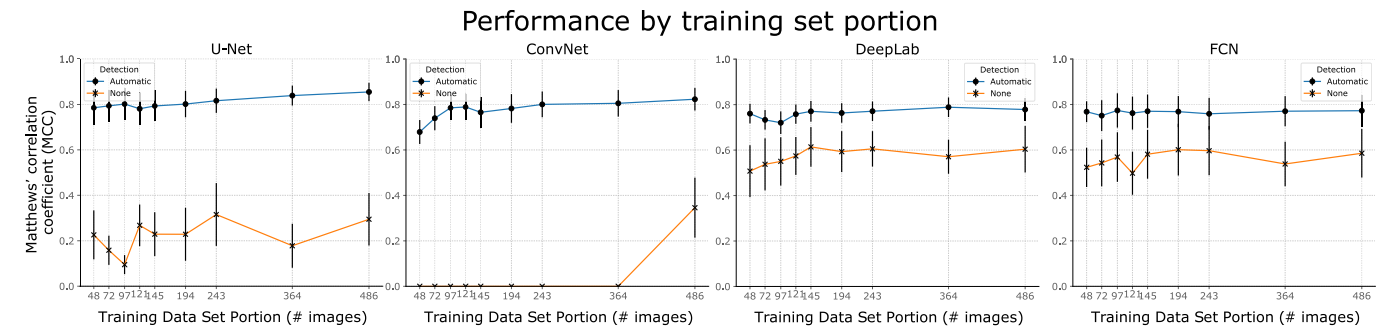

Fig. 4. Performance comparison of the DS approach (dots, blue) and no-detection (x, orange) in terms of Matthews' correlation coefficient (MCC, y-axis) when varying the amount of training set data (x-axis). The automatic detection of the DS approach enabled all the investigated segmentation models to preserve the performance when trained with reduced training set data.

5.2. Generalizability

We also evaluated the performance of the DS approach on wound images from multiple independent data sets. We found that training the segmentation models by only focusing on the wound-vs-skin problem provided all the models with improved generalization to images with different types of wounds and collected under different collection guidelines and environmental conditions. This was supported by the higher performance of all the models trained with both manual and automatic detection when tested on images with diabetic foot ulcers (SW-DFU, Table 2 and FUSC, Table 3), diabetic foot ulcers (SW-DFU, Table 2), digital ulcers (SW-SSD, Table 3), and other wound types (Medtec, SIH, Table 3). However, when the number of images from a new distribution is not sufficient to train the wound detection models, as it occurred in the case of the digital ulcers (SW-SSD, Table 3), the performance of the segmentation models may be negatively influenced. In fact, we found that the manual detection on images with systemic

sclerosis digital ulcers provided all segmentation models with higher segmentation performance than using the automatic detection (DS approach). To cope with lack of generalization due to low training data amount, more complex data augmentation techniques could be further explored to synthesize a variety of wounds (e.g. by using generative adversarial networks [57]). In addition, research efforts in wound imaging should follow the progress made in dermoscopy imaging, in which large collections of several skin lesion images are publicly available [58,59]. Wound imaging data sets should not only include images from diabetic foot wounds but also from other wound types, such as digital ulcers, post-surgical wounds, vascular ulcers, and burns.

Surprisingly, DeepLab and FCN models that were tested on the wound images of the SIH data set without detection showed higher segmentation performance than the DS approach (SIH, Table 3). The DeepLab and FCN models were built using backbone architectures that were pre-trained on larger non-medical data sets, while the U-Net and the ConvNet were trained from scratch using only wound images.
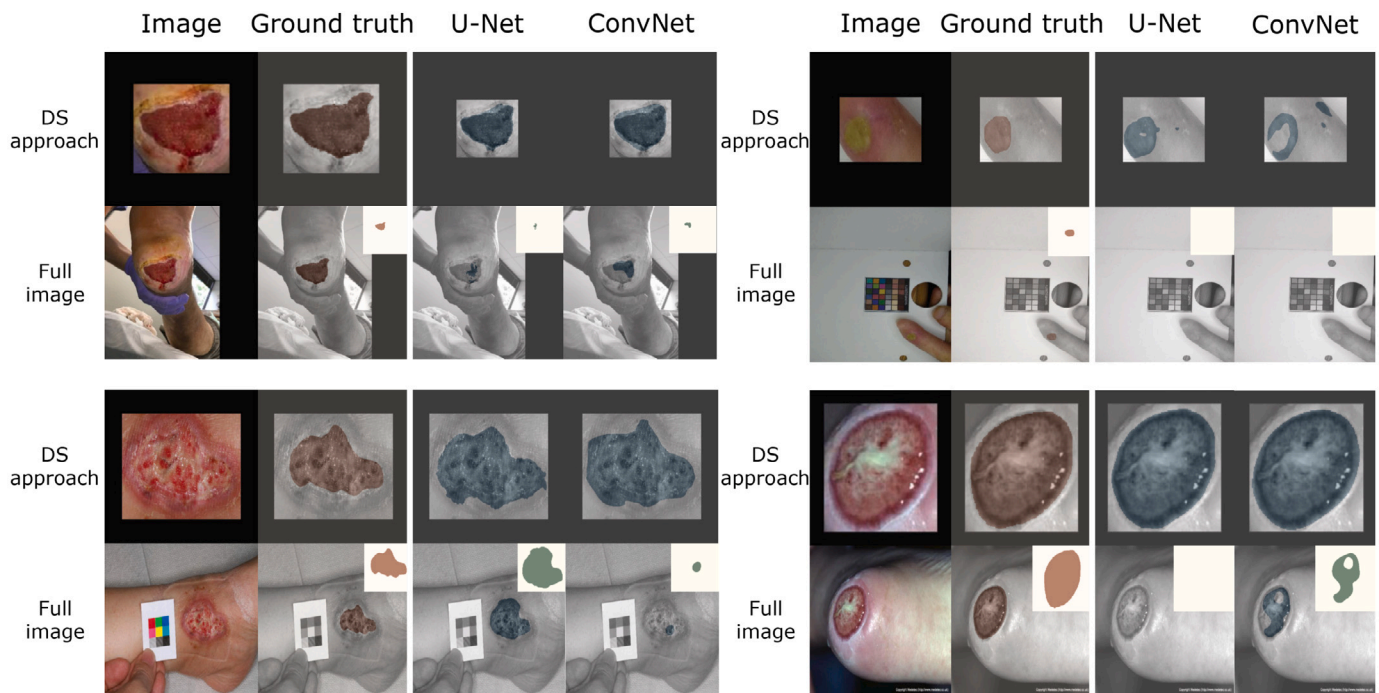
Fig. 5. Qualitative generalization performance. Comparison of segmentation masks obtained with the DS approach and with the segmentation of the full image produced by U-Net and Convnet, from (top, left) the foot ulcer segmentation challenge (FUSC), (top, right) SwissWOU — systemic sclerosis digital ulcers (SW-SSD), (bottom, left) secondary intention healing (SIH), and (bottom, right) Medtec data sets. Top right white square within each of the images corresponding to full image row depicts the ground truth (brown) and computed (green) segmentation maps. The isolation of the wound bed from the uninformative background pixels performed by the automatic detection of the DS approach enables more robust and accurate segmentation compared to the segmentation of the full image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This, in combination with the high amount of skin pixels of the SIH data set (Fig. 3) which results in a less heterogeneous background, could explain why the DeepLab and FCN models benefited from larger quantity of skin pixels around the wound. Nevertheless, considering the overall performance of the investigated segmentation models over all the analyzed data sets, applying the DS approach in combination with U-Net offered superior and more robust segmentation performance (Fig. 5). To further improve the segmentation performance of the DS, vision transformer architectures could be considered. Vision transformers combine image patch embedding and self-attention mechanisms as novel method to extract meaningful features from image data. Their innovative design outperformed traditional convolutional-based neural networks in several semantic segmentation tasks [60].

Our findings on segmentation accuracy, reduced training data, and out-of-distribution testing complement the work of Wang et al. [34], who developed a fully automatic method for wound image segmentation. In combination, these results provide the evidence that detecting and isolating the wound region from the background pixels is important for improving the accuracy of the segmentation and to generalize better to out-of-distribution wound images.

5.3. Impact of the data set size

The DS approach reduced the amount of training data for the segmentation models while preserving performance. Using less segmentation maps for training implied less time needed by trained personnel to manually select all the pixels belonging to the wound region. Despite reducing the number of segmentation maps, the labeling process for object detection was still needed. However, this process is in general faster because it only requires drawing a rectangle over the wound region. Therefore, it is a trade off between producing more labels for wound detection and less for segmentation in order to saving resources for manual data labeling. The advantage of preserving high segmentation performance with reduced training data makes the DS approach

a preferable choice compared to two-stage architectures such as the Mask R-CNN, which would primarily require segmentation maps to be trained for the segmentation task [61]. Moreover, future approaches could use unlabeled data with a semi-supervised learning approach such as the self-training [62], which has been successfully applied for segmentation of medical images [63].

5.4. Limitations

Despite the benefits of adopting an automatic wound detection step before the segmentation, cases of inaccurate wound detection must also be considered. The wound detector could produce bounding boxes with very large portions of the surrounding skin and even partial background or could lead to false positive and negative detections. The case of inaccurate but true positive detections could be mitigated by using bigger and richer data sets for the training of segmentation models. The risk of false positive detection could be solved by training the segmentation models with negative examples, such as images with region of healthy skin only or random background [41]. Furthermore, the case of false negative detection could be mitigated by introducing feedback mechanisms at the image collection time, which would notify the user to re-take the picture or change framing.

While we tested the performance of the DS approach with data from independent data sets, the study cohort of the SwissWOU data set used for training the segmentation models is likely not representative for all wound patients. It is evident that a validation study with larger and clinically representative cohort with multiple ethnicities and skin colors will be necessary to establish the performance of the DS approach for wound image segmentation. Additional clinical validation should also be conducted to evaluate whether the DS approach could be used to measure the wound bed area, which constitutes one of the most relevant parameters to assess the wound healing status and whether this process simplifies the clinical routines.

6. Conclusion

We presented a deep learning approach to improve generalization of wound image segmentation. We found that using the fusion of automated wound detection and segmentation improves the segmentation performance and enables the segmentation model to generalize well on out-of-distribution wound images. The DS approach preserved segmentation performance when data with limited size was used for training to segment wound images. We demonstrated experimentally that our approach can be used to segment wound images from smartphones or professional digital cameras without the need of user interaction or background standardization during image collection. Our results confirm that machine learning based image analysis could be implemented on mobile phones to support remote wound care in the future.

Declaration of competing interest

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.imu.2022.100884.

References

[1] Lindholm C, Searle R. Wound management for the 21st century: combining effectiveness and efficiency. Int Wound J 2016;13:5–15. http://dx.doi.org/10.1111/iwj.12623.

[2] Escandon J, Vivas AC, Tang J, Rowland KJ, Kirsner RS. High mortality in patients with chronic wounds. Wound Repair Regen 2011;19(4):526–8.

[3] Sen CK. Human wounds and its burden: An updated compendium of estimates. 2019, http://dx.doi.org/10.1089/wound.2019.0946.

[4] Han G, Ceilley R. Chronic wound healing: a review of current management and treatments. Adv Therapy 2017;34(3):599–610.

[5] Othman D. Negative pressure wound therapy literature review of efficacy, cost effectiveness, and impact on patients' quality of life in chronic wound management and its implementation in the United Kingdom. Plast Surg Int 2012;2012:1–6. http://dx.doi.org/10.1155/2012/374398.

[6] Khalil H, Cullen M, Chambers H, Steers N, Walker J. Implementation of a successful electronic wound documentation system in rural victoria, australia: a subject of collaboration and community engagement. Int Wound J 2014;11(3):314–8.

[7] Hess CT. The art of skin and wound care documentation. Adv Skin Wound Care 2005;18(1):43–53.

[8] Wang L, Pedersen PC, Strong DM, Tulu B, Agu E, Ignotz R. Smartphone-based wound assessment system for patients with diabetes. IEEE Trans Biomed Eng 2015;62(2):477–88. http://dx.doi.org/10.1109/TBME.2014.2358632.

[9] Goyal M, Reeves ND, Rajbhandari S, Yap MH. Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. IEEE J Biomed Heal Inform 2019;23(4):1730–41. http://dx.doi.org/10.1109/JBHI.2018.2868656.

[10] Zhang J, Mihai C, Tüshaus L, Scebba G, Distler O, Karlen W, et al. Wound image quality from a mobile health tool for home-based chronic wound management with real-time quality feedback: Randomized feasibility study. JMIR mHealth uHealth 2021;9(7):e26149. http://dx.doi.org/10.2196/26149.

[11] Sheehan P, Jones P, Caselli A, Giurini JM, Veves A. Percent change in wound area of diabetic foot ulcers over a 4-week period is a robust predictor of complete healing in a 12-week prospective trial. Diabetes Care 2003;26(6):1879–82. http://dx.doi.org/10.2337/diacare.26.6.1879.

[12] Lucas Y, Niri R, Treuillet S, Douzi H, Castaneda B. Wound size imaging: Ready for smart assessment and monitoring, Advances in wound careAhead of print.

[13] Rose-McGuckin K. A guide to good practice in wound management photography. Institute of Medical Illustrators - UK; 2019.

[14] Dou Q, Chen H, Yu L, Zhao L, Qin J, Wang D, Mok VC, Shi L, Heng PA. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Trans Med Imaging 2016;35(5):1182–95. http://dx.doi.org/10.1109/TMI.2016.2528129.

[15] Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A, Lian K, Kambhampati S, Kijowski R. Deep learning approach for evaluating knee MR images: Achieving high diagnostic performance for cartilage lesion detection. Radiology 2018;289(1):160–9. http://dx.doi.org/10.1148/radiol.2018172986.

[16] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88. http://dx.doi.org/10.1016/J.MEDIA.2017.07.005.

[17] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Lect. notes comput. sci. (Including subser. lect. notes artif. intell. lect. notes bioinformatics), Vol. 9351. Springer Verlag; 2015, p. 234–41. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[18] Lugagne J-B, Lin H, Dunlop MJ. DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. PLOS Comput Biol 2020;16(4):e1007673. http://dx.doi.org/10.1371/journal.pcbi.1007673.

[19] Xie Y, Zhang Z, Sapkota M, Yang L. Spatial clockwork recurrent neural network for muscle perimysium segmentation. In: Lect. notes comput. sci. (Including subser. lect. notes artif. intell. lect. notes bioinformatics). LNCS, vol. 9901, Springer Verlag; 2016, p. 185–93. http://dx.doi.org/10.1007/978-3-319-46723-8_22.

[20] Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, Cetin I, Lekadir K, Camara O, Gonzalez Ballester MA, Sanroma G, Napel S, Petersen S, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Trans Med Imaging 2018;37(11):2514–25. http://dx.doi.org/10.1109/TMI.2018.2837502.

[21] De Brébisson A, Montana G. Deep neural networks for anatomical brain segmentation. In: IEEE comput. soc. conf. comput. vis. pattern recognit. work., Vol. 2015-Octob. IEEE Computer Society; 2015, p. 20–8. http://dx.doi.org/10.1109/CVPRW.2015.7301312, arXiv:1502.02445.

[22] Moeskops P, Wolterink JM, van der Velden BHM, Gilhuijs KGA, Leiner T, Viergever MA, Išgum I. Deep learning for multi-task medical image segmentation in multiple modalities. LNCS, vol. 9901, Springer Verlag; 2017, p. 478–86. http://dx.doi.org/10.1007/978-3-319-46723-8_55, arXiv:1704.03379.

[23] Harrison AP, Xu Z, George K, Lu L, Summers RM, Mollura DJ. Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images. In: Lect. notes comput. sci. (Including subser. lect. notes artif. intell. lect. notes bioinformatics). LNCS, vol. 10435, Springer Verlag; 2017, p. 621–9. http://dx.doi.org/10.1007/978-3-319-66179-7_71, arXiv:1706.03702.

[24] Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: Hybrid densely connected unet for liver and tumor segmentation from CT volumes. IEEE Trans Med Imaging 2018;37(12):2663–74. http://dx.doi.org/10.1109/TMI.2018.2845918, arXiv:1709.07330.

[25] Attia M, Hossny M, Nahavandi S, Yazdabadi A. Skin melanoma segmentation using recurrent and convolutional neural networks. In: Proc. - Int. symp. biomed. imaging. IEEE Computer Society; 2017, p. 292–6. http://dx.doi.org/10.1109/ISBI.2017.7950522.

[26] Fauzi MFA, Khansa I, Catignani K, Gordillo G, Sen CK, Gurcan MN. Computerized segmentation and measurement of chronic wound images. Comput Biol Med 2015;60:74–85.

[27] Fauzi MFA, Khansa I, Catignani K, Gordillo G, Sen CK, Gurcan MN. Segmentation and management of chronic wound images: A computer-based approach. Cham: Springer; 2018, p. 115–34. http://dx.doi.org/10.1007/15695_2018_131.

[28] Wannous H, Treuillet S, Lucas Y. Supervised tissue classification from color images for a complete wound assessment tool. In: Annu. int. conf. ieee eng. med. biol. - Proc.. 2007, p. 6031–4. http://dx.doi.org/10.1109/IEMBS.2007.4353723.

[29] Wang L, Pedersen PC, Agu E, Strong DM, Tulu B. Area determination of diabetic foot ulcer images using a cascaded two-stage SVM-based classification. IEEE Trans Biomed Eng 2017;64(9):2098–109. http://dx.doi.org/10.1109/TBME.2016.2632522.

[30] Veredas F, Mesa H, Morente L. Binary tissue classification on wound images with neural networks and bayesian classifiers. IEEE Trans Med Imaging 2010;29(2):410–27. http://dx.doi.org/10.1109/TMI.2009.2033595.

[31] Song B, Sacan A. Automated wound identification system based on image segmentation and artificial neural networks. In: Proc. - 2012 IEEE int. conf. bioinforma. biomed. BIBM 2012. 2012, p. 619–22. http://dx.doi.org/10.1109/BIBM.2012.6392633.

[32] Veredas FJ, Luque-Baena RM, Martín-Santos FJ, Morilla-Herrera JC, Morente L. Wound image evaluation with machine learning. Neurocomputing 2015;164:112–22. http://dx.doi.org/10.1016/j.neucom.2014.12.091.

[33] Bhavani RR, Wiselin Jiji G. Image registration for varicose ulcer classification using KNN classifier. Int J Comput Appl 2018;40(2):88–97. http://dx.doi.org/10.1080/1206212X.2017.1395108.

[34] Wang C, Anisuzzaman DM, Williamson V, Dhar MK, Rostami B, Niezgoda J, Gopalakrishnan S, Yu Z. Fully automatic wound segmentation with deep convolutional neural networks. Sci Rep 2020;10(1):21897. http://dx.doi.org/10.1038/s41598-020-78799-w, arXiv:2010.05855.

[35] Wang C, Yan X, Smith M, Kochhar K, Rubin M, Warren SM, Wrobel J, Lee H. A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. In: 2015 37th annu. int. conf. IEEE eng. med. biol. soc.. IEEE; 2015, p. 2415–8. http://dx.doi.org/10.1109/EMBC.2015.7318881.

[36] Rother C, Kolmogorov V, Blake A. GrabCut - Interactive foreground extraction using iterated graph cuts. In: ACM SIGGRAPH 2004 pap. SIGGRAPH 2004. New York, New York, USA: ACM Press; 2004, p. 309–14. http://dx.doi.org/10.1145/1186562.1015720.

[37] Cui C, Thurnhofer-Hemsi K, Soroushmehr R, Mishra A, Gryak J, Domínguez E, Najarian K, López-Rubio E. Diabetic wound segmentation using convolutional neural networks. In: 2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2019, p. 1002–5.

[38] 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.

[39] Liu X, Wang C, Li F, Zhao X, Zhu E, Peng Y. A framework of wound segmentation based on deep convolutional networks. In: Proc. - 2017 10th int. congr. image signal process. biomed. eng. informatics, CISP-BMEI 2017, Vol. 2018-Janua. Institute of Electrical and Electronics Engineers Inc.; 2018, p. 1–7. http://dx.doi.org/10.1109/CISP-BMEI.2017.8302184.

[40] Dorileo ÉA, Frade MA, Rangayyan RM, Azevedo-Marques PM. Segmentation and analysis of the tissue composition of dermatological ulcers. In: CCECE 2010. IEEE; 2010, p. 1–4.

[41] Goyal M, Yap MH, Reeves ND, Rajbhandari S, Spragg J. Fully convolutional networks for diabetic foot ulcer segmentation. In: 2017 IEEE int. conf. syst. man, cybern.. IEEE; 2017, p. 618–23. http://dx.doi.org/10.1109/SMC.2017.8122675.

[42] Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. IEEE Trans Pattern Anal Mach Intell 2017;42(2):318–27, arXiv:1708.02002.

[43] Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: Common objects in context. In: Lect. notes comput. sci. (Including subser. lect. notes artif. intell. lect. notes bioinformatics). LNCS, vol. 8693, Springer Verlag; 2014, p. 740–55. http://dx.doi.org/10.1007/978-3-319-10602-1_48, arXiv:1405.0312.

[44] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. IEEE comput. soc. conf. comput. vis. pattern recognit. Vol. 2016-Decem. IEEE Computer Society; 2016, p. 770–8. http://dx.doi.org/10.1109/CVPR.2016.90, arXiv:1512.03385.

[45] Tan M, Pang R, Le QV. EfficientDet: Scalable and efficient object detection. In: Proc. IEEE comput. soc. conf. comput. vis. pattern recognit.. 2019, p. 10778–87, arXiv:1911.09070.

[46] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv arXiv:1704.04861.

[47] Kingma DP, Ba JL. Adam: A method for stochastic optimization. In: 3rd int. conf. learn. represent. ICLR 2015 - Conf. track proc. international conference on learning representations, ICLR. 2015, arXiv:1412.6980.

[48] Cassidy B, Reeves ND, Pappachan JM, Gillespie D, O'Shea C, Rajbhandari S, Maiya AG, Frank E, Boulton AJ, Armstrong DG, et al. The dfuc 2020 dataset: Analysis towards diabetic foot ulcer detection. Eur Endocrinol 2021;1(1):5–11.

[49] Clark A. Pillow (PIL Fork) documentation. 2015, visited on 2020-06-01. URL https://pillow.readthedocs.io/en/stable/.

[50] Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. 2017, arXiv. arXiv:1706.05587.

[51] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Bengio Y, LeCun Y, editors. 3rd international conference on learning representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference track proceedings. 2015.

[52] Medtec, Medetec Surgical Dressings and Wound Management Resource Centre - Home page, visited on 2021-03-16 (Dataset). URL http://www.medetec.co.uk/index.html.

[53] Yang S, Park J, Lee H, Kim S, Lee B-U, Chung K-Y, Oh B. Sequential change of wound calculated by image analysis using a color patch method during a secondary intention healing. PLoS One 2016;11(9):e0163092. http://dx.doi.org/10.1371/journal.pone.0163092.

[54] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(2):377–81. http://dx.doi.org/10.1016/j.jbi.2008.08.010.

[55] Boughorbel S, Jarray F, El-Anbari M. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS One 2017;12(6):e0177678. http://dx.doi.org/10.1371/journal.pone.0177678.

[56] Eelbode T, Bertels J, Berman M, Vandermeulen D, Maes F, Bisschops R, Blaschko MB. Optimization for medical image segmentation: Theory and practice when evaluating with dice score or Jaccard index. IEEE Trans Med Imaging 2020;39(11):3679–90. http://dx.doi.org/10.1109/tmi.2020.3002417, arXiv:2010.13499.

[57] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing 2018;321:321–31. http://dx.doi.org/10.1016/j.neucom.2018.09.013, arXiv:1803.01229.

[58] Codella NC, Gutman D, Celebi ME, Helba B, Marchetti MA, Dusza SW, Kalloo A, Liopyris K, Mishra N, Kittler H, 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). 2018, p. 168–72.

[59] Tschandl P, Rosendahl C, Kittler H. Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 2018;5(1):1–9. http://dx.doi.org/10.1038/sdata.2018.161.

[60] Zheng S, Lu J, Zhao H, Zhu X, Luo Z, Wang Y, Fu Y, Feng J, Xiang T, Torr PH, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 6881–6890.

[61] He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. 2017, p. 2961–9.

[62] Xie Q, Luong M-T, Hovy E, Le QV. Self-training with noisy student improves imagenet classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, p. 10687–98.

[63] Zhang L, Shi Y, Yao J, Bian Y, Cao K, Jin D, Xiao J, Lu L. Robust pancreatic ductal adenocarcinoma segmentation with multi-institutional multi-phase partially-annotated ct scans. In: International conference on medical image computing and computer-assisted intervention. Springer; 2020, p. 491–500.