Segmenting Skin Ulcers and Measuring the Wound Area Using Deep Convolutional Networks

Daniel Y. T. Chino[a], Lucas C. Scabora[a], Mirela T. Cazzolato[a], Ana E. S. Jorge[b], Caetano Traina-Jr.[a], Agma J. M. Traina[a]

[a] *Institute of Mathematical and Computer Sciences, University of Sao Paulo*
{chinodyt, caetano, agma}@icmc.usp.br; {lucascsb, mirelac}@usp.br
[b] *Department of Physical Therapy, Federal University of Sao Carlos*
anajorge@alumni.usp.br

Abstract

Background and Objectives: Bedridden patients presenting chronic skin ulcers often need to be examined at home. Healthcare professionals follow the evolution of the patients' condition by regularly taking pictures of the wounds, as different aspects of the wound can indicate the healing stages of the ulcer, including depth, location, and size. The manual measurement of the wounds' size is often inaccurate, time-consuming, and can also cause discomfort to the patient. In this work, we propose the Automatic Skin Ulcer Region Assessment (ASURA) framework to accurately segment the wound and automatically measure its size.
Methods: ASURA uses an encoder/decoder deep neural network to perform the segmentation, which detects the measurement ruler/tape present in the image and estimates its pixel density.
Results: Experimental results show that ASURA outperforms the state-of-the-art methods by up to 16% regarding the Dice score, being able to correctly segment the wound with a Dice score higher than 90%. ASURA automatically estimates the pixel density of the images with a relative error of 5%. When using a semi-automatic approach, ASURA was able to estimate the area of the wound in square centimeters with a relative error of 14%.
Conclusions: The results show that ASURA is well-suited for the problem of segmenting and automatically measuring skin ulcers.

Keywords: skin ulcer, image segmentation, deep convolutional neural networks, wound measurement

1. Introduction

Clinical environments are increasingly improving the capability of generating large amounts of images, exams, and related information. Advances in technologies such as cameras, storage infrastructure, and clinical equipment improve the information quality [1, 2]. Bedridden patients, due to specific health conditions, need to be examined at home. This is the case of patients presenting chronic skin wounds, referred to as skin ulcers.

Skin ulcers appear due to different reasons, including poor blood circulation, injuries, infections, tumors, and other skin conditions [3, 4, 5]. The visual appearance of the wounds can provide clinical signs that lead healthcare professionals during the diagnosis process. Venous skin ulcers are wounds with different healing stages, namely fibrin, granulation, callous and necrosis. Healthcare professionals and caretakers follow-up the healing evolution of the patients by regularly taking photographs of the wound. The healing time of each wound depends on multiple factors, including depth, location, patient age, local and systemic disease, as well as the wound size [6].

The introduction of Deep Learning (DL) techniques [7] has motivated several works to use Convolutional Neural Networks (CNNs) on the medical domain. Photograph images have been used to recognize melanomas by segmenting [8, 9, 10] or classifying [11, 12] them, for foot ulcer segmentation [13] and pressure ulcer segmentation and classification [14]. However, to the best of our knowledge, there are few works that deal with skin ulcer images using DL techniques [15]. Those works do not provide the wound measurement considering the real world unit where the image was taken from, as our proposed method does.

In this paper, we are particularly interested in automatically assessing the wound area of skin ulcer images. This task requires two major steps: to segment the ulcer region and to evaluate the wound area attributes. The image segmentation task refers to locating the boundary between the wound and the surrounding skin [16]. Measurements of ulcer attributes, such as their size, and their comparison with previous measurements, allow health professionals to assess patient health outcomes. The ulcer area estimation is usually performed manually, which can be a time consuming and inaccurate task [17], besides being a source of discomfort to the patients. Accurate and automatic wounds measurement strongly relies on well-segmented regions. Existing works lack the required accurate segmentation, as they focus more on the retrieval and classification tasks [4, 1, 18, 17, 19].

Based on such a scenario, in this work we propose the Automatic Skin Ulcer Region Assessment (ASURA) framework. ASURA uses CNNs and automatically detects and segments the ulcer wounds and the measurement tools depicted in the images, such as measurement rulers and/or tapes. ASURA automatically computes the wound size, helping healthcare professionals and caretakers in the analysis of the patient's image and healing evolution. Accordingly, the contributions of ASURA are two-fold:

- **Segmentation**: ASURA segments regions depicting skin ulcers and measurement rulers/tapes using deep CNN; and
- **Area measurement**: ASURA automatically computes the area of the wound in real-world units, based on the pixel density information and the segmented measurement ruler/tape.

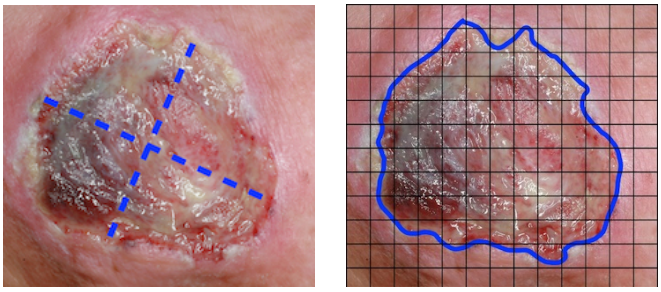
We provide an extensive experimental analysis, comparing our proposal with the state-of-the-art methods for skin ulcer segmentation. Regarding the wound area measurement, we compare the obtained results of ASURA with manually annotated segmentations, showing the high precision of our method.

Paper outline. Section 2 presents the challenges regarding the wound area estimation task. Section 3 discusses basic concepts used in this paper. Section 4 discusses the related work. Section 5 details the ASURA framework. Section 6 presents the materials and methods used in our evaluation. Section 7 presents the experimental results and the corresponding discussion. Finally, Section 8 concludes our work. For reproducibility, our code is open-sourced at `https://github.com/chinodyt/asura`.

2. Area Estimation Challenges

To follow-up the healing evolution of a wound, healthcare professionals need to quantify their visual characteristics. In this paper, we focus on how to measure its main attributes. Several non-automatic standard protocols exist to measure the wound's area [20]. The quickest and most used one is to calculate the area of the minimum bounding rectangle (MBR) using the longest length of the wound by the largest perpendicular width with a ruler/metric tape (Figure 1(a)). Another method covers the wound with a transparent plastic film, and the physician or a trained healthcare professional trace the border of the wound with a pencil marker. The area is calculated by counting how many squares the wound covers (Figure 1(b)). However, those methods

are invasive (demand some contact with the wound) and may be imprecise. One way to overcome this problem is to take advantage of digital images and automatize the process.



(a) MBR (b) Transparent film

Figure 1: Example of standard protocols to asses the wound area size.

Different from other medical images, such as CT (computerized tomography) scans, MRI (magnetic resonance imaging) and ultrasound, there is no specific equipment to capture images of skin wounds. Healthcare professionals take pictures from the whole wound and perform comparisons spanning distinct healing stages at different periods using different equipments. Usually, the pictures are taken using digital cameras or smartphones.

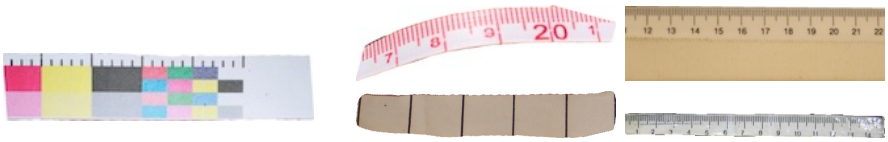


Figure 2: Example of measurement rulers/tapes.

The area of the wound can be calculated using an image processing software, *e.g.*, ImageJ [21]. However, the images may have different scales, since there is no standard protocol regarding the position of the camera while taking the pictures. For this problem, healthcare professionals use a reference object, such as rulers or metric tapes, while taking the picture. Figure 2 shows examples of measurement rulers/tapes that are used by healthcare professionals. Therefore, it is possible to estimate the area of the wound in a real-world unit. However, these tasks are tedious and can be inaccurate, when such measurement is manually done, as the number of patients grows. An automatic and accurate wound segmentation and measurement approach

is the main challenge we address in this paper. On the next section, we will discuss segmentation methods for wound regions in ulcer images.

3. Background

One of the concepts used in this work is to use deep convolutional neural networks to perform a semantically rich segmentation. Nowadays, deep CNN are largely used in image recognition tasks [22]. A CNN consists of a series of convolution operations, which encodes an image into a feature map, and can be used to perform image segmentation. Deep architectures designed for segmentation interpret the feature map as a segmentation mask. The Fully Convolutional Networks (FCN) [23] learns to make a pixel-wise prediction of the input image, which generates a mask as output. The SegNet [24] decodes the feature map by upscaling and using a series of convolutions. However, those networks require thousands of annotated training samples. To overcome this, Ronneberger *et al.* [25] proposed the U-Net, an encoder/decoder FCN able to deal with a smaller training set. In U-Net, the decoder receives a copy of the output of the activation layers and concatenates it with the upscaling tensor. In this way, U-Net can pass the spatial information lost in the encoder step to the corresponding decoder layers, improving the segmentation output.

Another segmentation method is the DeepLab [26], which uses atrous convolutions to upscale the features. Atrous convolutions allow enlarging the field of view of filters with a simpler architecture. Several extensions were proposed to improve the DeepLab model. First, they used atrous spatial pyramid pooling to allow the representation of objects in multiple scales. Then, they improved the capability of the DeepLab model to capture sharper object boundaries by using an encoder-decoder with atrous convolutions. The latest iteration, the DeepLabv3+ [26], includes an effective decoder to refine the segmentation. On the encoder, the max pooling operations are replaced with depth-wise separable convolutions. The decoder concatenates the encoded features upscaled by a factor of 4 with the corresponding low-level features.

Some FCN integrate feature maps from multiple convolutional layers to improve the image segmentation results. The Contextual Aggregating Network (CAN) [27] detects objects using context clues, and aggregates feature maps from intermediary convolutional layers to identify shapes and boundaries in the image. The Multi-Scale Deep Context Convolutional Network

(MDCCNet) [28] integrates the three deepest convolutional layers using Conditional Random Fields (CRF). By employing a densely connected CRF postprocessing, it was possible not only to detect the object shapes and boundaries, but also to eliminate false positives in the segmentation process. Finally, both CAN and MDCCNet are FCN-based approaches focused on segmenting generic objects, such as planes, cars, and buildings, which are different from images used in the medical and healthcare systems.

4. Related Works

There are a few works that deal with skin ulcer images. Dorileo *et al.* [4] proposed an image segmentation method, which is based on the analysis of the RGB channels of the image. Dorileo *et al.* took advantage of a controlled environment to obtain the images. Since all images had a blue background, they discarded the blue channel and also used the intensity channel of the HSI (hue, saturation, intensity) color space. Each channel is used to find a distinct type of tissue: fibrin, granulation, and necrotic. For each channel, the method automatically finds thresholds and process the discovered regions by focusing on blobs near the center of the image. The main problem of this method is the need for a controlled environment.

Another skin ulcer segmentation method was proposed by Seixas *et al.* [18]. Seixas *et al.* employed off-the-shelf classifiers to segment the ulcer images. They extracted pixel-wise color features, the mean value of the neighborhood of the pixel, and the difference of the pixel value and the mean beforehand mentioned. The authors segmented a training set of images to isolate the wound region.

There are also some works that use segmentation as an intermediary step for different tasks, such as classification and image retrieval. Pereyra *et al.* [1] employed texture features and a k-nearest-neighbor classifier, creating a confusion matrix to distinguish the images among granulation, fibrin, or mixed tissues. As a processing step, Pereyra *et al.* proposed a segmentation step based on a multivariate gaussian mixture model. The clusters were manually selected in a Graphical User Interface (GUI) to output the segmentation mask.

A few other methods use superpixels algorithms as intermediate steps to simplify the image complexity. Blanco *et al.* [17] proposed the Counting-Labels Similarity Measure (CL-Measure), a similarity measure based on color and shape features. Blanco *et al.* segment the image by classifying features

extracted from superpixels. Chino *et al.* [19] proposed Imaging Content Analysis for the Retrieval of Ulcer Signatures (ICARUS), an image retrieval based on bag-of-visual words. ICARUS segments the wound parts of the image to improve the retrieval precision. The segmentation is also based on superpixels classification.

Silva *et al.* [29] proposed 2PLA (**T**wo-**P**hase **L**earning **A**pproach), a superpixel-based approach for the segmentation of lower limbs presenting dermatological ulcers. The first phase of 2PLA removes the background region of the image using pixel-wise classification and mathematical morphology, resulting in an intermediary segmentation image of the limb presenting the ulcer. The second phase extracts superpixels from the segmented image, further grouping similar superpixels. As the output, 2PLA provides a pixel-wise area quantification of each resulting group of superpixels. Although 2PLA presents very promising results on the segmentation of limbs, it does not segment the wound region. Therefore, the quantification of each region is performed on all regions containing both healthy skin and wound tissues. This does not meet the requirement of measuring the wound area only, which is important for the follow-up on the evolution of dermatological ulcers by the specialists. Another drawback is that 2PLA was designed to work on images with a white or bluish homogeneous background, colors that are not expected to be within the visual aspect of dermatological ulcers or skin. They employed the dataset available in [3]. Our method, on the other hand, segments and quantifies the wound region, independently on the background color.

Blanco *et al.* [15] proposed QTDU, a deep-learning-based approach to analyze dermatological wounds using superpixels. QTDU uses CNNs for the wound segmentation, and rely on superpixel approaches to divide images into regions. Their method feeds the network with superpixel regions instead of single images. QTDU uses the CNN networks ResNet and InceptionV3. The authors added six new layers at the end of the underline CNN, where the final layer provides the label for the superpixel. The specialists that labeled their training data (using superpixel regions) picked images from the dataset at random, aiming at maximizing diversity regarding tissue dominance, skin color, age, and treatment [15]. The dataset employed is the same reported in works [29, 3], where every image have white and bluish color in the background. Their results on classification seem promising (with up to 0.986 of AUC), but they heavily depend on the superpixel approach, which should adequately adhere to the wound borders. Moreover,

they focus on the classification of the different wound tissues into one of the labels L = {Not Wound, Fibrosis, Necrosis, Granulation}. QTDU does not segment the rules/tapes present in the images, and the estimation of the wound area consists of counting the number of pixels inside the segmented area of each found tissue and checking this value proportionally to the number of pixels of the entire image.

Table 1 compares our proposed method with the existing methods for skin ulcer segmentation. The majority of methods are able to segment skin ulcer wounds, except for 2PLA, which focuses on segmenting the limbs with skin ulcers. Some methods use superpixels to reduce the complexity of the image while doing the segmentation, which can compromise the wound segmentation. One important aspect of ASURA is the detection of the measurement tools, which only the DL based methods can achieve after proper training. However, none of the aforementioned related works can measure the segmentation area of the wound using real-world units. Our proposed method, on the other hand, converts the size of the wound into squared centimeters (cm^2), which is the measure used in the daily routine of medical practitioners.

Table 1: Summary of different methods to segment skin ulcer wounds.

	Dorileo [4]	Seixas [18]	Pereyra [1]	CL-Measure [17]	ICARUS [19]	2PLA [29]	DL Based [25, 26]	QTDU [15]	ASURA
Wound Segmentation	✓	✓	✓	✓	✓		✓	✓	✓
Pixel-wise Segmentation	✓	✓	✓				✓		✓
Detect Measurement Tool							✓		✓
Area in Real World Units									✓

5. ASURA

In this section, we introduce the Automatic Skin Ulcer Region Assessment (ASURA) framework, which measures the area of wounds on skin ulcer images. ASURA takes advantage of a deep learning approach to segment the

skin ulcer wound. By analyzing objects such as measurement rulers/tapes on the images, ASURA is able to estimate the area of the ulcer wound in real-world units. ASURA works performing two steps: (A) Ulcer Segmentation and (B) Pixel Density Estimation. Figure 3 shows the ASURA's architecture. ASURA also offers an interactive GUI in which the user can analyze the provided automatic Pixel Density estimation. Alternatively, the user has the option of interacting with the system and setting a more suitable Pixel Density.

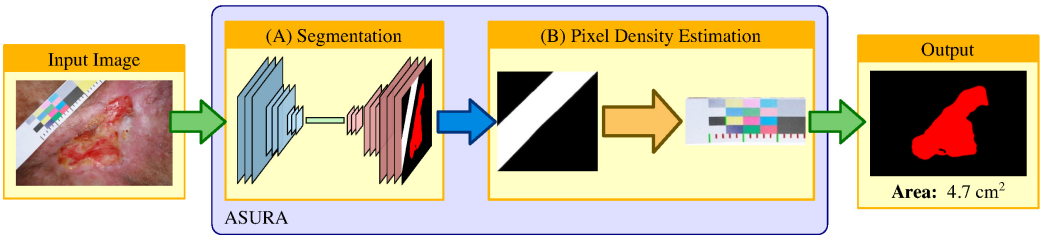


Figure 3: ASURA framework.

5.1. Ulcer Segmentation

In the segmentation step, ASURA receives a skin ulcer RGB image and outputs a segmentation mask with both the wound and the measurement ruler/tape. Ulcer Segmentation is based on a convolutional deep neural network, previously developed for image segmentation. Since the size of the training dataset is limited, ASURA uses an architecture model based on the U-Net [25], which has shown to be appropriate for this scenario.

Figure 4 shows the model of the architecture used. The network consists of an encoder and a decoder. First, ASURA receives as input an RGB image with arbitrary resolution. Since the input layer of the network is a tensor of size (512x512x3), the image is resized to a 512x512 resolution. The encoding phase consists of repeatedly applying two 3x3 padded convolutions followed by an exponential linear unit (ELU). A 2x2 max pooling then halves the tensor size. The decoder consists of repeatedly applying a 2x2 deconvolution, which halves the depth of the tensor. This tensor is then concatenated with the corresponding tensor on the encoding phase. Then, two 3x3 padded convolutions followed by an exponential linear unit (ELU) are applied once more. The output layer of the network is a 1x1 sigmoid convolution to map the 16 layers of the decoded tensor into the three classes (wound, measurement ruler/tape, background). In the final step, the output tensor is resized

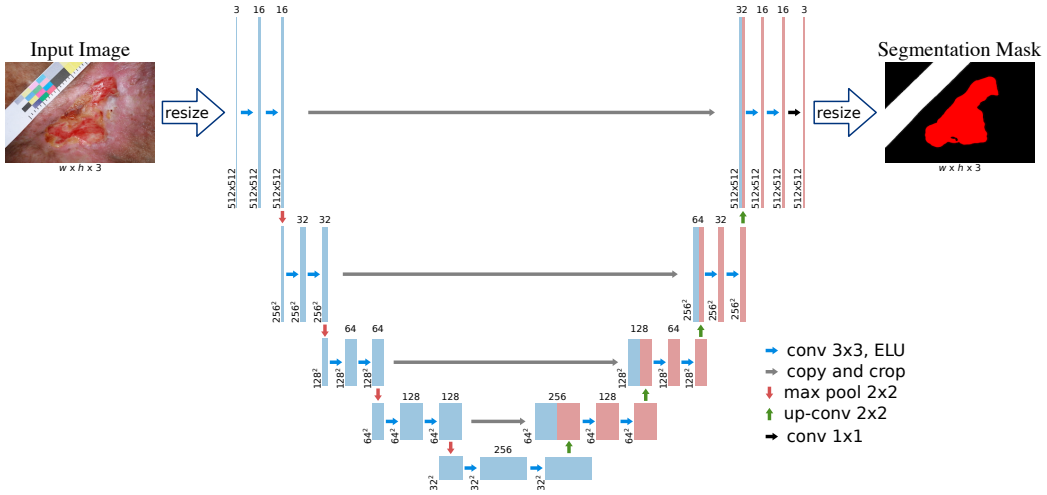


Figure 4: The architecture of the network used by ASURA. The blue tensors are encoders and the red tensors are decoders.

to the resolution of the input image. A heaviside step function is applied to the resized segmentation mask.

5.2. Pixel Density Estimation

After the Ulcer Segmentation step, ASURA processes the objects resembling the measurement rulers/tapes present in the image to estimate the Pixel Density (λ) of the image. With the segmentation mask and knowing the Pixel Density of the image, it is possible to estimate the area of the wound in real-world units. Figure 5 shows the steps used by ASURA to estimate λ.

The Pixel Density Estimation step receives as input the image and a segmentation mask of the ulcer wound, as well as the measurement ruler/tape obtained in the first step of ASURA. Although there are at least five different rulers/measurement tapes (Figure 2), they all belong to the same class. The following steps are applied to all rulers/measurement tapes. (a) Using the segmentation mask, ASURA crops the measurement ruler/tape and deletes the image's background (Figure 5(a)). (b) To simplify the measurement ruler/tape processing, ASURA finds the orientation of the ruler and rotates the image horizontally (Figure 5(b)). (c) On the next step, ASURA binarizes the image (Figure 5(c)). The image is converted to grayscale and passes through an auto-threshold method using the ISODATA algorithm. Then, a vertical edge detector filter is applied to the binarized image. (d) After

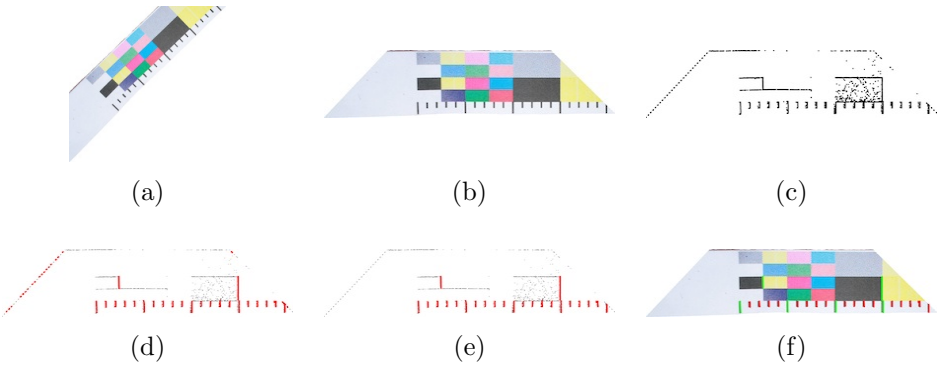


Figure 5: Pixel Density Estimation steps. (a) Find and crop the measurement ruler/tape in the image. (b) Rotate the measurement ruler/tape. (c) Binarize the image. (d) Detect lines in the image. (e) Keep only the most frequent parallel lines. (f) Group ticks by size and calculates the distance between ticks.

the edge detector filter, ASURA uses the Line Segment Detector (LSD) [30] to find the ticks of the measurement ruler/tape (Figure 5(d)). (e) With the ticks detected as line segments, ASURA groups them by its angle and keeps only the most frequent parallel ticks (Figure 5(e)). To do so, ASURA analyzes the distribution of the angles and considers only the angles within a threshold of 5 degrees of the angle with the highest probability. (f) As the last step, ASURA groups the ticks by size and calculates the distance in pixels between the ticks in each group (Figure 5(f)). Once again, ASURA analyzes the distribution of the calculated distances and returns the distance with the highest probability. It is important to note that some images may have more than one measurement ruler/tape. On these cases, these steps are repeated for each segmented measurement ruler/tape. Algorithm 1 shows how ASURA estimates the Pixel Density λ.

The area of the wound can be estimated using Equation 1, where $|Mask|$ is the size of the wound from the segmentation mask.

$$A = \frac{|Mask|}{\lambda^2} \tag{1}$$

5.3. Graphical User Interface (GUI)

ASURA has an interactive GUI that allows the user to browse the image and to analyze the segmentation of the outputted mask and the Pixel Density estimation. The interactive interface also allows the user to indicate a better Pixel Density estimation when needed.

Algorithm 1 Pixel Density Estimation algorithm.

Require: I: input image, $Mask$: segmentation mask
Ensure: $\boldsymbol{\lambda}$: distances in pixels between ticks

$ruler_I$, $ruler_{mask} \leftarrow$ cropRuler(I, $Mask$)
$ruler_I \leftarrow$ findAngleAndRotate($ruler_I, ruler_{mask}$)
$ruler_{bw} \leftarrow$ binarizeImage($ruler_I$)
$ruler_{edge} \leftarrow$ verticalLineFilter($ruler_{bw}$)
$lines \leftarrow$ detectLines($ruler_{edge}$)
$lines_0 \leftarrow$ getMostFrequentParallel($lines$)
$\boldsymbol{L} \leftarrow$ groupBySize($lines_0$)
$\boldsymbol{\lambda} \leftarrow \emptyset$
for all $l_i \in \boldsymbol{L}$ **do**
    $\lambda_i \leftarrow$ calculateDistance(l_i)
    Add λ_i to $\boldsymbol{\lambda}$
end for
return $\boldsymbol{\lambda}$

Figure 6 shows the ASURA's GUI. On the area highlighted in green (1), the user can see the input image and the output of the Ulcer Segmentation. On the area highlighted in blue (2), the user can draw a line on the measurement ruler/tape image to indicate the length of the real world unit he/she desires. If the user indicates the length, ASURA shows the segmented area in this real world unit. Finally, on the area highlighted in red (3), ASURA shows the Pixel Density estimation. ASURA marks the detected ticks and assumes the red line below the measurement ruler/tape as the estimated distance between the ticks. If ASURA detects more than one tick size, it uses a different color for each tick size. The user can input a value to multiply the estimated Pixel Density to estimate the segmented area in real-world units.

6. Materials and Methods

We evaluated the performance of ASURA to estimate the area of the wounds in ulcer images. To carry the evaluation of ASURA, we run two sets of experiments: Ulcer Segmentation and Pixel Density estimation. We implemented

Figure 6: ASURA's interactive graphical user interface.

ASURA in Python with the Keras library[1], using TensorFlow[2] backend. All experiments were carried out on a 4.20GHz Intel Core i7-7700k CPU with 16GB RAM and an NVIDIA Titan Xp with 12GB GDDR5X, running CentOS 7.

6.1. Datasets

We evaluated ASURA on two skin ulcer datasets: `ULCER` and `ULCER-2`. `ULCER` [3] is composed of 217 dermatological images that originated from both venous or arterial insufficiencies. The wounds were located on the inferior limbs with different sizes and healing stages. Only one wound per patient was included, and the majority of the patients' skin color was white. The `ULCER-2` is composed of 229 dermatological images taken from 23 patients of the Learning Health Center of the University of Sao Paulo, at Ribeirao Preto,

[1]https://keras.io/
[2]https://www.tensorflow.org/

also originated from both venous or arterial insufficiencies. The wounds were also located on the inferior limbs with different sizes and healing stages. For each patient, a series of images were taken along a period of 90 days, with an average of 10 images per patient. On both datasets, the images were taken using a digital camera. However, the acquisition procedure was different. While `ULCER` dataset followed a controlled protocol (same equipment, background, and measurement tape), `ULCER-2` has a more diverse set of images, with different backgrounds and measurement tools. An anonymization process removed personal data. For both datasets, experts manually segmented the wound region and the measurement ruler/tape to create a ground truth mask. Figure 7 shows some examples of the images and their respective masks. The red region on the ground truth mask is the wound area, and the white region is the measurement ruler/tape placement in the original image. We also considered the combination of both datasets (`ULCER-BOTH`). To evaluate ASURA, we used 5-fold cross-validation, and the images were randomly split between the folds such that images from one patient are present in only one fold.

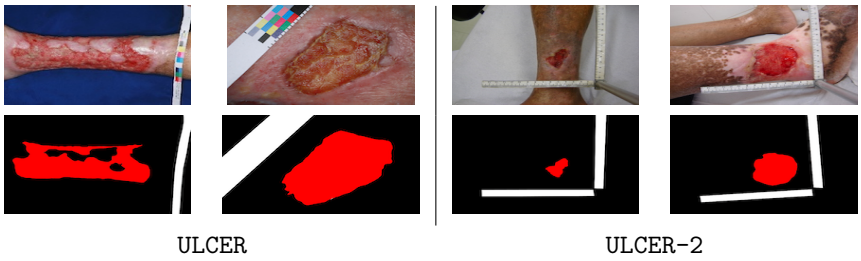


Figure 7: Example of skin ulcer images. The dataset images are on the top row and the ground truth masks are on the bottom row.

6.2. Data Augmentation

Deep learning models require a large amount of data to correctly learn patterns [7]. Since both `ULCER` and `ULCER-2` are small datasets, we used data augmentation to increase the robustness of ASURA. The images and masks were augmented using a series of random geometric transformations (translation, scale, and rotation). Each image was translated by a random value up to 10% of the width/height of the image. Each image was rotated by a random angle between $-15°$ and $15°$. Finally, each image was scaled up/down by a random value between 0.8 and 1.2. Points outside the image that are

now visible were filled with a background color (black). Figure 8 shows examples of these transformations. Each mask received the same geometric transformations of its corresponding image. Table 2 shows the number of original images and augmentation images for each dataset.

Figure 8: Examples of images generated by data augmentation.

Table 2: Number of images on each dataset.

Dataset	Size	Size After Augmentation
ULCER	217	1519
ULCER-2	229	1603
ULCER-BOTH	446	1784

6.3. Pixel Density Estimation

In this section, we detail the measures used to evaluate how well ASURA can estimate the area of the wound on a real-world unit of measurement, *e.g.*, squared centimeters (cm^2). The area A of a wound can be computed using the Pixel Density (λ). Thus, the area can be obtained using $A = |Mask|/\lambda^2$, where $Mask$ is the segmentation mask of the wound. On this experiment we estimated the Pixel Density in pixels per centimeter (*pixel* / cm). For evaluation purposes, an expert drew a one-centimeter line in each image with the ASURA GUI. By measuring the length in pixels of this line, it is possible to obtain the real Pixel Density (λ_{real}) of each image. The one-centimeter line is marked by a red line in the ASURA GUI.

Since ASURA can estimate more than one Pixel Density (λ) per measurement ruler/tape (distance between small ticks or distance between large

ticks), the best estimation is chosen (λ_{est}). It is important to note that sometimes the chosen λ_{chosen} can be equivalent to a fraction of the desired unit. In this case, λ_{chosen} must be multiplied by the corresponding factor, *e.g.*, if λ_{chosen} is equivalent to one millimeter, the estimated Pixel Density for one centimeter is $\lambda_{est} = 10 \times \lambda_{chosen}$. Using Equation 1 and the values of λ_{real} and λ_{est}, we can calculate four different areas in real-world unit of measurement:

- **Ground Truth Area (A_{gt})**: size of the wound on the ground truth mask and λ_{real};
- **Real Area (A_{real})**: size of the wound on the ground truth mask λ_{est};
- **Estimated Area (A_{est})**: size of the wound on the ASURA segmentation mask and λ_{est};
- **MBR Area (A_{mbr})**: size of the MBR (Section 2) on the ground truth mask and λ_{real}.

To evaluate the results obtained by ASURA, we calculated the relative error E in percentage, which can be calculated using Equation 2.

$$E_v = \frac{|v - \hat{v}|}{v} \times 100\% \tag{2}$$

where v can be any variable (λ_{est}, A_{real} and A_{est}), v is the true value of the variable and $\hat{v}$ is the estimated value of the variable.

7. Results and Discussion

In this section, we show the performance of ASURA to segment ulcer images and estimate the area of the wounds.

7.1. Ulcer Segmentation Evaluation

The first set of experiments evaluated how well ASURA segmented the skin ulcer images. We compared ASURA with DeepLabv3+ [26], QTDU [15], CL-Measure [17], ICARUS [19] and a Pixel Color Classification Segmentation (Color Classification). During the training of the DL based methods, we used the Dice Score as the objective function and the same amount of epochs for ASURA and DeepLabv3+. For the DeepLabv3+ method, we used the Xception as the backbone. For the QTDU, we used the ResNet to classify the

superpixels. For both CL-Measure and ICARUS we considered the superpixel classification step as a segmentation algorithm. For both CL-Measure and QTDU, we considered the different healing stages of the wound (fibrin, granulation, callous, and necrosis) as the same label (wounded tissue). Although Dorileo *et al.* [4] and Pereyra *et al.* [1] proposed ulcer segmentation algorithms, they were not directly compared with ASURA for the sake of fairness. Dorileo *et al.* was designed for a controlled environment, and every image needs a blue background. Since we are also considering images outside this scope, the results obtained by Dorileo *et al.* would be severely harmed. Pereyra *et al.* was not considered because it requires the manual selection of the correct clusters. For this step, we focused on automatic segmentation methods. To evaluate the overall effectiveness of all methods, we run each segmentation method in every image of each test dataset. Then, we calculated five measures: Jaccard Coefficient, Dice Score/F1-Score, Precision, and Recall. All results shown in this section are the average of every image. Table 3 shows the results obtained by all methods on the `ULCER`, `ULCER-2` and `ULCER-BOTH` datasets. Our experiments showed that ASURA outperformed the competitors on all datasets.

For each dataset, ASURA achieved values above 83% for the Jaccard Coefficient and above 89% for the other measurements. On the `ULCER`, ASURA was 19% better than QTDU and DeepLabv3+. While comparing with non DL based methods, ASURA was 37% better than ICARUS, 39% better than CL-Measure and 62% better than Color Classification on the Jaccard Coefficient. When comparing the Dice Score and F1-Score, ASURA was 12% better than DeepLabv3+, 13% better than QTDU, 27% better than the ICARUS, 29% better than CL-Measure and 52% better than Color Classification. According to the authors, for this dataset, Pereyra *et al.* [1] achieved a Jaccard Coefficient of only 56%. When comparing Precision, ASURA was 14% and 15% better than QTDU and DeepLabv3+ respectively. Comparing Recall, ASURA was 6% and 9% better than DeepLabv3+ and QTDU respectively.

A similar behavior occurred with `ULCER-2` dataset. ASURA was 28% better than QTDU, 44% better than DeepLabv3+, 48% better than ICARUS, 64% better than CL-Measure and 57% better than Color Classification on the Jaccard Coefficient. When comparing the Dice Score and F1-Score, ASURA was 20% better than QTDU, 34% better than DeepLabv3+, 41% better than ICARUS, 58% better than CL-Measure and 48% better than Color Classification. When comparing Precision, ASURA and DeepLabv3+ achieved a similar result. However, ASURA was able to achieve a Recall 45% better

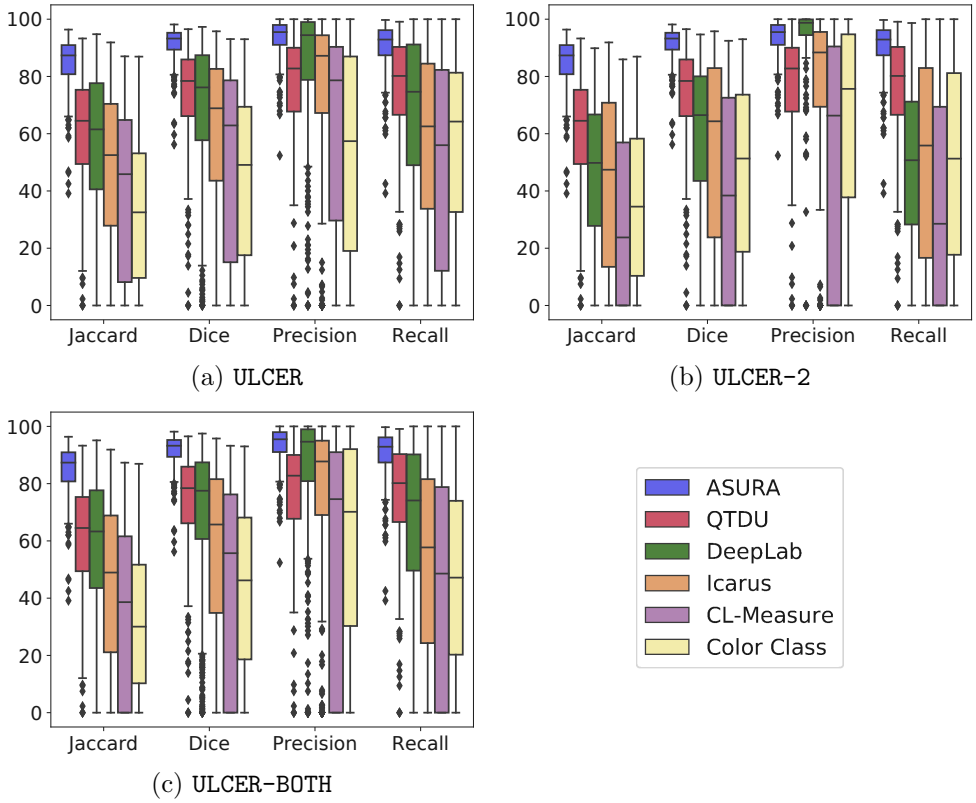
Table 3: Evaluation of the segmentation methods on each dataset. The highlighted line is our proposal and the **bold** values are the best results. All values are in percentage.

	Method	Jaccard	Dice / F1-Score	Precision	Recall
ULCER	ASURA	**83 ± 19**	**89 ± 16**	**91 ± 14**	**90 ± 17**
	QTDU	67 ± 23	77 ± 21	78 ± 22	81 ± 19
	DeepLab	68 ± 20	78 ± 18	78 ± 21	84 ± 16
	CL-Meas.	50 ± 24	63 ± 25	72 ± 26	64 ± 29
	ICARUS	52 ± 23	65 ± 23	76 ± 25	63 ± 26
	Col. Class.	31 ± 22	43 ± 26	42 ± 30	64 ± 23
ULCER-2	ASURA	**84 ± 10**	**91 ± 7**	**93 ± 7**	**90 ± 10**
	QTDU	60 ± 20	73 ± 19	77 ± 20	75 ± 21
	DeepLab	47 ± 24	60 ± 25	92 ± 17	49 ± 26
	CL-Meas.	30 ± 29	38 ± 35	50 ± 42	38 ± 36
	ICARUS	44 ± 30	54 ± 32	74 ± 33	51 ± 34
	Col. Class.	36 ± 26	47 ± 30	63 ± 35	50 ± 33
BOTH	ASURA	**84 ± 16**	**90 ± 13**	**92 ± 12**	**91 ± 13**
	QTDU	64 ± 21	76 ± 20	79 ± 21	77 ± 20
	DeepLab	64 ± 22	75 ± 21	83 ± 19	76 ± 25
	CL-Meas.	39 ± 28	50 ± 32	61 ± 36	50 ± 35
	ICARUS	46 ± 26	58 ± 28	75 ± 30	55 ± 31
	Col. Class.	31 ± 23	43 ± 27	58 ± 35	45 ± 29

than DeepLabv3+.

While processing the combined dataset (ULCER-BOTH), ASURA was still able to correctly segment the wound regions in the skin ulcer images. ASURA was 23% better than QTDU and DeepLabv3+, 45% better than ICARUS, 54% better than CL-Measure and 63% better than Color Classification on the Jaccard Coefficient. When comparing the Dice Score and F1-Score, ASURA was 16% better than QTDU and DeepLabv3+, 36% better than ICARUS (the second best method), 44% better than CL-Measure and 52% better than Color Classification. When comparing Precision, ASURA was 9% better

than DeepLabv3+ and 14% better than QTDU. And when comparing Recall, ASURA was 15% better than QTDU and 17% better than DeepLabv3+.



(a) `ULCER` (b) `ULCER-2`

(c) `ULCER-BOTH`

Figure 9: Box plot for the evaluation of the segmentation methods on each dataset, the Y axes are in percentage. The black diamonds represent outlier values from the box plot visualization.

Figure 9 shows a box plot summarizing the distribution of the results. The outliers are the values outside of the interquartile range times 1.5, as is the usual approach. It is important to notice that, for every measure, ASURA's box plots are shorter and with higher values. On the Jaccard Coefficient, the majority of ASURA values were above 80%, while the second-best result (ICARUS) had the majority of the values below 75%. On the Dice/F1-Score, the majority of ASURA values were above 90%, while on ICARUS, the majority was below 85%.

Figure 10 gives examples of the segmentation output from the `ULCER` and `ULCER-2` datasets. On both datasets, ASURA had an output similar to the

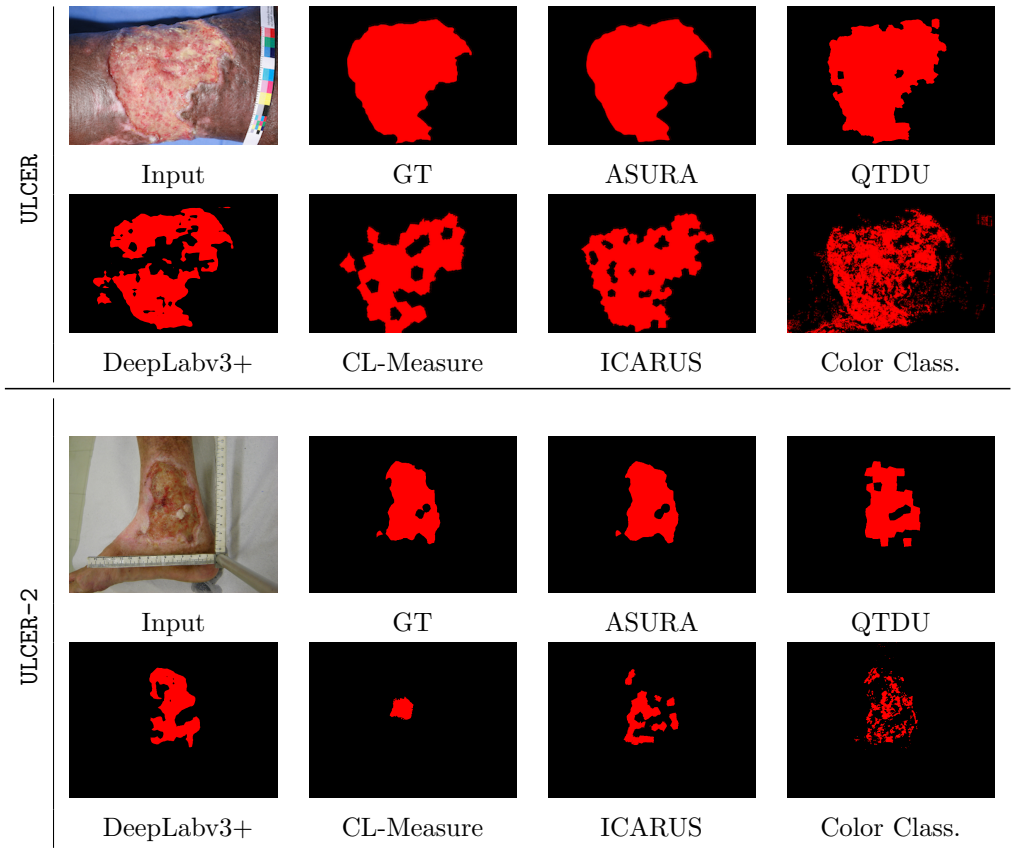


Figure 10: Ulcer segmentation of images from ULCER and ULCER-2.

Ground Truth (GT). For both images, although QTDU and DeepLabv3+ had good results, they still missed segmenting some regions of the wound. We can note that all methods based on superpixels (QTDU, CL-Measure and ICARUS) had problems to segment the wound due to the miss-classification of their superpixels. Nonetheless, it is possible to note that the use of CNNs made a huge improvement in the superpixels classification. DeepLabv3+ faced problems to correctly segment the wound due to the small amount of training data, which favored the U-Net based architecture used by ASURA. Since Color Classification is pixel-wise, it is unable to segment the whole wound correctly. Thus some pixels inside the region were not considered as part of the wound.

Figure 11 exemplifies bad segmentation outputs generated by ASURA for ULCER and ULCER-2 datasets. On the top row of Figure 11, ASURA

considered parts of the wound as healthy/background. A reason for this mistake was the presence of a shine region on that part of the wound. ASURA also had a poor performance while segmenting images with small wounds (bottom row of Figure 11). For this image, ASURA achieved a Jaccard Coefficient of 25.85% and a Dice Score of 41.08%. A reason for this poor performance on images with small wounds is because ASURA has to resize the images to the input tensor size (512×512) and later resize the output tensor to the original size. The wound on this image has 147 pixels, and the ASURA output has only 38 pixels, which is not sufficient for the superpixel segmentation method to work properly.

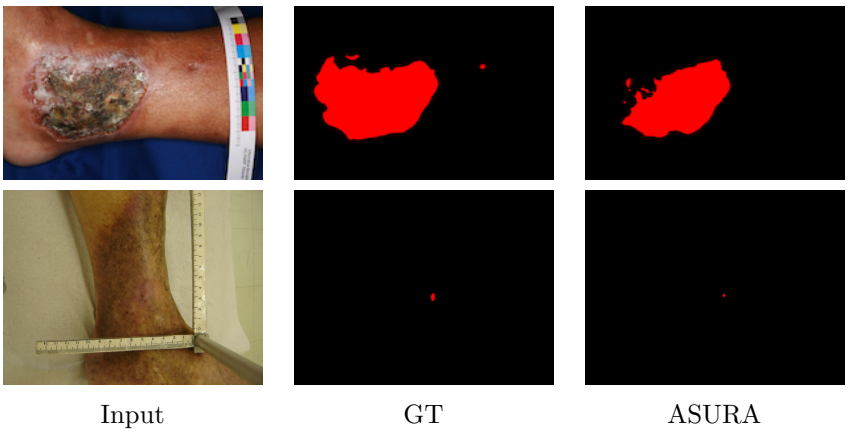


Figure 11: Cases where ASURA had a poor segmentation performance.

7.2. Pixel Density Estimation Evaluation

In this experiment, we evaluated how well ASURA can estimate the area of the wound on a real-world unit of measurement, *e.g.*, squared centimeters (cm^2). We calculated the relative error for every image in the test set. The results shown in this section are the average over every image.

Table 4 shows the relative error of the estimated Pixel Density (λ_{est}), real area (A_{real}), estimated area (A_{est}) and the MBR area (A_{mbr}). By calculating the relative error of A_{real}, we can estimate the impact of the ASURA's Pixel Density estimation, the relative error of A_{est} shows how well ASURA can estimate the wound area in cm^2 and the relative error of A_{mbr} shows how the simplest manual estimation is less reliable than ASURA. ASURA had the best result on `ULCER-2`, estimating λ_{est} with a relative error of 4.7%, the

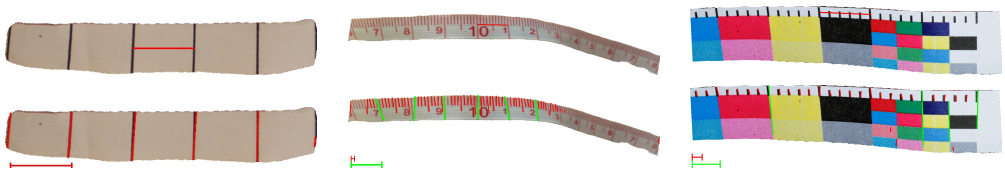
error while calculating the area A_{real} was of 11.2% and the area A_{est} had an error of 17.2%. On ULCER, ASURA was able to estimate λ_{est} with a relative error of 8%. Using the λ_{est} to calculate the A_{real}, ASURA had a relative error of 22.2% and the estimated area A_{est} had an error of 33.9%. ULCER-BOTH had errors between ULCER-2 and ULCER. Moreover, ASURA was up to 31% better than the area estimation using the ruler method (A_{mbr}).

Table 4: Relative error (%) of the Pixel Density and area estimation.

Dataset	Relative Error (%)			
	λ_{est}	A_{real}	A_{est}	A_{mbr}
ULCER	8.0	22.2	33.9	52.7
ULCER-2	4.7	11.2	17.2	48.1
ULCER-BOTH	6.3	16.6	30.5	50.4

Figure 12 shows some examples of Pixel Density estimations. Figure 12(a) shows a measurement tape where ASURA correctly estimated a λ equivalent to 1.0 cm. The estimated Pixel Density is $\lambda_{est} = 116$ pixels/cm and the ground truth is $\lambda_{real} = 115$ pixels/cm. Figure 12(b) shows a measurement tape where ASURA estimated two different λ, one for the smaller ticks (λ_{red}) and one for the larger ticks (λ_{green}). On this measurement tape, the smaller ticks are the millimeters (mm), and the green ticks are the centimeters. For this measurement tape $\lambda_{real} = 100$ pixels/cm, the estimated $\lambda_{red} = 11$ pixels/mm and $\lambda_{green} = 101$ pixels/cm. Figure 12(c) shows a measurement tape with more details (colored squares). Even with a more complex object, ASURA was able to correctly estimate the Pixel Density. Once again, ASURA estimated two variations of the Pixel Density, $\lambda_{red} = 47$ pixels/(2×mm) and $\lambda_{green} = 231$ pixels/cm, the ground truth is $\lambda_{real} = 240$ pixels/cm.

However, ASURA was not able to correctly estimate the Pixel Density on some images. While trying to detect the ticks on some images, ASURA wrongly detected the vertical lines on the text as ticks. Other images also have a bright spot on regions of the measurement tape, leading ASURA to fail in detecting the ticks on those regions. Measurements of the Pixel Density performed by ASURA with a high relative error can be manually fixed by the user (specialist). Table 5 shows the relative errors with the aid of an expert. By replacing the wrong estimations, we had a reduction of up to 2.3% on the relative error of the Pixel Density on all datasets.



(a) (b) (c)

Figure 12: Pixel Density estimation on different measurement tapes. The red horizontal lines on the top row are the expected length corresponding to 1cm. The red and green lines on the bottom row are the ASURA's estimated distance between the ticks with the same colors, as described in Section 5.2.

We reduced the relative error on the area A_{real} by 9.4%, 3% and 6.2% on the `ULCER`, `ULCER-2` and `ULCER-BOTH` respectively. The relative error on the area A_{est} is reduced by 9%, 2.3% and 6% on the `ULCER`, `ULCER-2` and `ULCER-BOTH` respectively. Although the user must input a new value of Pixel Density on some estimations, the majority of the images have less than 20% of relative error. All datasets have more than 90% of estimations with a relative error lower than 20%, the `ULCER` has 92.4%, the `ULCER-2` has 96.3% and the `ULCER-BOTH` has 94.4%. This shows that by replacing only a few estimations, we can improve the area estimation by up to 9%. When comparing with the error of area A_{mbr}, ASURA was 27%, 33% and 26% better than the manual estimation on the `ULCER`, `ULCER-2` and `ULCER-BOTH` respectively.

Table 5: The relative error, in percentage, of the Pixel Density and area estimation with the aid of a human expert.

Dataset	Relative Error (%)		
	λ_{est}	A_{real}	A_{est}
`ULCER`	5.7	12.7	25.0
`ULCER-2`	3.7	8.2	14.8
`ULCER-BOTH`	4.7	10.4	24.5

8. Conclusions

In this paper, we proposed the ASURA framework to segment and estimate the wound area of a skin ulcer image. ASURA uses an encoder/decoder deep neural network to segment the wound from the image. The comparison

of ASURA with methods built on traditional image processing algorithms, such as CL-Measure and ICARUS, has revealed that the deep learning approaches presented way better results. ASURA also detects the measurement ruler/tape present in the image and automatically estimates the image's pixel density. This allows an accurate size measurement of the wounds. Accordingly, the main contributions of this paper are:

- **Segmentation**: ASURA was able to correctly segment the wound regions with a Dice score greater than 90%, being up to 16% better than the state-of-the-art methods (QTDU and DeepLabv3+) in both datasets (`ULCER-BOTH`); and
- **Area measurement**: ASURA was able to automatically estimate the pixel density of the images with a relative error of 5.6% and semi-automatically able to estimate the wound area in cm^2 within a relative error of 12.1%. To the best of our knowledge, this is a novel resource, highly desired by the physicians and healthcare specialists.

Ethical Approval

The ulcer images used in this study were obtained with the permission of a researcher who has undertaken a clinical trial approved by the ethical research committee of Ribeirao Preto Medical School, from the University of Sao Paulo.

Conflict of interests

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

Acknowledgements

This research was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, by grants #2014/25125-3, #2016/17330-1, #2016/17078-0, #2018/24414-2 from the São Paulo Research Foundation (FAPESP) and CNPq. The authors would like to thank the NVIDIA Corporation for the donation of a Titan Xp card. The authors also would like to thank Marco A. C. Frade, Ederson A. G. Dorileo, and co-authors for providing the `ULCER` dataset.

References

[1] L. C. Pereyra, S. M. Pereira, J. P. Souza, M. A. C. Frade, R. M. Rangayyan, P. M. Azevedo-Marques, Characterization and pattern recognition of color images of dermatological ulcers - a pilot study, Comp. Science Journal of Moldova 22 (2) (2014) 211–235.
URL http://www.math.md/publications/csjm/issues/v22-n2/11672/

[2] M. T. Cazzolato, L. S. Rodrigues, L. C. Scabora, G. F. Zaboti, G. Q. Vasconcelos, D. Y. Chino, A. E. S. Jorge, R. L. F. Cordeiro, C. Traina-Jr, A. J. M. Traina, A DBMS-based framework for content-based retrieval and analysis of skin ulcer images in medical practice, in: Proceedings of the 34th Brazilian Symposium on Databases, SBC, 2019, pp. 109–120.
URL https://sol.sbc.org.br/index.php/sbbd/article/view/8812

[3] E. A. G. Dorileo, M. A. C. Frade, A. M. F. Roselino, R. M. Rangayyan, P. M. Azevedo-Marques, Color image processing and content-based image retrieval techniques for the analysis of dermatological lesions, in: IEEE Eng. Med. Biol. Society., 2008, pp. 1230–1233. doi:10.1109/IEMBS.2008.4649385.

[4] E. A. G. Dorileo, M. A. C. Frade, R. M. Rangayyan, P. M. Azevedo-Marques, Segmentation and analysis of the tissue composition of dermatological ulcers, in: CCECE, 2010, pp. 1–4. doi:10.1109/CCECE.2010.5575143.

[5] P. Gholami, M. A. Ahmadi-pajouh, N. Abolftahi, G. Hamarneh, M. Kayvanrad, Segmentation and measurement of chronic wounds for bioprinting, IEEE Journal of Biomedical and Health Informatics 22 (4) (2018) 1269–1277. doi:10.1109/JBHI.2017.2743526.

[6] L. M. Morton, T. J. Phillips, Wound healing and treating wounds: Differential diagnosis and evaluation of chronic wounds, Journal of the American Academy of Dermatology 74 (4) (2016) 589 – 605. doi:https://doi.org/10.1016/j.jaad.2015.08.068.

[7] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12, Curran Associates Inc., USA, 2012, pp. 1097–1105.
URL http://dl.acm.org/citation.cfm?id=2999134.2999257

[8] Y. Yuan, M. Chao, Y. Lo, Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance, IEEE Trans. on Medical Imaging 36 (9) (2017) 1876–1886. doi:10.1109/TMI.2017.2695227.

[9] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, T.-S. Kim, Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks, Computer Methods and Programs in Biomedicine 162 (2018) 221 – 231. doi:https://doi.org/10.1016/j.cmpb.2018.05.027.

[10] P. Tang, Q. Liang, X. Yan, S. Xiang, W. Sun, D. Zhang, G. Coppola, Efficient skin lesion segmentation using separable-unet with stochastic weight averaging, Computer Methods and Programs in Biomedicine 178 (2019) 289 – 301. doi:https://doi.org/10.1016/j.cmpb.2019.07.005.
URL http://www.sciencedirect.com/science/article/pii/S0169260719306145

[11] J. Kawahara, G. Hamarneh, Multi-resolution-tract cnn with hybrid pretrained and skin-lesion trained layers, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2016, pp. 164–171.

[12] L. Yu, H. Chen, Q. Dou, J. Qin, P. Heng, Automated melanoma recognition in dermoscopy images via very deep residual networks, IEEE Trans. on Medical Imaging 36 (4) (2017) 994–1004. doi:10.1109/TMI.2016.2642839.

[13] M. Goyal, M. H. Yap, N. D. Reeves, S. Rajbhandari, J. Spragg, Fully convolutional networks for diabetic foot ulcer segmentation, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, 2017, pp. 618–623.

[14] S. Zahia, D. Sierra-Sosa, B. Garcia-Zapirain, A. Elmaghraby, Tissue classification and segmentation of pressure injuries using convolutional

neural networks, Computer Methods and Programs in Biomedicine 159 (2018) 51 – 58. doi:https://doi.org/10.1016/j.cmpb.2018.02.018.

[15] G. Blanco, A. J. Traina, C. T. Jr., P. M. Azevedo-Marques, A. E. Jorge, D. de Oliveira, M. V. Bedo, A superpixel-driven deep learning approach for the analysis of dermatological wounds, Computer Methods and Programs in Biomedicine 183 (2020) 1–9. doi:10.1016/j.cmpb.2019.105079.

[16] F. Navarro, M. Escudero-Viñolo, J. Bescós, Accurate segmentation and registration of skin lesion images to evaluate lesion change, IEEE Journal of Biomedical and Health Informatics 23 (2) (2019) 501–508. doi:10.1109/JBHI.2018.2825251.

[17] G. Blanco, M. V. N. Bedo, M. T. Cazzolato, L. F. D. Santos, A. E. S. Jorge, C. Traina, P. M. Azevedo-Marques, A. J. M. Traina, A label-scaled similarity measure for content-based image retrieval, in: 2016 IEEE International Symposium on Multimedia (ISM), 2016, pp. 20–25. doi:10.1109/ISM.2016.0014.

[18] J. L. Seixas, S. Barbon, R. G. Mantovani, Pattern recognition of lower member skin ulcers in medical images with machine learning algorithms, in: IEEE International Symposium on Computer-Based Medical Systems, 2015, pp. 50–53. doi:10.1109/CBMS.2015.48.

[19] D. Y. T. Chino, L. C. Scabora, M. T. Cazzolato, A. E. S. Jorge, C. Traina, A. J. M. Traina, Icarus: Retrieving skin ulcer images through bag-of-signatures, in: IEEE International Symposium on Computer-Based Medical Systems, 2018, pp. 82–87. doi:10.1109/CBMS.2018.00022.

[20] C. Little, J. McDonald, M. Jenkins, P. McCarron, An overview of techniques used to measure wound area and volume, Journal of wound care 18 (6) (2009) 250–253.

[21] W. Rasband, ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA (1997–2018).
URL `https://imagej.nih.gov/ij/`

[22] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 4278 – 4284.

[23] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[24] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (12) (2017) 2481–2495. doi:10.1109/TPAMI.2016.2644615.

[25] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[26] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision – ECCV 2018, Springer International Publishing, Cham, 2018, pp. 833–851.

[27] D. Cong, Q. Zhou, J. Cheng, X. Wu, S. Zhang, W. Ou, H. Lu, CAN: contextual aggregating network for semantic segmentation, in: International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 1892–1896.

[28] Q. Zhou, W. Yang, G. Gao, W. Ou, H. Lu, J. Chen, L. J. Latecki, Multi-scale deep context convolutional neural networks for semantic segmentation, World Wide Web 22 (2) (2019) 555–570. doi:10.1007/s11280-018-0556-3.

[29] W. S. Silva, D. L. Jasbick, R. E. Wilson, P. M. Azevedo-Marques, A. J. M. Traina, L. F. D. Santos, A. E. S. Jorge, D. de Oliveira, M. V. N. Bedo, A two-phase learning approach for the segmentation of dermatological wounds, in: 32nd IEEE International Symposium on Computer-Based Medical Systems, CBMS 2019, Cordoba, Spain, June 5-7, 2019, 2019, pp. 343–348. doi:10.1109/CBMS.2019.00076.

[30] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, G. Randall, Lsd: a line segment detector, Image Processing Online 2 (2012) 35–55.