



Article Convolution Neural Network with Coordinate Attention for Real-Time Wound Segmentation and Automatic Wound Assessment

Yi Sun^{1,2}, Wenzhong Lou^{1,2,*}, Wenlong Ma¹, Fei Zhao¹ and Zilong Su¹

- ¹ National Key Laboratory of Electro-Mechanics Engineering and Control, School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100010, China
- ² Beijing Institute of Technology Chongqing Innovation Center, Chongqing 401120, China
- * Correspondence: louwz@bit.edu.cn; Tel.: +86-135-0113-7973

Abstract: Background: Wound treatment in emergency care requires the rapid assessment of wound size by medical staff. Limited medical resources and the empirical assessment of wounds can delay the treatment of patients, and manual contact measurement methods are often inaccurate and susceptible to wound infection. This study aimed to prepare an Automatic Wound Segmentation Assessment (AWSA) framework for real-time wound segmentation and automatic wound region estimation. Methods: This method comprised a short-term dense concatenate classification network (STDC-Net) as the backbone, realizing a segmentation accuracy-prediction speed trade-off. A coordinated attention mechanism was introduced to further improve the network segmentation performance. A functional relationship model between prior graphics pixels and shooting heights was constructed to achieve wound area measurement. Finally, extensive experiments on two types of wound datasets were conducted. Results: The experimental results showed that real-time AWSA outperformed state-of-the-art methods such as mAP, mIoU, recall, and dice score. The AUC value, which reflected the comprehensive segmentation ability, also reached the highest level of about 99.5%. The FPS values of our proposed segmentation method in the two datasets were 100.08 and 102.11, respectively, which were about 42% higher than those of the second-ranked method, reflecting better real-time performance. Moreover, real-time AWSA could automatically estimate the wound area in square centimeters with a relative error of only about 3.1%. Conclusion: The real-time AWSA method used the STDC-Net classification network as its backbone and improved the network processing speed while accurately segmenting the wound, realizing a segmentation accuracy-prediction speed trade-off.

Keywords: convolutional neural networks; coordinate attention; wound segmentation; area assessment

1. Introduction

Wound treatment in modern emergency environments, such as battlefields, fire disasters, and earthquakes, has distinct characteristics: more wounded but fewer medical personnel, fewer medical resources, and difficult medical evacuation. These factors make large-scale wound care extremely difficult in emergency situations [1]. Meanwhile, body surface wounds, such as scratches and burns, can lead to infection and poor blood circulation if not treated in time or amputation in severe cases [2–4]. Generally, the area and depth of the wound are manually estimated by medical staff [5]. Another method is to use a regular video camera to photograph the wound with a reference scale (such as a tape or a ruler) and upload it. Medical experts can judge the boundary and area of the wound and decide on its treatment. However, the manual segmentation of the wound area is complex and time-consuming, which delays the treatment of many wounded patients, and the contact wound measurement method is prone to wound infection. Therefore, a real-time and accurate tool is needed to assist in emergency medical care. The sequelae



Citation: Sun, Y.; Lou, W.; Ma, W.; Zhao, F.; Su, Z. Convolution Neural Network with Coordinate Attention for Real-Time Wound Segmentation and Automatic Wound Assessment. *Healthcare* **2023**, *11*, 1205. https:// doi.org/10.3390/healthcare11091205

Academic Editor: Oriana Simonetti

Received: 19 February 2023 Revised: 3 April 2023 Accepted: 12 April 2023 Published: 23 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). caused by the wound can be minimized by judging the state of the wound, transmitting wound information in time, and taking targeted treatment measures.

As technological advances in smartphones, computing storage devices, and clinical devices have improved the quality of image information [6,7], the computer-aided automatic segmentation and measurement of wound size have become new methods for accurate wound assessment. In particular, artificial intelligence technology has proved its efficiency and high performance in automatic image classification via machine learning methods. Artificial intelligence effectively removes a large amount of redundant information, analyzes and judges the state of the wound according to the wound image data, and assists telemedicine experts in preparing the best treatment plan so as not to miss the "golden 30 min" of emergency treatment. Deep learning (DL) is an extension of machine learning that mainly focuses on the automatic extraction and classification of image features and has achieved great success in many applications, especially in healthcare [8,9]. The introduction of DL techniques has motivated several researchers to use convolutional neural networks (CNNs) in the medical domain [10]. CNNs are a powerful tool for image processing owing to their good feature representation capability [11–14]. Photographic images have been used to recognize melanomas by segmenting [15–17] or classifying them; they have also been used for foot ulcer segmentation [18-20] and pressure ulcer segmentation and classification [21]. However, few studies have dealt with wound segmentation using DL techniques. Such studies do not provide the real-time segmentation of wound images and non-contact wound measurement without the help of gauges, but the proposed real-time Automatic Wound Segmentation Assessment (AWSA) framework addresses these concerns.

In this study, we performed the real-time automatic assessment of body surface wounds, which could help to rapidly assess wounds in many patients under emergency medical care and provide targeted treatment during the first 30 min of admission. This task mainly involved two steps: automatic wound segmentation and the assessment of the wound area. Wound image segmentation was conducted to locate the boundary between the wound and the surrounding skin [22]. The measurement of the wound area is usually performed manually, which is time-consuming and inaccurate and causes discomfort to patients [23–25]. Accurate and automatic wound measurement mostly relies on well-segmented wound regions. Previous studies lacked the required accurate segmentation, as they focused more on the retrieval and classification tasks.

We prepared the real-time AWSA framework to address the shortcomings of the presently used methods. Real-time AWSA uses CNNs and automatically detected and segmented wound areas delineated in images. It automatically calculates the area of the unknown wound by building a functional model between the prior graphics pixels and the graphics shooting heights. This method could help medical experts quickly obtain patients' wound information without touching the wound or using measuring tools (such as measurement rulers or tapes). We used a novel and efficient short-term dense concatenate (STDC) network structure that removed structural redundancy to address these problems. Specifically, the basic module of the STDC network was formed by gradually reducing the dimensionality of feature maps and using their aggregation for image representation. In the decoder, we proposed a detail aggregation module by integrating the learning of spatial information into low-level layers in a single-stream manner. The low-level and deep features were fused to predict the final segmentation results. Finally, STDC-Net was used as the backbone to achieve a state-of-the-art speed–accuracy trade-off in real-time semantic segmentation by adding a coordinated attention mechanism.

The contributions of real-time AWSA are two-fold:

- The improved STDC-Net architecture with the pretrained weight model as the encoder layer achieved a trade-off between wound segmentation accuracy and prediction speed.
- 2. The coordinated attention mechanism was proposed to better obtain the global receptive field and encode the accurate location information so that the network could

locate the target of interest more accurately and further improve the performance of the proposed network.

3. The wound area estimation without contact and without measurement tools was realized by constructing a functional relationship model between prior graphic pixels and image shooting heights.

This study involved an extensive experimental analysis comparing the proposed method with the currently used state-of-the-art segmentation methods, demonstrating the highly accurate real-time performance of real-time AWSA. Regarding wound area measurement, we compared the results obtained using the real-time AWSA method with manual segmentations based on methods used in previous studies, revealing the high-level accuracy of real-time AWSA.

Paper outline: The challenges facing surface wound area estimation and the basic concepts used in this study are discussed in Section 1. Section 2 discusses related studies. Section 3 details the real-time AWSA framework. Section 4 presents the materials and methods used in this study. Section 5 presents the experimental results and the corresponding discussion. Section 6 summarizes this study.

2. Related Works

To date, several studies have been conducted on wound segmentation and wound area estimation. For example, Chino et al. proposed Automatic Skin Ulcer Region Assessment (ASURA), a segmentation method for ulcer wounds based on U-Net [26]. However, some defects were found. First, the simple skip-connection between the encoder and the decoder did not account for the importance of different channels, which may have reduced the accuracy. Second, as the network became deeper and wider, the redundancy of the network structure made the segmentation task more time-consuming and difficult to optimize.Moreover, the ASURA system proposed by Chino et al. was used for the segmentation and measurement of wound area in real-world units (cm²) [27].

The SegNet decodes the feature map by upscaling and using a series of convolutions [28,29]. However, these networks require thousands of annotated training samples. Ronneberger et al. made a breakthrough in medical image segmentation using U-Net based on FCN to overcome this problem [30]. In U-Net, the decoder receives a copy of the output of the activation layers and concatenates it with the upscaling tensor. In this manner, U-Net can pass on the spatial information lost in the encoder step to the corresponding decoder layers, improving the segmentation output.

Many U-Net-based variants have emerged in recent years [31,32]. ASURA used U-Net as the network backbone to perform image segmentation and achieved decent accuracy. Meanwhile, ASURA automatically measured the wound size and adjusted the measurements manually through the app. However, this method was time-consuming for segmentation tasks due to redundant network structures and did not provide real-time performance when a large number of wound images needed to be processed in emergency medical care. Furthermore, measurement tools (measurement rulers/tapes) were required to estimate the wound area in the wound measurement task.

Dorileo et al. proposed an image segmentation method based on the analysis of the RGB channels of the image [33]. As all images had a blue background, they discarded the blue channel and used the intensity channel of the hue, saturation, and intensity (HSI) color space. For each channel, Dorileo et al.'s method helped automatically find thresholds and process the discovered regions by focusing on blobs near the center of the image. The main issue with this method was the need for a controlled environment.

Blanco et al. proposed QTDU, a deep-learning-based approach to analyze dermatological wounds using superpixels [34]. QTDU used CNNs for wound segmentation and relied on superpixel approaches to divide images into regions. However, QTDU did not segment the rules/tapes present in the images, and the estimation of the wound area involved counting the number of pixels inside the segmented area of each identified tissue and checking this value proportionally with the number of pixels of the entire image. Seixas et al. employed off-the-shelf classifiers to segment wound images. They extracted pixel-wise color features, the mean value of the neighborhood of the pixel, and the difference in the pixel value and the mean beforehand. They segmented a training set of images to isolate the wound region.

Pereyra et al. proposed a segmentation process based on a multivariate Gaussian mixture model. The clusters were manually selected in a graphical user interface (GUI) to output the segmentation mask. Blanco et al. proposed the Counting-Labels Similarity Measure (CL-Measure), which focused on retrieving skin wound images based on visual similarity. Chino et al. proposed Imaging Content Analysis for the Retrieval of Ulcer Signatures (ICARUS), which was based on superpixels combined with Bags of Visual Signatures. It focused on the content-based retrieval of ulcer images and presented higher-quality results than CL-Measure.

Dastjerdi et al. proposed another method for semi-automatic wound segmentation and area measurement. It used both 2D and 3D representations, processing a single photo or a video, respectively. The 2D photo could be taken using a digital camera or smartphone, with a flexible paper ruler placed near the wound for size measurement. The segmentation started by roughly outlining the region of interest around the ulcer. Then, a trained random forest model calculated a probability map of each pixel belonging to the wound or healthy skin. A binary mask containing the wound area was created over the probability map by employing Otsu's threshold. The ruler was segmented to calculate the ratio between pixels and centimeters. However, this segmentation method lacked real-time applications, and wound area estimation required the aid of a ruler.

CNNs are mainly used for image recognition tasks. They consist of a series of convolution operations that encode an image into a feature map and can be used to perform image segmentation. The fully convolutional network (FCN) is a pioneering work of CNN in image segmentation.

DeepLab is another DL model for image segmentation. It employs dilated atrous convolutions to upscale the low-level features and enlarges the field of view of filters using a simple architecture. DeepLabv3+ is its latest version, which has an effective decoder to refine the segmentation, replacing the maximum pooling operations with depth-wise separable convolutions. The DeepLabv3+ decoder concatenates the encoded features, which are upscaled by a factor of four, with the corresponding features.

A comparison of our method with current state-of-the-art wound segmentation methods is presented in Table 1. Almost all methods could segment the wound. Some methods used superpixels in semantic segmentation to reduce the complexity of the image, which might impact the effect of wound segmentation. Real-time AWSA uses a novel and efficient STDC-Net classification network as a backbone to achieve high-precision wound segmentation with a high FPS while adding a coordinated attention mechanism to achieve the optimal speed–accuracy trade-off. However, none of the aforementioned methods realized the estimation of wound area without external measurement tools. This study measured unknown and irregular wounds by constructing the relationship model framework between prior graphics pixels and shooting heights, which could not only measure the wound area in real-world units but also realize the estimation of wound area without contact and without measurement tools.

	Wound Segmentation	Real-Time Segmentation	Detect Measurement Tool	Area Assessment without Measuring Tools
Dorlieo	\checkmark			
Seixas	\checkmark			
Pereyra				
CL-Measure	\checkmark			
ICARUS	\checkmark			
Dastjerdi	\checkmark			
ASURA (U-Net)	\checkmark		\checkmark	
Real-Time AWSA	\checkmark	\checkmark	\checkmark	\checkmark

Table 1. Summary of different methods to segment wounds.

3. Real-Time AWSA

We prepared a real-time wound segmentation and wound area estimation framework, real-time AWSA, that automatically measures wound area in images. Real-time AWSA uses deep CNNs to segment wounds. The functional relationship model between prior graphics pixels and image shooting heights was constructed to automatically measure the segmented wound area, which not only realized the estimation of wound area in real-world units but also avoided the use of measurement tools. Real-time AWSA works based on the following two main steps: (1) an automatic segmentation of surface wounds and (2) the construction of a functional relationship model between prior graphics pixels and graphics shooting heights to automatically estimate the wound area. Figure 1 shows the real-time AWSA framework. Real-time AWSA also offers an interactive GUI in which the user can obtain the predicted segmentation mask for the wound and interact with the GUI to obtain the estimated wound area.





3.1. Real-Time Wound Segmentation

In the segmentation task, real-time AWSA received RGB images of the surface wound and output segmentation masks with the wounds. The wound segmentation process was based on a convolutional deep neural network developed for image segmentation. Due to the limited training dataset and considering the tradeoff between segmentation accuracy and speed, real-time AWSA used the STDC-Net architecture, which could address the issues of the possible trade-off between segmentation accuracy and speed [35].

Figure 2 shows the improved network architecture in this study. The network consisted of an encoder and a decoder. First, real-time AWSA received, as input, an RGB image with an arbitrary resolution. As the input layer of the network was a tensor of size $512 \times 512 \times 3$, the image was resized to a 512×512 resolution. The network architecture consisted of six stages, in addition to an input layer and prediction layer. Generally, stages 1–5 downsampled the spatial resolution of the input with a stride of two, and stage 6 output the prediction logits by one ConvX, one global average pooling layer, and two fully connected

layers. Each ConvX consisted of one convolutional layer, one batch normalization layer, and one ReLU activation layer. Stages 1 and 2 are usually regarded as low-level layers for appearance feature extraction. We used only one convolutional block each in stages 1 and 2, which proved to be effective. The number of STDC modules in stages 3 to 5 was carefully tuned in our network. The first STDC module in each of these three stages down-sampled the spatial resolution with a stride of two. The following STDC modules in each stage kept the spatial resolution unchanged. We used the attention refine module to refine the combination features of stages 3 to 5. We adopted the feature fusion module in BiSeNet for the final semantic segmentation prediction [28] to fuse the 1/8 down-sampled feature from stage 3 in the encoder and the counterpart from the decoder. We set the output channel number as 1024 and carefully tuned the channel number of the remaining stages until reaching a good trade-off between accuracy and efficiency. Further, a coordinated attention module was added before and after the feature fusion module, which further improved the network prediction ability. Finally, the output tensor was resized to the resolution of the input image.



Figure 2. Overview of the real-time AWSA network. ARM denotes attention refine module, and FFM denotes feature fusion module.

3.2. Wound Area Estimation

After wound segmentation, real-time AWSA estimated the wound area in real-world units by constructing a functional relationship model between prior graphic pixels and the shooting heights of the image. Figure 3 shows the steps of real-time AWSA for estimating the wound area S_w . Algorithm 1 shows how real-time AWSA estimated the wound area S_w .

Algorithm 1 Wound area estimation.

Initialization: S_{pi} = area of prior graphics Require: I: input image, Mask: segmentation mask **Output:** S_w = the area of the wound Begin 1. λ_i : obtained prior graphics pixels (**I**,**Mask**) 2. $\lambda_1, \lambda_2, \ldots, \lambda_n$: pixels of different shooting heights h_1, h_2, \ldots, h_n 3. $\lambda = f(h)$: polynomial fitting of shooting height and prior graphics pixels 4. **if** $(h_i < h_1)$ or $(h_i > h_n)$: return none 5. elif $h_1 < h_i < h_n$: 6. $S_w = \varphi$ 7. for h_i in range (h_1, h_n) do 8. λ_w = wound image pixels obtained 9. calculate the wound area $S_w = (S_{pi} \times \lambda_w)/\lambda_i$ 10. end for 11. end if 12. return S_w



Figure 3. Functional relationship model between prior graphics pixels and shooting heights to estimate the wound area in real-world units.

Prior graphics took regular shapes such as triangles or squares with known areas. A smartphone was used to carry a laser ranging sensor to photograph prior graphics whose area S_{pi} was known. The shooting height varied from the minimum height h_1 to the maximum height h_n , and the distance was equally divided, as shown in Figure 3. The images were taken using the same smartphone with the same resolution. The number of pixels occupied by the prior graphics in the image changed with the shooting height, and each shooting height h corresponded to a pixel number λ . Next, the discrete relationship between the image shooting heights $h_1, h_2, h_3, \ldots, h_n$ and the number of pixels occupied by the 2D prior graphics $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_i$ was constructed using a polynomial fitting method, i.e., $\lambda = f(h)$. From this function, we could determine the number of pixels λ_i of the image corresponding to h_i at any point in the range h_1 to h_n . Last, we took a wound image with an unknown area S_w at a height greater than h_1 and lower than h_n , and then the wound area S_w could be obtained as follows:

$$S_w = \frac{S_{pi}}{\lambda_i} \times \lambda_w \tag{1}$$

where λ_w is the number of pixels in the image occupied by the wound at the shooting height h_i .

3.3. Graphical User Interface (GUI)

The real-time AWSA framework contains an interactive GUI that allows users to view the original wound image, the ground truth of the wound, and the predicted MASK of the wound after image segmentation. The interactive interface also allows the user to obtain the segmented wound area in real time after obtaining the predicted mask of the wound. Figure 4 shows the GUI of the real-time AWSA framework. The input wound image is below the Original Image heading. The ground truth of the wound is under the Label Image heading, while the prediction mask after wound segmentation is under the Prediction Image heading. The function selection of the interactive interface is on the left side of the GUI. Users can select any wound image and load the model to complete the semantic segmentation of the wound region. Real-time AWSA enables automatic wound area estimation. The user enters the height at which the wound image was taken, and real-time AWSA measures the area of the segmented wound in real-world units.



Figure 4. Real-time AWSA interactive graphical user interface.

4. Materials and Methods

The performance of real-time AWSA was evaluated to verify its rapid wound area estimation. Two sets of experiments were conducted for performance evaluation: real-time wound segmentation and wound area estimation. All experiments were conducted using a 4.20-GHz Intel Core i7-12700F CPU, 32 GB RAM, and 12 GB NVIDIA GTX 3060Ti GPU. Further, we implemented real-time AWSA in Python based on the PyTorch1.8.1 framework, and the development software used was Pycharm 3.7.

4.1. Datasets and Pre-Processing

Real-time AWSA was evaluated on the self-built wound dataset WOUND. It consisted of 661 images of different wound types, including scratches, cuts, and bruises, mainly on the arms, legs, and upper body. Among them, 535 images were used as the training set, and 126 images were used as the test set. In the images, the wounds were located all over the body, and some patients had multiple wounds. The image dataset WOUND was obtained from the National Trauma Database and Chengde County Hospital. All images in the dataset were captured by digital cameras. For the dataset WOUND, the experts manually segmented the wound area to create the ground truth mask. Figure 5 shows several wound

images and their respective ground truth masks, where the gray area is the wound. We augmented the dataset and compared our method with currently used state-of-the-art wound image segmentation methods to evaluate the performance of real-time AWSA.



Figure 5. Example wound images. The dataset images are on the top row, and the ground truth masks are on the bottom row.

DL models require a large amount of data for training to improve [10]. As WOUND was a small dataset, a data augmentation technique was used to improve the robustness of real-time AWSA. A series of methods, such as random flipping, random cropping, Gaussian noise, and adjusting brightness, were applied to enhance the number of images and masks. The flip angle of the image was randomly selected between 0° and 360°, and the image was randomly cropped to one third of the height or width. Figure 6 shows the results of image data augmentation; each transformed image had its corresponding mask. Table 2 depicts the quantitative comparison of original and enhanced images.



Original Image

Variations

Figure 6. Examples of images generated by data augmentation.

Table 2. Number of images in each dataset.

Training Dataset	Size	Size after Augmentation
WOUND-1	274	1826
WOUND-2	261	1722

4.2. Wound Area Estimation

In this section, we detail the ability of real-time AWSA to estimate wound area in real-world units. Real-time AWSA was also evaluated in regard to its ability to estimate wound area in real-world units using the scale relationship between prior graphics with a

known area and a wound with an unknown area, without considering pixel density. The number of pixels in the image was calculated using a computer, and the prior graphics were taken using the same smartphone as for the wound, with no resolution difference. The wound area in real-world units was calculated using Equation (1):

- Real area (S_{wReal}): the area of the ground truth mask.
- Estimated area (S_{wEst}): the estimated area of the wound region.
- MBR area (S_{wMbr}): wound area determined using a manual measurement method.

To evaluate the ability of real-time AWSA to estimate the wound area, we used Equation (2) to calculate the percentage error *E* for evaluating the ability of real-time AWSA to estimate wound area, where *s* is the true area and \hat{s} is the estimated area.

$$E = \frac{|s - \hat{s}|}{S} \times 100\% \tag{2}$$

4.3. Experimental Details

A gradient descent decay operation was used to find the learning rate during the training. The initial learning rate was set to 0.01, and the batch size during training was set to 8. Furthermore, we employed model training using pretrained weights and compared it with training from scratch, as shown in Figure 7. In terms of validation loss, training with pretrained weights converged to 0.015 after around 20 epochs, while training from scratch fluctuated around 0.04 even at the end of training. From epoch 3, the validation dice value of the model with pretrained weights showed better performance than the training from scratch. As depicted in the figure, the validation dice value of the pretrained weight strategy was stable at around 0.995, while the training strategy from scratch was stable at around 0.981.



Figure 7. Comparison of training using pretrained weights versus training from scratch along with training epochs. (**Left**): comparison of validation loss between two training strategies. (**Right**): comparison of validation dice between two training strategies.

5. Results and Discussion

The performance of real-time AWSA in terms of segmenting wound images and estimating wound area is detailed in this section.

5.1. Testing Metrics

The calculation formula for wound segmentation accuracy is as follows:

$$mAP = \frac{1}{|Q_R|} \sum_{q \in Q_R} AP(q) \tag{3}$$

mAP is an important indicator to measure the accuracy of segmentation, where Q_R represents the number of verification datasets.

m-IoU is the average of the intersection-over-union ratio of the real label and the predicted segmentation. The larger the ratio, the more accurate the segmentation. The formula is as follows:

$$mIoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} + p_{ii}}$$
(4)

where p_{ii} represents predicting category *i* as category *j*.

The recall represents the ratio of the predicted wound area to the real wound area. The closer the ratio is to 1, the more accurate the segmentation. The formula is as follows:

$$Recall = TP/(TP + FN)$$
(5)

The dice score is the harmonic mean of precision and recall and reflects the segmentation accuracy.

5.2. Segmentation Performance of Different Network Structures on Two Wound Datasets

We adopted the STDC-Net classification network as the backbone of the segmentation model in this study. The models were divided into STDC-Net813 and STDC-Net1446 based on the complexity of the model. Experiments were conducted on two wound datasets, WOUND-1 and WOUND-2, to demonstrate the effectiveness of our adapted model. Considering STDC-Net as the benchmark, the models were divided into eight categories based on whether they used pretrained weights or added a coordinated attention mechanism. Each segmentation method was implemented on each image in each test dataset to evaluate the effectiveness of all models. Then, we calculated six metrics: mAP, mIoU, recall, dice score, FPS, and AUC. Table 3 presents the results obtained for all networks using WOUND-1 and WOUND-2. As depicted in the table, the model using pretrained weights had better segmentation accuracy, and the network segmentation accuracy was further improved using the coordinated attention mechanism. All indicators were improved, but the FPS was reduced by about 9%. Further, the performance of STDCNet_CA813_Pretrain was slightly lower than that of STDCNet_CA1446_Pretrain in terms of dice score and AUC. On comparing the FPS, that of STDCNet_CA813_Pretrain was found to be 25% higher than that of STDCNet_CA1446_Pretrain, reflecting the higher processing speed of the network. Because of the complexity of the network, STDCNet_CA813_Pretrain was slightly inferior to STDCNet813 in terms of processing speed, but was significantly better than STDCNet813 in terms of segmentation accuracy. Therefore, considering all aspects, we selected STDCNet_CA813_Pretrain as the model for real-time AWSA.

WOUND-1							
Model	mAP	m-IoU	Recall	Dice Score	FPS	AUC	
STDCNet813	0.8981	0.8274	0.9766	0.7541	111.78 *	0.9813	
STDCNet813_Pretrain	0.9321	0.8736	0.9837	0.8331	110.66	0.9932	
STDCNet_CA813	0.9062	0.8444	0.9866	0.7621	100.12	0.9841	
Real-Time AWSA (STDCNet_CA813_Pretrain)	0.9481 *	0.8928 *	0.9877 *	0.8473	101.08	0.9938	
STDCNet1446	0.8749	0.8011	0.9695	0.6831	84.26	0.9779	
STDCNet1446_Pretrain	0.9333	0.8713	0.9824	0.8401	85.32	0.9934	
STDCNet_CA1446	0.8904	0.8026	0.9672	0.7047	78.02	0.9784	
STDCNet_CA1446_Pretrain	0.9462	0.8852	0.9831	0.8496 *	76.49	0.9951 *	
WOUND-2							
Model	mAP	m-IoU	Recall	Dice Score	FPS	AUC	
STDCNet813	0.8935	0.8441	0.9773	0.7632	111.33 *	0.9783	
STDCNet813_Pretrain	0.9377	0.8739	0.9896	0.8347	109.38	0.9938	
STDCNet_CA813	0.8995	0.8427	0.9881	0.7597	101.58	0.9841	
Real-Time AWSA (STDCNet_CA813_Pretrain)	0.9477 *	0.8944 *	0.9883 *	0.8495	102.11	0.9949	
STDCNet1446	0.8767	0.8019	0.9672	0.6822	82.27	0.9822	
STDCNet1446_Pretrain	0.9339	0.8762	0.9864	0.8451	85.88	0.9954	
STDCNet_CA1446	0.8924	0.8077	0.9692	0.7067	78.59	0.9761	
STDCNet_CA1446_Pretrain	0.9471	0.8905	0.9882	0.8511 *	77.24	0.9952 *	

Table 3. Evaluation of the segmentation methods for each dataset. The values marked with * are the best results. All values are percentages.

5.3. Comparison of Real-Time AWSA with State-of-the-Art Methods

Next, we compared real-time AWSA with state-of-the-art models, mainly including ASURA (U-Net) and DeepLabv3+. Chino et al. proved that ASURA using U-Net as its backbone outperformed CL-Measure, superpixel-based ICARUS, and DL-based QTDU. Therefore, we mainly used ASURA with U-Net as the backbone and the general image semantic segmentation model DeepLabv3+ as the comparison object. As depicted in Table 4, our model performed best on the mAP, m-IoU, recall, dice score, FPS, and AUC metrics. Specifically, the mAP, m-IoU, recall, and dice score values of real-time AWSA on the two wound datasets were about 0.14%, 2.7%, 0.44%, and 0.14% higher than those of the second-ranked method, ASURA (U-Net), respectively. This confirmed that our improved model had a better segmentation capability. Meanwhile, the FPS values of our method for the two wound datasets were 100.08 and 102.11, which were about 42% higher than those of the second-ranked method, Deeplabv3+, reflecting better real-time performance. Further, the AUC best reflected the overall segmentation performance of the models. In this study, our network achieved AUC levels of 0.9938 and 0.9949, demonstrating the robustness of our method.

Figure 8 presents the segmentation outputs from the WOUND-1 and WOUND-2 datasets. Using both datasets, the output of real-time AWSA was extremely close to the ground truth. Due to the relatively small size of the training datasets, Deeplabv3+ faced issues in correctly segmenting wounds. When we compared the detailed visual results of ASURA with U-Net as its backbone and the findings of our method, as shown in Figure 8, we found that it achieved significant results but still lacked enough semantic information and had obvious false-positive segmentation. These findings proved that our method was more robust.

WOUND-1						
Model	mAP	m-IoU	Recall	Dice Score	FPS	AUC
Deeplabv3+	0.8931	0.8293	0.9757	0.7604	58.45	0.9811
ASURA(U-Net)	0.9319	0.8690	0.9833	0.8463	55.04	0.9889
Real-Time AWSA	0.9451 *	0.8928 *	0.9877 *	0.8473 *	100.08 *	0.9938 *
WOUND-2						
Model	mAP	m-IoU	Recall	Dice Score	FPS	AUC
Deeplabv3+	0.8947	0.8255	0.9713	0.7604	58.74	0.9788
ASURA(U-Net)	0.9332	0.8697	0.9839	0.8471	54.22	0.9898
Real-Time AWSA	0.9477 *	0.8944 *	0.9883 *	0.8485 *	102.11 *	0.9949 *

Table 4. Results of different algorithms for WOUND-1 and WOUND-2 datasets. The bold values are our proposed model, and the values marked with * are the best results. All values are percentages.



Figure 8. Wound segmentation of images from WOUND-1 and WOUND-2.

As the AUC value, which is essentially the area under the receiver operating characteristic (ROC) curve, can comprehensively reflect the segmentation capability, we compared the ROC curves of the different models. Figure 9 shows that our network outperformed the others for both datasets.



Figure 9. ROC curves of different models for wound segmentation. (Left): WOUND-1, (Right): WOUND-2.

5.4. Wound Area Estimation

We evaluated the ability of real-time AWSA to measure wound area in real-world units, such as cm². We calculated the wound areas in the test dataset that had already been measured by medical experts at the hospital, and these measured wounds served as standards. Meanwhile, we calculated the error of the manual measurement method and the method proposed in this study relative to the expert standard. The results presented in this section are the average of all test images.

Figure 10 shows the automatic wound area estimation system, consisting of a height platform, a smart phone, a ranging laser sensor, and a personal computer. First, a smart-phone was used to carry laser ranging sensors to collect images of prior graphics at different heights. In this study, the shooting height ranged from 100 to 500 mm, with an interval of 10 mm. According to the collected image data, the polynomial fitting method was used to build the relationship model between prior graphics pixels and shooting heights (Figure 8). In the next step, the wound images were captured at any height from 100 mm to 500 mm using the same method. The same camera was used to acquire prior images and wound images with the same number of pixels. Finally, the segmented wound area was calculated using Equation (1).

Figure 11 shows some examples of area estimation. We measured the area of the wound images from two test datasets using three area estimation methods and obtained the relative error against the gold standard of human experts. As depicted in Table 5, the relative errors of the area estimates using the MBR method for the two datasets were 30.5% and 33.9%, respectively, and the errors using the thin-film edge labeling method were 8.1% and 6.3%, respectively. Although the thin-film edge labeling method demonstrated very small errors, the wound was not of a regular shape, and the wound edge might not have covered a full square grid, reducing the accuracy of the area estimation. Moreover, the counting of squares is time-consuming and laborious, delaying wound treatment. Notably, the errors in wound area estimation using the method proposed in this study were 3.7% and 3.1%, showing the best estimation results. Importantly, this method did not rely on measurement tools and exhibited better real-time performance.



Figure 10. Automatic wound area estimation system, consisting of a height platform, a smart phone, a ranging laser sensor, and a personal computer.



Figure 11. Area estimation for two wound datasets using the real-time AWSA framework.

Table 5. Relative error (%) of the three wound area estimation methods compared to the human expert gold standard.

Datasets	Relative Error (%)						
	Number	Proposed Method	Square Counting	MBR Method			
WOUND-1	30	3.1	6.3	30.5			
WOUND-2	30	3.7	8.1	33.9			

6. Conclusions

In this study, we explored methods for evaluating large-scale wounds in emergency situations and proposed the real-time AWSA framework to automatically segment wound images and estimate the area of a wound. Real-time AWSA used the STDC-Net classification network as its backbone, eliminated structural redundancy, adopted a pretrained weight model, and improved the network processing speed while accurately segmenting the wound, realizing a segmentation accuracy-prediction speed trade-off. A coordinated attention mechanism was introduced to further improve the network segmentation performance. Moreover, we constructed a functional relationship model between prior graphics pixels and shooting heights to perform wound area measurements without contact and measurement tools. We evaluated real-time AWSA using two wound datasets, WOUND-1 and WOUND-2, and found that the accuracy was greatly improved compared with the current state-of-the-art methods. The experimental results showed that real-time AWSA outperformed the state-of-the-art methods in terms of mAP, mIoU, recall, and dice score. The AUC value, which most reflected the comprehensive segmentation capability, also reached the highest level of about 99.5%. The FPS values of our proposed segmentation method in the two wound datasets were 100.08 and 102.11, respectively, which were about 42% higher than those of the second-ranked method, reflecting better real-time performance. Further, real-time AWSA could automatically estimate the wound area in square centimeters with relative errors of only 3.7% and 3.1% in the two test datasets, respectively, showing the best estimation results.

The method proposed in this study could quickly process a large number of collected wound images for trauma treatment in emergency environments, areas with scarce medical resources, or trauma patients with limited mobility. The main tasks included the automatic segmentation of the wound area and automatic estimation. The wound information could be uploaded and sent to telemedicine experts to achieve immediate treatment and real-time wound care. The current disadvantage is that the method in this paper was mainly aimed at determining the two-dimensional area of a wound, without considering the curvature factor and depth information. In addition, the deep learning method may lose some semantic information in areas where the color transition of the wound is not clear. In the future, deep learning combined with 3D reconstruction could deal with more complex wounds and solve these problems, which is our current research focus.

Author Contributions: Y.S. and W.L. designed the study. W.M. and F.Z. conducted the review of relevant literature. Y.S. constructed the network architecture and the area estimation method and wrote the manuscript. Z.S. wrote the application for the ethics approval. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [Natural Science Foundation of Chongqing] grant number [2020ZX 1200048] And The APC was funded by [Natural Science Foundation of Chongqing].

Institutional Review Board Statement: This article does not relate to any studies with human participants performed by any of the authors.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created.

Acknowledgments: The authors thank the National Trauma Database and Chengde County Hospital for providing wound data.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Burnette, G. Information: The battlefield of the future. Surf. Warf. 1995, 20, 8.
- Mirzaalian-Dastjerdi, H.; Topfer, D.; Rupitsch, S.J.; Maier, A.K. Measuring surface area of skin lesions with 2d and 3d algo-rithms. *Int. J. Biomed. Imaging* 2019, 2019, 4035148. [CrossRef] [PubMed]

- 3. Shirley, T.; Presnov, D.; Kolb, A. A lightweight approach to 3d measurement of chronic wounds. J. WSCG 2018, 26, 67–74. [CrossRef]
- Wagh, A.; Jain, S.; Mukherjee, A.; Agu, E.; Pedersen, P.C.; Strong, D.; Tulu, B.; Lindsay, C.; Liu, Z. Semantic segmentation of smartphone wound images: Comparative analysis of ahrf and cnn-based approaches. *IEEE Access* 2020, *8*, 181590–181604. [CrossRef]
- 5. Shetty, R.; Sreekar, H.; Lamba, S.; Gupta, A.K. A novel and accurate technique of photographic wound measurement. *Indian J. Plast. Surg. Off. Publ. Assoc. Plast. Surg. India* **2012**, *45*, 425–429. [CrossRef]
- Pereyra, L.C.; Pereira, S.M.; Souza, J.P.; Frade, M.A.C.; Rangayyan, R.M.; Azevedo-Marques, P.M. Characterization and pattern recognition of color images of dermatological ulcers—A pilot study. *Comp. Sci. J. Mold.* 2014, 22, 211–235.
- Cazzolato, M.T.; Rodrigues, L.S.; Scabora, L.C.; Zaboti, G.F.; Vasconcelos, G.Q.; Chino, D.Y.T.; Jorge, A.E.S.; Cordeiro, R.L.F.; Traina-Jr, C.; Traina, A.J.M. 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, Fortaleza, Brazil, 7–10 October 2019; pp. 109–120.
- Sethy, P.K.; Behera, S.K. Detection of coronavirus disease (COVID-19) based on deep features. *Preprints* 2020, 2020, 2020030300.
 Hassantabar, S.; Ahmadi, M.; Sharifi, A. Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray
- Hassandolf, S., Annadi, M., Sharin, A. Diagnosis and detection of infected fissue of COVID-19 patients based on fung x-ray image using convolutional neural network approaches. *Chaos Solitons Fractals* 2020, 140, 110170. [CrossRef]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems—Volume 1, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
- Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional net-works for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 640–651. [CrossRef]
- 12. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890. [CrossRef]
- 13. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* 2017, arXiv:1706.05587.
- Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818. [CrossRef]
- 15. Yuan, Y.; Chao, M.; Lo, Y. Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance. *IEEE Trans. Med. Imaging* **2017**, *36*, 1876–1886. [CrossRef] [PubMed]
- 16. Al-masni, M.A.; Al-antari, M.A.; Choi, M.-T.; Han, S.-M.; Kim, T.-S. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput. Methods Programs Biomed.* **2018**, *162*, 221–231. [CrossRef] [PubMed]
- 17. Tang, P.; Liang, Q.; Yan, X.; Xiang, S.; Sun, W.; Zhang, D.; Coppola, G. Efficient skin lesion segmentation using separable-U-Net with stochastic weight averaging. *Comput. Methods Programs Biomed.* **2019**, *178*, 289–301. [CrossRef]
- Kawahara, J.; Hamarneh, G. Multi-Resolution-Tract CNN with Hybrid Pretrained and Skin-Lesion Trained Layers. In International Workshop on Machine Learning in Medical Imaging; Springer: Cham, Switzerland, 2016; pp. 164–171.
- 19. Yu, L.; Chen, H.; Dou, Q.; Qin, J.; Heng, P. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans. Med. Imaging* **2017**, *36*, 994–1004. [CrossRef]
- Goyal, M.; Yap, M.H.; Reeves, N.D.; Rajbhandari, S.; Spragg, J. Fully convolutional networks for diabetic foot ulcer segmentation. In Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 618–623.
- 21. Zahia, S.; Sierra-Sosa, D.; Garcia-Zapirain, B.; Elmaghraby, A. Tissue classification and segmentation of pressure injuries using convolutional neural networks. *Comput. Methods Programs Biomed.* **2018**, *159*, 51–58. [CrossRef] [PubMed]
- 22. Navarro, F.; Escudero-Violo, M.; Bescs, J. Accurate segmentation and registration of skin lesion images to evaluate lesion change. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 501–508. [CrossRef]
- Blanco, G.; Bedo, M.V.N.; Cazzolato, M.T.; Santos, L.F.D.; Jorge, A.E.S.; Traina, C.; Azevedo-Marques, P.M.; Traina, A.J.M. A label-scaled similarity measure for content-based image retrieval. In Proceedings of the 2016 IEEE International Symposium on Multimedia (ISM), San Jose, CA, USA, 11–13 December 2016; pp. 20–25. [CrossRef]
- Seixas, J.L.; Barbon, S.; Mantovani, R.G. Pattern recognition of lower member skin ulcers in medical images with machine learning algorithms. In Proceedings of the IEEE International Symposium on Computer Based Medical Systems, Sao Carlos, Brazil, 22–25 June 2015; pp. 50–53. [CrossRef]
- Chino, D.Y.T.; Scabora, L.C.; Cazzolato, M.T.; Jorge, A.E.S.; Traina, C.; Traina, A.J.M. ICARUS: Retrieving skin ulcer images through bag-of-signatures. In Proceedings of the IEEE International Symposium on Computer-Based Medical Systems, Karlstad, Sweden, 18–21 June 2018; pp. 82–87. [CrossRef]
- 26. Little, C.; McDonald, J.; Jenkins, M.G.; McCarron, P. An overview of techniques used to measure wound area and volume. *J. Wound Care* **2009**, *18*, 250–253. [CrossRef]
- 27. Chino, D.Y.T.; Scabora, L.C.; Cazzolato, M.T.; Jorge, A.E.S.; Traina-Jr, C.; Traina, A.J.M. Segmenting skin ulcers and measuring the wound area using deep convolutional networks. *Comput. Methods Programs Biomed.* **2020**, *191*, 105376. [CrossRef]

- Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-ResNet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284.
- 29. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
- Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
- Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1520–1528. [CrossRef]
- Cazzolato, M.T.; Ramos, J.S.; Rodrigues, L.S.; Scabora, L.C.; Chino, D.Y.T.; Jorge, A.E.; de Azevedo-Marques, P.M.; Traina, C., Jr.; Traina, A.J. The UTrack framework for segmenting and measuring dermatological ulcers through telemedicine. *Comput. Biol. Med.* 2021, 134, 104489. [CrossRef]
- Dorileo, E.A.G.; Frade, M.A.C.; Rangayyan, R.M.; Azevedo-Marques, P.M. Segmentation and analysis of the tissue composition of dermatological ulcers. In Proceedings of the CCECE, Calgary, AB, Canada, 2–5 May 2010; pp. 1–4. [CrossRef]
- Blanco, G.; Traina, A.J.M.; Traina, C., Jr.; Azeve-do-Marques, P.M.; Jorge, A.E.S.; de Oliveira, D.; Bedo, M.V.N. A superpixel-driven deep learning approach for the analysis of dermatological wounds. *Comput. Methods Programs Biomed.* 2020, 183, 105079. [CrossRef] [PubMed]
- 35. Fan, M.Y.; Lai, S.Q.; Huang, J.S. Rethinking BiSeNet for Real-time Semantic Segmentation. arXiv 2021, arXiv:2104.13188. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.