

Article

FUSeG: The Foot Ulcer Segmentation Challenge

Chuanbo Wang^{1,†}, Amirreza Mahbod^{2,*,†} , Isabella Ellinger³ , Adrian Galdran⁴, Sandeep Gopalakrishnan⁵, Jeffrey Niezgoda⁶ and Zeyun Yu¹

¹ Big Data Analytics and Visualization Laboratory, Department of Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA

² Research Center for Medical Image Analysis and Artificial Intelligence, Department of Medicine, Danube Private University, 3500 Krems an der Donau, Austria

³ Institute for Pathophysiology and Allergy Research, Medical University of Vienna, 1090 Vienna, Austria

⁴ Department of Computing and Informatics, Bournemouth University, Bournemouth BH12 5BB, UK

⁵ Wound Healing and Tissue Repair Laboratory, School of Nursing, College of Health Professions and Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA

⁶ Advancing the Zenith of Healthcare (AZH) Wound and Vascular Center, Milwaukee, WI 53211, USA

* Correspondence: amirreza.mahbod@dp-uni.ac.at

† These authors contributed equally to this work.

Abstract: Wound care professionals provide proper diagnosis and treatment with heavy reliance on images and image documentation. Segmentation of wound boundaries in images is a key component of the care and diagnosis protocol since it is important to estimate the area of the wound and provide quantitative measurement for the treatment. Unfortunately, this process is very time-consuming and requires a high level of expertise, hence the need for automatic wound measurement methods. Recently, automatic wound segmentation methods based on deep learning have shown promising performance; yet, they heavily rely on large training datasets. A few wound image datasets were published including the Diabetic Foot Ulcer Challenge dataset, the Medetec wound dataset, and WoundDB. Existing public wound image datasets suffer from small size and a lack of annotation. There is a need to build a fully annotated dataset to benchmark wound segmentation methods. To address these issues, we propose the Foot Ulcer Segmentation Challenge (FUSeG), organized in conjunction with the 2021 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). It contains 1210 pixel-wise annotated foot ulcer images collected over 2 years from 889 patients. The submitted algorithms are reviewed in this paper and the dataset can be accessed through the Foot Ulcer Segmentation Challenge website.

Keywords: semantic segmentation; chronic wounds; foot ulcers; wound segmentation; challenge; benchmark



Citation: Wang, C.; Mahbod, A.; Ellinger, I.; Galdran, A.; Gopalakrishnan, S.; Niezgoda, J.; Yu, Z. FUSeG: The Foot Ulcer Segmentation Challenge. *Information* **2024**, *15*, 140. <https://doi.org/10.3390/info15030140>

Academic Editor: Xiaoshuang Shi

Received: 31 January 2024

Revised: 25 February 2024

Accepted: 27 February 2024

Published: 1 March 2024



Copyright: © 2024 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/>).

1. Introduction

Chronic non-healing and acute wounds represent a difficult challenge for healthcare systems, affecting millions of patients globally [1]. In the United States, the cost prediction for wound care treatment is estimated to be between USD 28.1 and USD 96.8 billion [2]. In contrast to acute wounds, chronic wounds fail to predictably proceed through the common phases of healing in an orderly and timely fashion; thus, hospitalization and additional care are necessary but increase the costs of the health services by billions annually [3]. The outpatient costs (USD 9.9–USD 35.8 billion) of wound care are reported to be significantly higher than inpatient costs (USD 5.0–USD 24.3 billion) [4]. However, the access and quality of care to chronic wound patients are often limited in primary and rural healthcare settings. A vast majority of chronic wound patients usually have other health conditions such as diabetes, obesity, and circulation problems. The shortage of well-trained wound care clinicians also worsens the situation. Accurate and fast measurement of the wound area is critical to the management and evaluation of chronic wounds to monitor the wound healing

process and to determine future interventions. Clinicians typically measure wounds by their length, width, and depth in clinical practices. Wound length and width are measured with a ruler guide and the depth is measured with Q-tips. However, manual measurement is time-consuming and often inaccurate, which can cause a negative impact on patients. Reliable automatic wound segmentation from images would enable the automation of measurement of the wound area as well as efficient data entry into the electronic medical record to enhance patient care. With the recent advances in deep learning, image-based semantic segmentation of wounds offers a comprehensive characterization of the wound. However, deep-learning methods impose an even larger burden of manual effort than most manual measurements, since they need to be trained on large datasets with pixel-wise labeled images. A few wound datasets are publicly available. The Diabetic Foot Ulcer (DFU) Challenge dataset [5] contains 15,683 DFU image patches. Unfortunately, the images are not labeled with segmentation masks. The Medetec wound dataset [6] consists of free stock images of all types of open wounds such as venous leg ulcers, arterial leg ulcers, pressure ulcers (pressure sores), and diabetic ulcers. Despite covering almost all wound types, the dataset only contains 341 images, with manual segmentation masks provided for 152 images. However, the size of the dataset is far from sufficient for training deep learning models. WoundDB [7] contains 188 sets of wound photographs where each set includes four modalities: RGB image, thermal image, stereo image, and depth map. The wounds are fully labeled with outlined boundaries. Like Medetec, the number of images in WoundDB is not sufficient for training deep segmentation models.

2. Materials and Methods

2.1. Dataset

2.1.1. Data Collection

We collaborated with the Advancing the Zenith of Healthcare (AZH) Wound and Vascular Center to build a chronic wound dataset. Data were collected over 2 years from October 2019 to April 2021 at the center and contain 1210 foot ulcer images taken from 889 patients during multiple clinical visits. The raw images were taken by Canon SX 620 HS digital camera and iPad Pro under uncontrolled illumination conditions, with various backgrounds. The images (some examples shown in Figure 1) are randomly split into three subsets: a training set with 810 images, a validation set with 200 images, and a test set with 200 images. Of course, the annotations of the test set are kept private. We confirm that the data collected were de-identified and in accordance with relevant guidelines and regulations and the patient's informed consent is waived by the institutional review board of the University of Wisconsin-Milwaukee. An electronic laboratory notebook was not used. All images were zero-padded and had a unique size of 512×512 pixels.

2.1.2. Annotation

Deep learning models learn the annotations of the training dataset during training. Thus, the quality of annotations is essential. Automatic annotation generated with computer vision algorithms is not ideal when deep learning models are trained to learn how human experts recognize the wound region. For efficiency, initial annotation masks were firstly manually proposed for the raw images by nonexpert annotators using photoshop, where the masks are created as a new layers on top of the original images. Graphics tablets were used as the hardware for drawing the boundary of wounds on the mask layer. These raw proposals were sent to the AZH woundcare center as photoshop documents (.psd) files to be further reviewed and modified by the annotation team, which consists of two MDs with over 20 years of wound care experience and two medical assistants who followed up on the patient visits. For tricky cases and disagreements, the annotation team would discuss about the questionable areas of the wounds in a meeting and make final decisions on whether these areas should be labeled as wound or non-wound. Tissues found in a foot ulcers could be categorized as callous, eschar, fibrin, granulation, necrotic, neodermis, tendon, muscle, bone, etc. During our annotation process, all non-epithelialized tissues

(e.g., granulation, slough, eschar, subcutaneous, tendon, muscle, bone) were annotated as wounds. The annotation masks in the finalized psd. files were exported as png. images, where the wound areas were labeled with pixel intensities of 1 and the non-wound areas were labeled with 0. The AZH Wound and Vascular Center, Milwaukee, WI, consented to make our dataset publicly available.



Figure 1. The challenge dataset consists of 1210 foot ulcer images taken from 889 patients. The first row contains some examples of the raw images collected. The second row consists of the corresponding segmentation masks.

2.2. Challenge Design

2.2.1. Infrastructure and Timeline

The homepage of our challenge was hosted on the Grand Challenge platform and the dataset is stored in our repository on GitHub. We created an individual webpage to present the leaderboard. The challenge was announced on 24 February 2021, and the training set was published on 9 March 2021. In early July we made the testing set available for participants to perform sanity checks and the submission was closed on 15 July 2021. The testing results and rankings for FUSeg2021 were published on 16 August 2021, 2 months before our workshop in MICCAI 2021.

2.2.2. Submission and Evaluation

Each participating team was asked to submit a Docker container that contains their algorithm and the prediction code. We generated segmentation predictions using the code and re-produced the results to be evaluated on our GPU server and all results were evaluated under the same software and hardware environment. The Dice coefficient [8] was used to quantify the performances of submitted algorithms:

$$\text{Dice} = \frac{2 \times \text{True positives}}{2 \times \text{True positives} + \text{False negatives} + \text{False positives}} \quad (1)$$

where true positives, false negatives, and false positives represent the corresponding number of pixels. Over 100 researchers registered for the FUSeg2021 challenge and eight teams successfully submitted their algorithms. It is also worth mentioning that using external datasets for pre-training was allowed for participants, as long as they clearly mentioned the used datasets in their approach.

In Section 3, we provide an in-depth description of the methods of the top three ranked teams.

2.3. Top Three Submissions

2.3.1. First Place

This submission [9] was made by Amirreza Mahbod (Danube Private University, Medical University of Vienna), Rupert Ecker (TissueGnostics GmbH), and Isabella Ellinger (Medical University of Vienna). Two CNN models were used, namely U-Net [10] and LinkNet [11]. Instead of using the models in their plain forms, pre-trained CNNs were used

in the decoder parts of the models. For LinkNet, a pre-trained EfficientNetB1 model [12] was used, and EfficientNetB2 model [12] was utilized for U-Net. LinkNet with EfficientNetB1 backbone comprised of around 8.6 million parameters (32.9 Mb), and U-Net with EfficientNetB2 consisted of around 14.3 million parameters (54.5 Mb). As shown in Figure 2, the entire Medetec dataset was also used for pre-training. To train the models, random scaling, random rotations, vertical and horizontal flipping, and brightness and contrast shifts were used as augmentation techniques, as suggested in [13]. Each model was trained for 80 epochs with a learning rate (LR) scheduler that reduced the LR by 90% after every 25 epochs. The initial LR was set to 0.001. The batch size was set to 4 and full-size images were used to train the models. Adam optimizer [14] and a combination of Dice loss and Focal loss [15] were also adopted for model training. For each dataset, five-fold cross-validation was exploited and the best models based on the segmentation scores of the validation sets were saved to be used in the inference phase. To boost the segmentation performance, three distinct ensembling strategies were used, namely 5-fold cross-validation, test time augmentation (TTA), and result fusion from the two exploited models. All these techniques have shown promising results for medical image classification and segmentation tasks in former studies [13,16–18]. Instead of using the entire training set to train a single model, it was divided randomly into five subsets. Then, five sub-models were trained based on the derived subsets (i.e., for each of the sub-models, four subsets were used for training and the hold-out set was used for validation). In the inference phase, the test images were sent to all five derived sub-models and then the segmentation results were averaged. TTA was used in the inference phase for better segmentation performance. Degree rotations of 0, 90, 180, and 270, as well as horizontal flipping, were applied in TTA. Since two distinct models (LinkNet with EfficientNetB1 backbone and U-Net with EfficientNetB2 backbone) were trained, their results were fused in the inference phase for a given test image, as shown in Figure 2. To form the final segmentation masks for the test images, first, the fused prediction probability vectors were binarized using a 0.5 threshold. Two post-processing steps were also applied, namely filling the holes, and removing very small detected objects, with identical settings as described in [19].

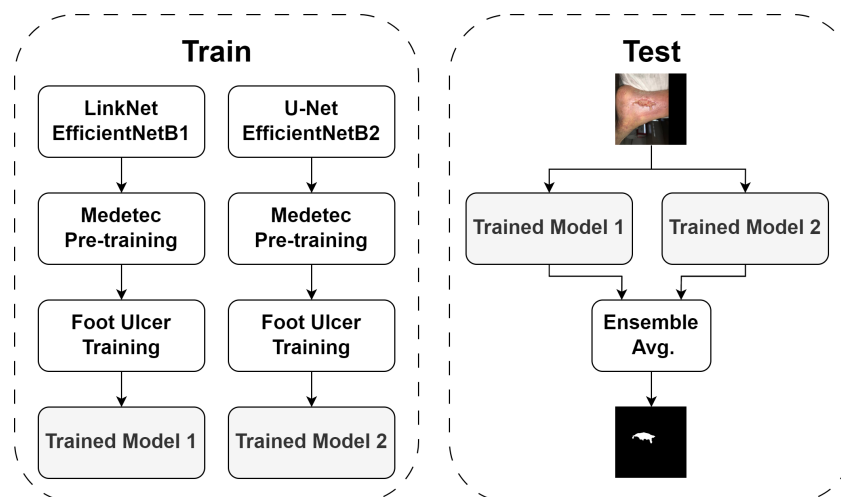


Figure 2. The architecture of Mahbod’s model. In the training phase, LinkNet and U-Net were used to obtain two trained models whose predictions were ensembled in the testing phase.

2.3.2. Second Place

This submission was made by Yichen Zhang from Huazhong University of Science and Technology. In this approach, the Harmonic Densely Connected Network (HarDNet-MSEG) [20] was applied to the wound dataset. As shown in Figure 3, the architecture consisted of an encoder backbone and a decoder. Specifically, HarDNet-68 was adopted as the encoder. HarDNet-68 consisted of around 17.6 million parameters (67.1 Mb) and the repeated application of HarDNet blocks, batch normalizations (BNs), and Max pooling

layers. The arrangement of the layers followed the standard *Conv-BN-ReLU* order to enable the folding of BN. For the layer distribution, instead of concentrating on stride-16, which is adopted by most of the CNN models, stride-8 was used to have the most layers in the HarDNet-68. This improved the local feature learning, benefiting small-scale object awareness. The employed HarDNet blocks were the more efficient version of dense blocks, originally presented in [21]. In each HarDNet block, layer k was connected to layer $k - 2n$ if $2n$ divides k , where n is a non-negative integer and $k - 2n \geq 0$. Under this connection scheme, once layer $2n$ was processed, layers 1 through $2n - 1$ can be flushed from the memory, reducing the concatenation cost significantly. The decoder was implemented as a cascaded partial decoder [22]. It turned out that the shallow features with high resolution occupied computing resources, and the deep information could also represent the spatial details of the shallow information relatively well. Thus, in the cascaded partial decoder architecture, the shallow features were discarded and allowed for more computing on the features of the deeper layers. We also used receptive field blocks (RFBs) in the decoder part of the model to enhance the deep feature learning process [23]. At the same time, the aggregation of features maps at different scales were employed by adding appropriate convolution and skip connections in the dense aggregation module. During training, multi-scale training and standard image augmentations were applied. Further details about the building blocks can be found in [20].

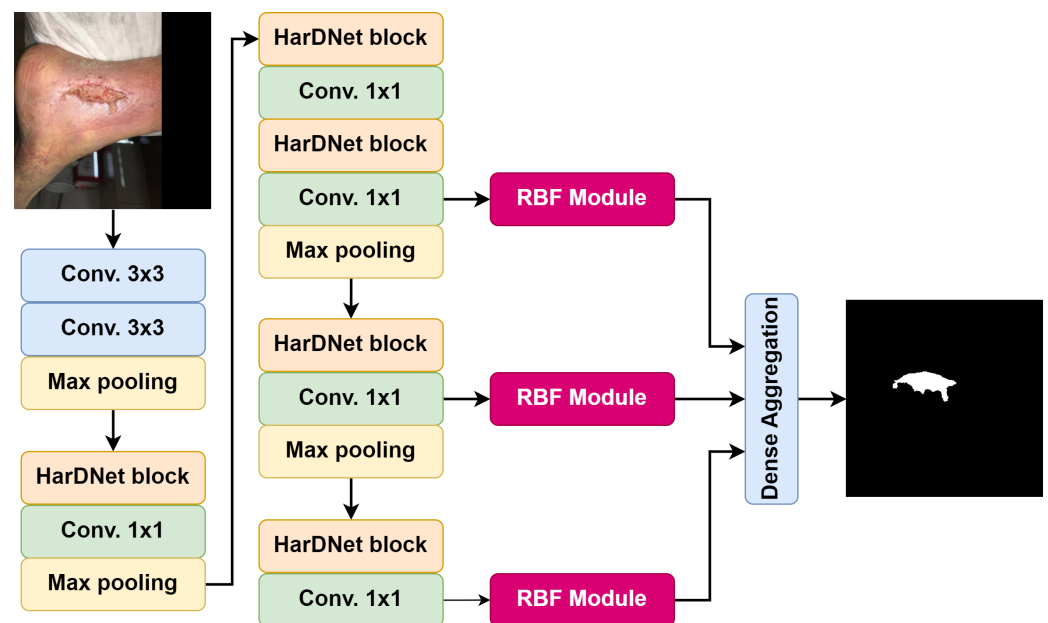


Figure 3. The architecture of HarDNet-MSEG [20]. Conv.: convolutional layer. RBF: receptive field block.

2.3.3. Third Place

This submission [24] was made by Adrian Galdran of the University of Bournemouth. In this work, a framework for semantic segmentation was proposed based on the sequential use of two encoder–decoder networks. The architecture consists of two segmentation networks stacking in a sequential manner, where the second network receives as input the concatenation of the prediction from the first one with the original frame, as shown in Figure 4. This way, the output of the first network acts as an attention mechanism that provides the second network with a map of interesting locations on which the second network should focus. Double encoder–decoders are a direct extension of encoder–decoder architectures in which two encoder–decoder networks are sequentially combined. In this approach, the input RGB image, the first network, and the second network are denoted with x , $E^{(1)}$, and $E^{(2)}$, respectively. In a double encoder–decoder, the output $E^{(1)}(x)$ of the

first network is provided to the second network together with x so that it can act as an attention map that allows $E^{(2)}$ to focus on the most interesting areas of the image:

$$E(x) = E^{(2)}(x, E^{(1)}(x)), \tag{2}$$

where x and $E^{(1)}(x)$ are stacked so that the input to $E^{(2)}$ has four channels instead of the three channels corresponding to the RGB components of x . There are some choices to be made in this framework, specifically about the structure of the encoder and decoder sub-networks within $E^{(1)}$ and $E^{(2)}$. Note that $E^{(1)}$ and $E^{(2)}$ do not need to share the same architecture, although in this work, feature pyramid network with ResNeXt-50 backbone [25,26] was selected for both sub-networks with around 26.4 million parameters (100.8 Mb) for each sub-network.

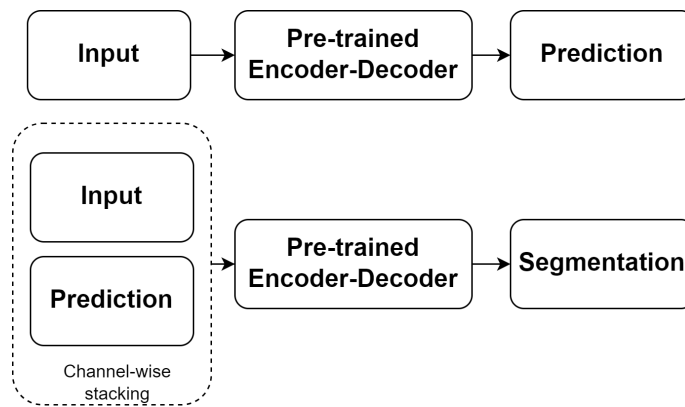


Figure 4. Pre-trained double encoder–decoder network [24]. The second network receives input as the original raw image concatenated with the prediction of the first network. This allows the second network to better focus on the region of interest in the image, improving its segmentation accuracy.

3. Results

In this section, we report the performances of the top three submitted algorithms. Besides the Dice coefficient, we also calculate the precision and recall to demonstrate the performances in detail. As shown in Table 1, the ensemble network (Rank 1), consisting of a U-Net and a LinkNet, achieves the highest precision of 91.55% and the highest Dice coefficient of 88.8%. The Dice coefficient of HarDNet-MSEG (Rank 2) was evaluated to be 87.57% on the testing set of the wound dataset. Rank 2’s entry was rated with the highest recall of 86.31%, but the lowest precision of 88.87%, which resulted in its overall Dice score being relegated to second place. The lowest Recall and Dice scores were determined for the double encoder–decoder model of Rank 3. In detail, the precision, recall, and Dice scores were 90.03%, 86.91%, and 84% respectively.

In total, 103 registered users from 27 different countries downloaded our dataset, and eight teams successfully submitted their algorithms to us.

Table 1. Performance evaluation of the top three submitted algorithms (Dice score was used for the ranking in the challenge).

Rank	Precision (%)	Recall (%)	Dice (%)
First Rank	91.55	86.22	88.80
Second Rank	88.87	86.31	87.57
Third Rank	90.03	84.00	86.91

4. Discussion

4.1. Overall Segmentation Performance

In the annotation process, one of the medical assistants in the annotation team reported that the wound segmentation was difficult for even human experts. One of the challenges is that boundaries between epithelial tissue and granulation tissue are often ambiguous due to their similar color and the fact that epithelial tissue forms on top of granulation tissue. Figure 5 shows some of these cases where deep learning models failed to predict accurate segmentation masks.

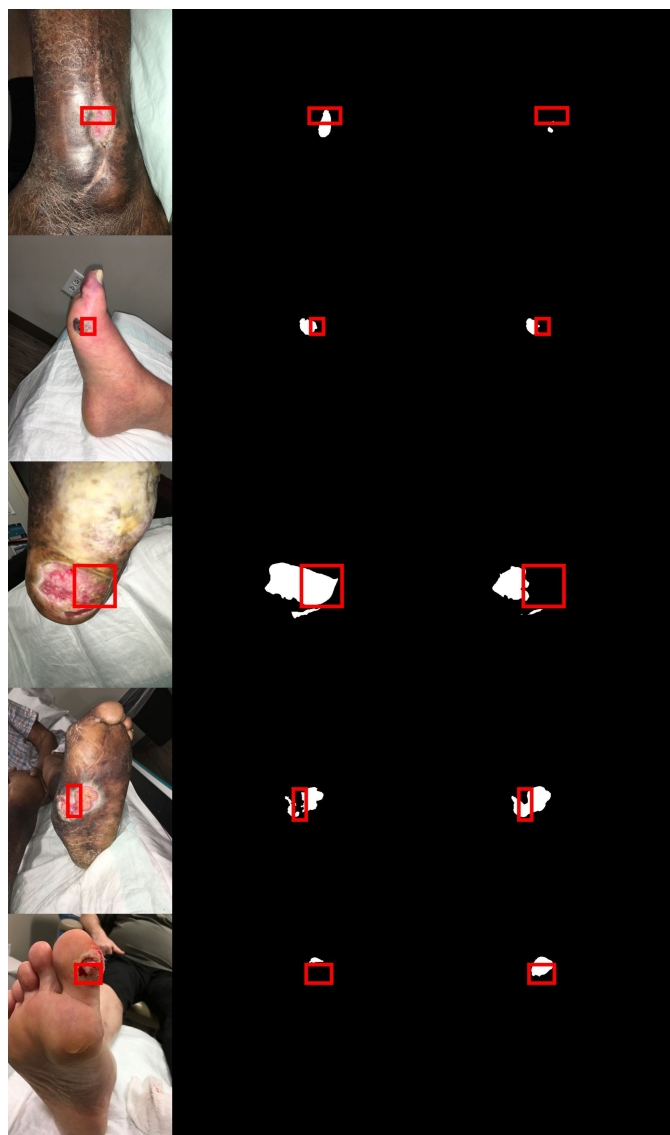


Figure 5. Cases where boundaries between epithelial tissue and granulation tissue are ambiguous. The left column shows the original images. The middle column shows the ground truth label and the right column shows the segmentation predictions.

The overall performance of the submitted algorithms is promising.

In general, all submitted algorithms were capable of segmenting the wound region quite well (Figure 6), with Dice scores of over 80%. According to the literature [27–29], Dice scores over 70% are considered acceptable for medical image segmentation, and all three winning approaches delivered significantly better performances. The best Dice score of 88.8% and a surprising precision of 91.55% were reached by the top team. For the top team, the same approach was applied to the Chronic Wound dataset (a subset of the FuSeg

challenge dataset). As reported in [9], the performance is slightly worse in the FuSeg dataset (precision of 91.55%, recall of 86.22%, and dice of 88.80%) compared to the Chronic Wound dataset (precision of 92.68%, recall of 91.80%, and dice of 92.07%), implying a slight overfitting for the top-ranked approach.



Figure 6. Demonstration of the predicted segmentation masks from the top three submissions. Each row in the top-down order are the original images, ground truth masks, and the model predictions submitted by Mahbod, Zhang, and Galdran, respectively.

In general, segmenting the relatively bigger wounds with clear boundaries worked well when the wound beds are properly cleansed and the dead tissue are removed, and reasonably well for cases where infection, slough, or other impediments were present. As shown in Figure 7, segmenting small isolated areas of the wound with ambiguous boundaries was the most difficult task.

4.2. U-Net Architecture

U-Net is still a popular option for the participants as a subnetwork in ensemble networks or the encoder part of their architecture. For example, U-Net is used as one of the subnetworks pre-trained with carefully selected models in the submission by Mahbod. Galdran also makes use of U-Net by stacking two sequential models where the input of the second network is the concatenation of the original input and the prediction from the first network. Among the submissions, we also find other variants of U-Net such as U-Net with atrous spatial pyramid pooling layers [30] and U-Net with the addition of residual connections [31]. As mentioned in Section 2.3, all three winning solutions were based on CNNs. Transformer-based models have been recently applied to various computer vision

tasks such as classification and segmentation [32,33]; models such as vision transformers (ViTs) [34] or data-efficient image transformers (DeITs) [35] have also been applied to foot ulcer segmentation, but their performance compared to the CNN-based solutions proposed in the FuSeg challenge could be analyzed in future studies. Similarly, previously untested models such as the Segmentation Transformer (SETR) [36] could be adapted and applied for foot ulcer segmentation.

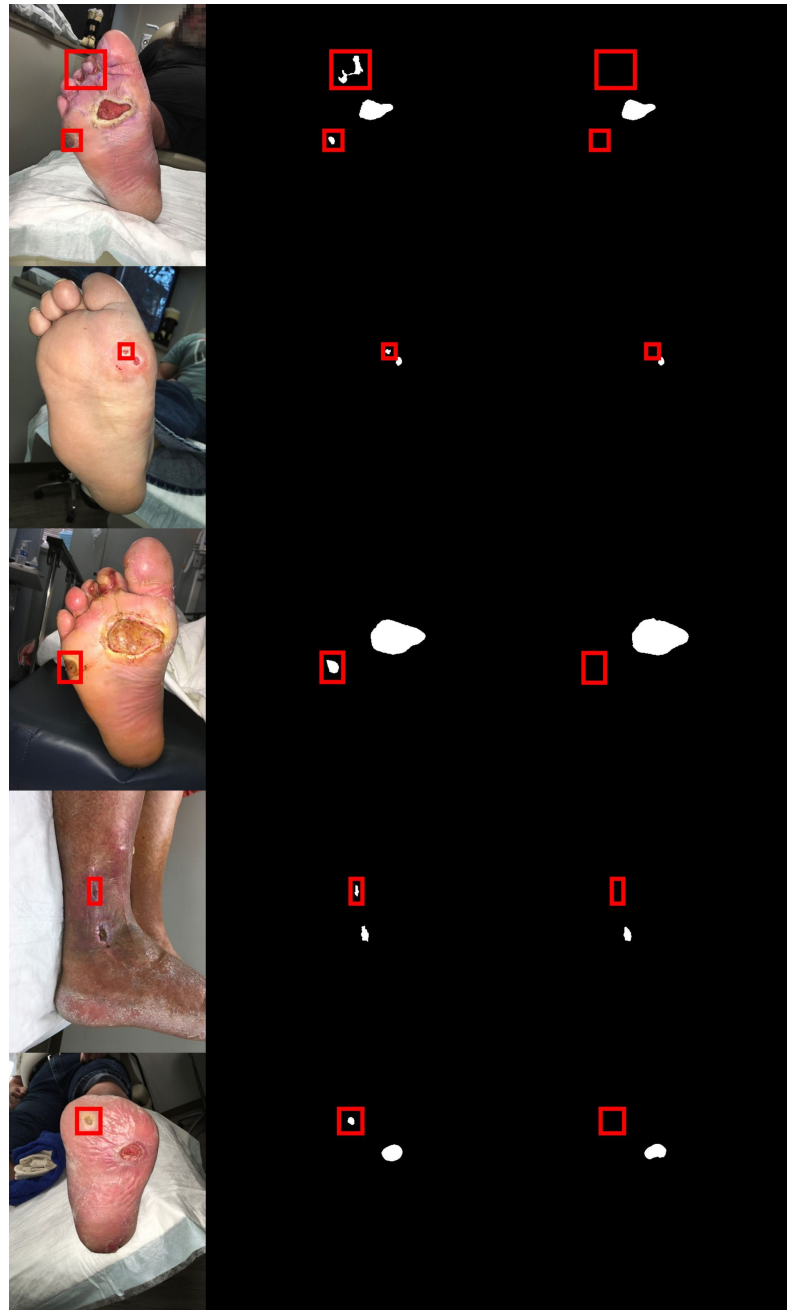


Figure 7. Cases with small isolated wound areas. The left column shows the original images. The middle column shows the ground truth label and the right column shows the segmentation predictions.

4.3. Ensembling

When comparing the two groups of models (ensemble networks vs. single networks), we observed that fusing predictions from different networks always outperforms the single networks in our dataset. This observation agrees with the theory of ensemble learning, ensembles that combine multiple networks tend to yield more generalized and robust

results when there is significant diversity among the networks. In this challenge, we see two of the main ensemble learning methods, stacking and boosting. Both algorithms perform better than the individual networks, namely U-Net and LinkNet, when applied to solve the same problem. However, it should be noted that using techniques such as ensembling or TTA could increase the inference time drastically, as shown in previous studies [37,38]. Therefore, both computational resources and intended applications should be considered carefully when employing such techniques.

5. Conclusions and Future Works

In this paper, we presented the Foot Ulcer Segmentation challenge. We built a large pixel-wise annotated wound image dataset that is manually labeled by experts. Currently, only foot ulcer images were annotated and included in the challenge dataset as these wounds tend to be smaller than other types of chronic wounds, which makes it easier and less time-consuming to manually annotate the pixel-wise segmentation masks. In the future, we plan to create larger image libraries to include all types of chronic wounds, such as venous leg ulcers, pressure ulcers, and surgery wounds as well as non-wound reference images.

We used our dataset to create the FuSeg challenge and evaluated the submitted wound segmentation methods. Although wound segmentation is a difficult task, our results suggest a few takeaways:

- Current state-of-the-art algorithms can accurately segment the wound area, with Dice scores reaching 88.8% and precision reaching 91.55%.
- There is still room for improvement for newer deep learning models to be applied on wound segmentation. For example, transformer-based methods and semi-supervised methods.
- From the predictions generated by the submitted algorithms, we observed the challenges in distinguishing between epithelial tissue and granulation tissue and segmenting small isolated wound regions.
- We also observed the superiority of ensemble networks over individual networks applied to our dataset.

Author Contributions: Conceptualization, C.W.; methodology, C.W., A.M. and A.G.; validation, C.W., A.M. and A.G.; formal analysis, C.W.; investigation, C.W. and A.M.; resources, C.W.; data curation, C.W.; writing—original draft preparation, C.W.; writing—review and editing, C.W., A.M., A.G., I.E., J.N., S.G. and Z.Y.; visualization, C.W.; supervision, I.E., S.G., J.N. and Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The research was partially funded by UWM Research Foundation Catalyst Grant and UWM Discovery and Innovation Grant program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: We confirm that the data collected were de-identified and in accordance with relevant guidelines and regulations and the patient's informed consent is waived by the institutional review board of the University of Wisconsin-Milwaukee.

Data Availability Statement: The used dataset in this article can be found at: <https://fusc.grand-challenge.org/FUSeg-2021/> (accessed on 31 January 2024).

Acknowledgments: C.W. wish to acknowledge the great help provided by the doctors and medical assistants in the AZH Wound and Vascular Center. The assistance provided by Behrouz Rostami in the proposal of initial annotations was greatly appreciated. Adrian Galdran and Amirreza Mahbod (two challenge participants) were partially funded by the Marie Skłodowska-Curie Fellowship, No. 892297 and the Austrian Research Promotion Agency (FFG), No. 872636, respectively.

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

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
AZH	Advancing the Zenith of Healthcare
BN	Batch Normalization
DFU	Diabetic Foot Ulcer
FUSeg	Foot Ulcer Segmentation
HarDNet	Harmonic Densely Connected Network
LR	Learning Rate
MICCAI	International Conference on Medical Image Computing and Computer Assisted Intervention
TTA	Test Time Augmentation

References

- Frykberg, R.G.; Banks, J. Challenges in the Treatment of Chronic Wounds. *Adv. Wound Care* **2015**, *4*, 560–582. [CrossRef]
- Sen, C.K. Human Wounds and Its Burden: An Updated Compendium of Estimates. *Adv. Wound Care* **2019**, *8*, 39–48. [CrossRef] [PubMed]
- Branski, L.K.; Gauglitz, G.G.; Herndon, D.N.; Jeschke, M.G. A review of gene and stem cell therapy in cutaneous wound healing. *Burns* **2009**, *35*, 171–180. [CrossRef] [PubMed]
- Nussbaum, S.R.; Carter, M.J.; Fife, C.E.; DaVanzo, J.; Haught, R.; Nusgart, M.; Cartwright, D. An Economic Evaluation of the Impact, Cost, and Medicare Policy Implications of Chronic non-healing Wounds. *Value Health* **2018**, *21*, 27–32. [CrossRef] [PubMed]
- Yap, M.H.; Cassidy, B.; Pappachan, J.M.; O’Shea, C.; Gillespie, D.; Reeves, N.D. Analysis Towards Classification of Infection and Ischaemia of Diabetic Foot Ulcers. In Proceedings of the EMBS International Conference on Biomedical and Health Informatics, Athens, Greece, 27–30 July 2021; pp. 1–4. [CrossRef]
- Thomas, S. Medetec Wound Database. 2014. Available online: <http://www.medetec.co.uk/files/medetec-image-databases.html> (accessed on 22 January 2024).
- Kręcichwost, M.; Czajkowska, J.; Wijata, A.; Juszczak, J.; Pyciński, B.; Biesok, M.; Rudzki, M.; Majewski, J.; Kostecki, J.; Pietka, E. Chronic wounds multimodal image database. *Comput. Med. Imaging Graph.* **2021**, *88*, 101844. [CrossRef] [PubMed]
- Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [CrossRef]
- Mahbod, A.; Schaefer, G.; Ecker, R.; Ellinger, I. Automatic Foot Ulcer Segmentation Using an Ensemble of Convolutional Neural Networks. In Proceedings of the 26th International Conference on Pattern Recognition, Montreal, QC, Canada, 21–25 August 2022; pp. 4358–4364. [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. [CrossRef]
- Chaurasia, A.; Culurciello, E. LinkNet: Exploiting encoder representations for efficient semantic segmentation. In Proceedings of the IEEE Visual Communications and Image Processing, St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4. [CrossRef]
- Tan, M.; Le, Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Chaudhuri, K., Salakhutdinov, R., Eds.; Proceedings of Machine Learning Research PMLR, PMLR; Volume 97, pp. 6105–6114.
- Mahbod, A.; Schaefer, G.; Ecker, R.; Ellinger, I. Pollen grain microscopic image classification using an ensemble of fine-tuned deep convolutional neural networks. In Proceedings of the International Conference on Pattern Recognition, Virtual, 10–15 January 2021; Springer: Berlin/Heidelberg, Germany, 2021; pp. 344–356. [CrossRef]
- Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
- Zhang, Y.; Liu, S.; Li, C.; Wang, J. Rethinking the Dice Loss for Deep Learning Lesion Segmentation in Medical Images. *J. Shanghai Jiaotong Univ. Sci.* **2021**, *26*, 93–102. [CrossRef]
- Moshkov, N.; Mathe, B.; Kertesz-Farkas, A.; Hollandi, R.; Horvath, P. Test-time augmentation for deep learning-based cell segmentation on microscopy images. *Sci. Rep.* **2020**, *10*, 5068. [CrossRef] [PubMed]
- Tsiknakis, N.; Savvidaki, E.; Manikis, G.C.; Gotsiou, P.; Remoundou, I.; Marias, K.; Alissandrakis, E.; Vidakis, N. Pollen Grain Classification Based on Ensemble Transfer Learning on the Cretan Pollen Dataset. *Plants* **2022**, *11*, 919. [CrossRef] [PubMed]
- Gaillochet, M.; Desrosiers, C.; Lombaert, H. TAAL: Test-Time Augmentation for Active Learning in Medical Image Segmentation. In Proceedings of the Data Augmentation, Labelling, and Imperfections, Singapore, 22 September 2022; Nguyen, H.V., Huang, S.X., Xue, Y., Eds.; Springer: Cham, Switzerland, 2022; pp. 43–53. [CrossRef]
- Wang, C.; Anisuzzaman, D.M.; Williamson, V.; Dhar, M.K.; Rostami, B.; Niezgodna, J.; Gopalakrishnan, S.; Yu, Z. Fully automatic wound segmentation with deep convolutional neural networks. *Sci. Rep.* **2020**, *10*, 21897. [CrossRef] [PubMed]
- Huang, C.H.; Wu, H.Y.; Lin, Y.L. HarDNet-MSEG: A simple encoder–decoder polyp segmentation neural network that achieves over 0.9 mean dice and 86 fps. *arXiv* **2021**, arXiv:2101.07172.

21. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [[CrossRef](#)]
22. Wu, Z.; Su, L.; Huang, Q. Cascaded Partial Decoder for Fast and Accurate Saliency Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
23. Liu, S.; Huang, D.; Wang, A. Receptive Field Block Net for Accurate and Fast Object Detection. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
24. Galdran, A.; Carneiro, G.; Ballester, M.A.G. Double encoder–decoder Networks for Gastrointestinal Polyp Segmentation. In Proceedings of the International Conference on Pattern Recognition, Virtual, 10–15 January 2021; pp. 293–307. [[CrossRef](#)]
25. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
26. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5987–5995.
27. Bartko, J.J. Measurement and Reliability: Statistical Thinking Considerations. *Schizophr. Bull.* **1991**, *17*, 483–489. [[CrossRef](#)] [[PubMed](#)]
28. Boehringer, A.S.; Sanaat, A.; Arabi, H.; Zaidi, H. An active learning approach to train a deep learning algorithm for tumor segmentation from brain MR images. *Insights Imaging* **2023**, *14*, 141. [[CrossRef](#)] [[PubMed](#)]
29. Zijdenbos, A.; Dawant, B.; Margolin, R.; Palmer, A. Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE Trans. Med. Imaging* **1994**, *13*, 716–724. [[CrossRef](#)] [[PubMed](#)]
30. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [[CrossRef](#)] [[PubMed](#)]
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
32. Azad, R.; Kazerouni, A.; Heidari, M.; Aghdam, E.K.; Molaie, A.; Jia, Y.; Jose, A.; Roy, R.; Merhof, D. Advances in medical image analysis with vision Transformers: A comprehensive review. *Med. Image Anal.* **2024**, *91*, 103000. [[CrossRef](#)] [[PubMed](#)]
33. He, K.; Gan, C.; Li, Z.; Reik, I.; Yin, Z.; Ji, W.; Gao, Y.; Wang, Q.; Zhang, J.; Shen, D. Transformers in medical image analysis. *Intell. Med.* **2023**, *3*, 59–78. [[CrossRef](#)]
34. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
35. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jegou, H. Training data-efficient image transformers & distillation through attention. In Proceedings of the 38th International Conference on Machine Learning, Virtual, 18–24 July 2021; Meila, M., Zhang, T., Eds.; Proceedings of Machine Learning Research PMLR, PMLR; Volume 139, pp. 10347–10357.
36. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.; et al. Rethinking Semantic Segmentation From a Sequence-to-Sequence Perspective With Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 6881–6890.
37. Mahbod, A.; Dorffner, G.; Ellinger, I.; Woitek, R.; Hatamikia, S. Improving generalization capability of deep learning-based nuclei instance segmentation by non-deterministic train time and deterministic test time stain normalization. *Comput. Struct. Biotechnol. J.* **2024**, *23*, 669–678. [[CrossRef](#)] [[PubMed](#)]
38. Wen, Y.; Tran, D.; Ba, J. BatchEnsemble: An alternative approach to efficient ensemble and lifelong learning. *arXiv* **2020**, arXiv:2002.06715.

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.