IMPROVING LIMITED SUPERVISED FOOT ULCER SEGMENTATION USING CROSS-DOMAIN AUGMENTATION STRATEGIES

Shang-Jui Kuo[†,⋆]   *Po-Han Huang*[†,⋆]   *Chia-Ching Lin*[†]   *Jeng-Lin Li*[†]   *Ming-Ching Chang*[‡]

[†] Inventec Corporation
[‡] University at Albany, State University of New York

{kuo.raysj, huang.po-han, lin.jacob, li.johncl}@inventec.com, mchang2@albany.edu

ABSTRACT

Diabetic foot ulcers pose health risks, including higher morbidity, mortality, and amputation rates. Monitoring wound areas is crucial for proper care, but manual segmentation is subjective due to complex wound features and background variation. Expert annotations are costly and time-intensive, thus hampering large dataset creation. Existing segmentation models relying on extensive annotations are impractical in real-world scenarios with limited annotated data. In this paper, we propose a cross-domain augmentation method named *TransMix* that combines *Augmented Global Pre-training* (*AGP*) and *Localized CutMix Fine-tuning* (*LCF*) to enrich wound segmentation data for model learning. *TransMix* can effectively improve the foot ulcer segmentation model training by leveraging other dermatology datasets not on ulcer skins or wounds. *AGP* effectively increases the overall image variability, while *LCF* increases the diversity of wound regions. Experimental results show that *TransMix* increases the variability of wound regions and substantially improves the Dice score for models trained with only 40 annotated images under various proportions.

Index Terms— foot ulcer segmentation, data augmentation, CutMix, transfer learning, pre-training.

1. INTRODUCTION

Foot ulcers, a common complication associated with diabetes, represent a significant public health concern due to their substantial impact on morbidity, mortality, and the elevated risk of lower limb amputations [1]. Precise delineation of wound areas plays a pivotal role in comprehensive wound management and ongoing healing assessment. Manual skin wound segmentation is a traditionally employed approach that demands a high level of expertise and experience [2]. This requirement arises from the multifaceted etiologies and diverse visual characteristics of wounds. However, the inherent subjectivity and time-intensive nature of manual segmentation procedures pose considerable challenges within clinical practice [3, 4]. Consequently, there exists a pressing need to develop an automatic segmentation model for efficient and accurate ulcer wound analysis.

Research on automatic wound segmentation has advanced greatly with Deep Neural Networks (DNN), notably the convolutional neural network (CNN) and Transformer architectures. Visual object detection and segmentation networks such as the Mask R-CNN [5] have been applied for automatic wound segmentation [6]. The lightweight model in [7] utilizes residual attention to capture local patch details. The recent FUSegNet [3] refines the U-Net [8] architecture by integrating spatial and channel Squeeze-and-Excitation modules during the decoding phase. In addition, ensemble techniques are broadly employed to overcome the limited discriminative ability of the architecture designs [3, 9, 10]. Despite the advancement, these model designs and ensemble strategies focus on fine-tuning network architectures, without directly manipulating the raw image data. These approaches still face limitations when it comes to effectively learning and segmenting diverse types of wounds. Furthermore, over-parameterized models may struggle to accurately capture subtle wound patterns, especially when there is a scarcity of available labeled training data [11].

Previous studies on foot ulcer segmentation have typically neglected real-world scenarios, characterized by a scarcity of labeled training data due to the time-intensive nature of the data collection and annotation processes. In addressing the challenge of data scarcity, cross-domain learning approaches are developed in medical image analysis. Most research focuses on learning from different sources of data that belong to the same set of labeled classes. The domain gap arising from differences in imaging sites for MRI data can be eliminated this way [12]. Other transfer learning approaches extend beyond these domain gaps to encompass different label types. For instance, the diabetic retinopathy method is expanded in [13] to recognize multiple ocular diseases, where their glioma identification model can be used to diagnose other brain tumors. In automatic wound segmentation, it is impractical to collect a sufficiently large amount of images with annotations that cover the required variability for DNN model training. Although the transfer learning

---

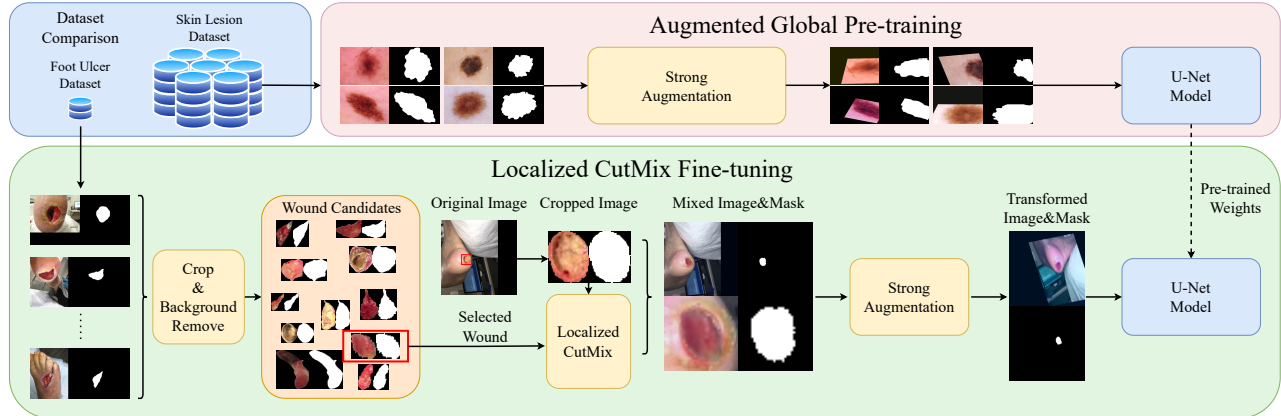⋆: The first two authors equally contribute to this work.

Fig. 1.   **Overview of the proposed *TransMix* framework.** *Augmented Global Pre-training* (*AGP*) enables initial model learning from other larger skin datasets. *Localized CutMix Fine-tuning* (*LCF*) expands the diversity of wound patches via CutMix synthesis. Our approach can effectively overcome the issue of data scarcity during model training.

across classes can partially work for wound segmentation, the application-dependent domain gap remains the bottleneck for effectiveness. The main difficulty in developing an automatic foot ulcer segmentation model is still the scarcity of suitable wound image data. To this end, developing effective methods that can learn from limited labeled data is imperative.

In this work, we propose a cross-domain wound segmentation method based on data augmentation via image transformation and local patch mixing, named *TransMix*; see Fig. 1. *TransMix* consists of two model training stages: *Augmented Global Pre-training* (*AGP*) and *Localized CutMix Fine-tuning* (*LCF*). To verify the effectiveness of learning from limited labels, our model is first trained on the large skin lesion dataset HAM10000 [14, 15] from Harvard Dataverse. The model with learned global knowledge regarding skin segmentation is then fine-tuned on a smaller foot ulcer dataset, where the augmented wound regions with diversified wound appearances can be effectively learned. *AGP* enables the learning of global knowledge such as camera viewpoint variations. In *LCF*, a more local and refined augmentation using CutMix [16] is adopted to expand the variability of the wound regions. Instead of randomly replacing patches in the image, we carefully specify the wound region for fine-grained augmentation. The proposed *AGP* and *LCF* pipeline work jointly to narrow down domain gaps across datasets, enabling our method to robustly transfer the learned knowledge in the scenario of limited annotation data. We evaluate our proposed framework on the FUSeg dataset [17] and achieve an 85.26% Dice score using only 40 labeled images, which achieved a 10.43% improvement compared to the baseline method.

2. METHOD

We focus on the wound segmentation task with limited training data. We consider transfer learning from a *source* domain with a larger dataset to a *target* domain with limited data samples. Specifically, for model pre-training in the source domain, let the source skin dataset D_S consist of image-mask

pairs $\{\mathbf{x}_i^S, \mathbf{m}_i^S\}_{i=1}^M$, where the size M of the source dataset D_S is large enough. Let the target foot ulcer wound dataset D_T consist of image-mask pairs $\{\mathbf{x}_i, \mathbf{m}_i\}_{i=1}^N$, where $\mathbf{m}_i$ denotes the $\mathbb{R}^{W \times H}$ segmentation ground truth of the corresponding image $\mathbf{x}_i$ with the same dimension. We assume the size N of the target dataset to be small (*e.g.*, $N \in \{40, 81\}$). The goal is to derive a model that can accurately segment the wound region from the test images in the domain of D_T.

Dataset.   We use a real-world foot ulcer wound segmentation benchmark, FUSeg [17] as the dataset for the target domain. FUSeg contains 810 foot ulcer images along with their ground-truth masks as a training set and 200 images as a validation set. To simulate the data scarcity scenarios for evaluation purposes, we limit the access to only 40 and 81 randomly selected training samples (which corresponds to 5% and 10% of the original training set, respectively). For the source domain, we utilize the large-scale HAM10000 dataset [14] Harvard Dataverse, as the source dataset. HAM10000 consists of 10,015 skin lesion images along with ground-truth masks.

2.1. Cross-Domain Transfer Strategies

Our cross-domain wound segmentation framework in Fig. 1 consists of two training stages: *Augmented Global Pre-training* (*AGP*) and *Localized CutMix Fine-tuning* (*LCF*). The focus of *AGP* is to extract global knowledge of skin lesion segmentation from the larger source-domain data. *LCF* then perform location-aware CutMix with background removal to enhance wound variability to fine-tune the wound segmentation model on the limited available target-domain samples. Both *AGP* and *LCF* perform image data augmentations, from coarse to fine, with different purposes: *AGP* focuses on global viewpoint variations, while *LCF* attends to local wound patch variabilities.

2.2. Augmented Global Pre-training

To mitigating the large discrepancy between the source and target domains, effective data augmentation is necessary for

Table 1. Evaluation of *Augmented Global Pre-training* (*AGP*) and *Localized CutMix Fine-tuning* (*LCF*) on the FUSeg validation dataset in terms of precision, recall, and Dice scores in %.

Experiments Setting #	Pre-trained	Method	40 Samples				81 Samples			
			Precision	Recall	IoU	Dice	Precision	Recall	IoU	Dice
A	-	U-Net	67.99	83.18	59.78	74.83	**86.98**	73.24	66.00	79.52
B	✓	+ Pre-training	84.06	55.92	50.56	67.16	82.98	71.16	62.09	76.61
C	✓	+ AGP	84.99	79.60	69.79	82.21	84.87	86.50	74.95	85.68
D	-	+ CutMix	80.16	75.16	63.37	77.58	76.68	**90.53**	70.98	83.03
E	-	+ CutMix + Background Removal	75.11	77.85	61.89	76.46	84.18	87.29	74.99	85.71
F	-	+ CutMix + Location-aware Pasting	85.57	72.38	64.51	78.43	85.98	86.26	75.72	86.12
G	-	+ LCF	**86.75**	82.39	73.18	84.51	84.79	87.53	75.65	86.14
H	✓	+ TransMix	82.50	**88.22**	**74.31**	**85.26**	84.83	88.51	**76.44**	**86.65**

knowledge transfer during model training. As shown in the snapshots in Fig. 1, HAM10000 contains higher-quality images taken with consistent camera view angle and distance, illumination conditions and backgrounds. The skin lesion regions are mostly circular in shape, occupying a large portion at the center. In contrast, many images in the FUSeg dataset are taken with various conditions and qualities, with arbitrary wound locations, shapes and sizes. To this end, we perform intense spatial-level transformations including global perspective transformations and optical distortion. We also perform strong pixel-level transformations including RGB & HSV shifts, brightness & contrast, to simulate illumination changes. A U-Net [8] model is pre-trained to learn the segmentation of skin regions, which will be transferred to the foot ulcer domain for wound segmentation.

2.3. Localized CutMix Fine-tuning

With the pre-trained model using the large-scale HAM10000 dataset, we perform fine-tuning to the FUSeg dataset for wound segmentation. In a situation with very limited training data, the model can hardly capture diverse types of wounds. Therefore, our proposed *TransMix* approach includes a CutMix strategy to diversify the visual appearances of the wounds during the fine-tuning steps.

We first obtain the bounding boxes' position of the wounds in all the training images using the corresponding annotated masks and then crop the bounding boxes along with their masks as the wound candidates. The background of each wound candidate is removed based on the corresponding mask to synthesize a more realistic wound appearance without an obvious artificially generated cropping boundary.

In each batch of data, we specify the position of the wound in target images and randomly paste other wound candidates on the specified region. As the wound size can be dramatically different across images, we deliberately select the wound candidates with similar wound sizes to the target wound region and resize the wound candidates to match the target wound size. When we paste a wound candidate to a specified region, part of the wounds in the target image remains, and other parts of the pasted wound candidate might

exceed the specified boundary due to the different shapes. We regard the union regions of the remained wounds and the pasted wound candidates to generate a new segmentation mask for the newly synthesized image. Through multiple iterations in the fine-tuning phase, the same wound region is augmented with multiple wound candidates which enhances the robustness of segmentation model learning against various wound appearances. Aside from the localized wound CutMix, we also perform the strong whole-image transformation mentioned in §2.2 to augment the training data perspectives.

3. EXPERIMENTAL RESULTS

We set up two experiments to evaluate our proposed framework in terms of the comparison to other segmentation approaches and the analysis of different numbers of accessible training data. The results are reported in §3.1 and §3.2, respectively. We include the compared models and implementation details in the following sections.

Models for comparison. The settings of the compared models for §3.1 are described below and correspond to the results in Table 1. We adopt *U-Net* (Setting A) as a strong baseline model which has been widely used for automated wound segmentation [18]. Notably, FUSegNet [3] using an EfficientNet-B7 [19] backbone with ensemble techniques, achieved a state-of-the-art Dice score (92.70%) on the FUSeg dataset. We kept the architecture but avoided such a cumbersome backbone model for the data-scarce scenario while still achieving a comparable Dice score (91.08%) on the full set evaluation with the ResNet-50 backbone.

Pre-training (Setting B) is performed using the HAM10000 dataset with naive fine-tuning. This baseline is a common practice in the medical field in dealing with limited training data [20] although most past studies are limited to using segmentation datasets that have the same set of labeled classes. *AGP* (Setting C) utilizes our proposed pre-training approach (§2.2) using naive fine-tuning.

We also compare model training approaches without pre-training as follows. CutMix (Setting D) is applied based on the original paper [16]. We further remove the background of the wound candidates and specify the wound location in
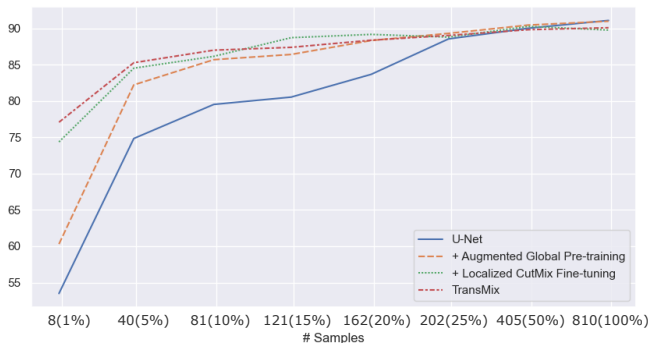
Fig. 2. We evaluate the model fine-tuned using different numbers of FUSeg training samples, and report the Dice scores (%) on the FUSeg validation set.

Setting E and Setting F. *LCF* (Setting G) is our proposed approach in §2.3 without pre-training.

Ultimately, our proposed *TransMix* (Setting H) combines both *AGP* and *LCF*.

Implementation Details. We use the dataset described at §2 and U-Net with ResNet-50 encoder as our backbone. We used Adam optimizer with 5e-4 learning rate, batch size of 4, and training with 300 epochs, the learning rate is decayed by a factor of 10 every 100 epochs. We perform data augmentation at two levels using the widely-used toolbox of Albumentations [21]. At the spatial level, we use several augmentations including grid distortion, optical distortion, perspective transform, and affine transform to align the whole image angle across the dataset. At the pixel level, we applied some transformations such as RGB shift, brightness contrast, and HUE saturation value transformations to align illumination conditions.

3.1. Results Under Limited Supervision

We compare our proposed *TransMix* method against other segmentation approaches under limited supervision. Table 1 lists the performance comparisons on the FUSeg validation set using only 40 or 81 training samples. As can be seen, our framework (*TransMix* in Setting H) consistently demonstrates significantly superior performance compared to baseline methods. Notably, our framework achieves Dice scores of 85.26% and 86.65%, respectively, showing 10.43% and 7.13% improvements compared to the baseline method (*U-Net* in Setting A).

Effects of the Augmented Global Pre-training. As mentioned in §2.2, it is essential to eliminate the domain gap between the source and target domains for effective knowledge transferring. As can be seen from Setting B in Table 1, simple pre-training on HAM10000 for transfer learning could result in significant performance degradation as compared to the baseline (Setting A). In contrast, the model pre-trained by the proposed *Augmented Global Pre-training* strategy (Setting C)

achieves Dice scores of 82.21% and 85.68% after fine-tuning on 40 and 81 FUSeg training samples, respectively, showing 7.38% and 6.16% improvements compared to the baseline. This demonstrates the necessity of our strong augmentation strategy in eliminating the domain gap to better benefit from the pre-training process, especially in data-scarce scenarios.

Effects of the Localized CutMix Fine-tuning. As mentioned in §2.3, it is crucial to synthesize diverse wound samples with reasonable locations and realistic appearances for useful data augmentation. We conduct an ablation study to demonstrate the effectiveness of the proposed *LCF* in fulfilling this goal. Settings D-G in Table 1 list the performance comparisons under different CutMix strategies during fine-tuning. We employ the original CutMix [16] technique (Setting D), and further remove the background of wound candidates to be pasted (Setting E). Both of these settings show improvements compared to the baseline (Setting A). However, since the position for pasting is randomly selected, these settings may result in unrealistic synthesized images. Furthermore, we paste wound candidates based on the wound locations in the target image without performing background removal (Setting F) to verify the effectiveness of location-aware pasting. This approach resulted in an improvement compared to Setting D, highlighting the importance of a reasonable pasting position. Finally, we simultaneously employ background removal and location-aware pasting (Setting G), which achieves further improvements of 6.93% and 3.11% in Dice scores, respectively.

3.2. Analysis on the Amount Training Data

In Fig. 2, we examine the change of Dice score with varying data volumes, using 40 (5%), 81 (10%), 202 (25%), 410 (50%), and 810 (100%) images to simulate the data accessibility in the real world. Our proposed *TransMix* brings about large benefits when we only obtain less than 200 wound samples. The trend shown in Fig. 2 suggests the benefits of *TransMix*, while the margin becomes minor if the target dataset is large enough. With very limited labeled data (e.g., $N < 40$), the model performance could highly depend on the quality of the sampled data.

4. CONCLUSION

Diabetic foot ulcers patients require continual wound monitoring to manage their recovery conditions. However, current automatic segmentation approaches assume the accessibility of large-scale labeled data, which is impractical in clinical practice. To our best knowledge, this work is the first study addressing the limited supervised wound segmentation problem. The novel cross-domain framework enables the use of other datasets with skin images for global knowledge transfer and augments wound variability in a localized manner.

Future work: we will apply few-shot and zero-shot learning to enhance model usability in the real world.

5. REFERENCES

[1] Rebecca Sorber and Christopher J Abularrage, "Diabetic foot ulcers: Epidemiology and the role of multidisciplinary care teams," in *Seminars in vascular surgery*. Elsevier, 2021.

[2] Xiaohui Liu, Changjian Wang, Fangzhao Li, Xiang Zhao, En Zhu, and Yuxing Peng, "A framework of wound segmentation based on deep convolutional networks," in *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 2017.

[3] Mrinal Kanti Dhar, Taiyu Zhang, Yash Patel, and Zeyun Yu, "FUSegNet: A deep convolutional neural network for foot ulcer segmentation," *arXiv preprint arXiv:2305.02961*, 2023.

[4] Dhanesh Ramachandram, Jose Luis Ramirez-GarciaLuna, Robert DJ Fraser, Mario Aurelio Martínez-Jiménez, Jesus E Arriaga-Caballero, Justin Allport, et al., "Fully automated wound tissue segmentation using deep learning on mobile devices: cohort study," *JMIR mHealth and uHealth*, 2022.

[5] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask R-CNN," in *IEEE/CVF International Conference on Computer Vision*, 2017.

[6] PL Muñoz, R Rodríguez, and N Montalvo, "Automatic segmentation of diabetic foot ulcer from mask region-based convolutional neural networks," *Journal of Biomedical Research and Clinical Investigation*, 2020.

[7] Shahzad Ali, Arif Mahmood, and Soon Ki Jung, "Lightweight encoder-decoder architecture for foot ulcer segmentation," in *International Workshop on Frontiers of Computer Vision*. Springer, 2022.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer Assisted Intervention*. Springer, 2015.

[9] Amirreza Mahbod, Gerald Schaefer, Rupert Ecker, and Isabella Ellinger, "Automatic foot ulcer segmentation using an ensemble of convolutional neural networks," in *International Conference on Production Research*, 2022.

[10] Pin Xu, Xin Wu, Yanyi Li, Ejaz Ul Haq, Jianping Yin, and Kuan Li, "Ensemble learning for diabetic foot ulcer segmentation based on dfuc2022 dataset," in *International Conference on Electronic Information Technology and Computer Engineering*, 2022.

[11] Xiaohui Liu, Changjian Wang, Fangzhao Li, Xiang Zhao, En Zhu, and Yuxing Peng, "A framework of wound segmentation based on deep convolutional networks," in *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, 2017.

[12] Mathilde Bateson, Hoel Kervadec, Jose Dolz, Hervé Lombaert, and Ismail Ben Ayed, "Source-free domain adaptation for image segmentation," *Medical Image Analysis*, 2022.

[13] Yi Zhou, Lei Huang, Tao Zhou, and Ling Shao, "Cct-net: category-invariant cross-domain transfer for medical single-to-multiple disease diagnosis," in *IEEE/CVF International Conference on Computer Vision*, 2021.

[14] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific data*, 2018.

[15] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, et al., "Human–computer collaboration for skin cancer recognition," *Nature Medicine*, 2020.

[16] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *IEEE/CVF International Conference on Computer Vision*, 2019.

[17] Chuanbo Wang, Amirreza Mahbod, Isabella Ellinger, Adrian Galdran, Sandeep Gopalakrishnan, Jeffrey Niezgoda, and Zeyun Yu, "FUSeg: The foot ulcer segmentation challenge," in *arXiv 2201.00414*, 2022.

[18] Gaetano Scebba, Jia Zhang, Sabrina Catanzaro, Carina Mihai, Oliver Distler, Martin Berli, and Walter Karlen, "Detect-and-segment: A deep learning approach to automate wound image segmentation," *Informatics in Medicine Unlocked*, 2022.

[19] Mingxing Tan and Quoc Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International Conference on Machine Learning*. PMLR, 2019.

[20] Jialin Peng and Ye Wang, "Medical image segmentation with limited supervision: a review of deep network models," *IEEE Access*, 2021.

[21] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, 2020.