Refined mixup augmentation for diabetic foot ulcer segmentation

David Jozef Hresko[1], Jakub Vereb[1], Valentin Krigovsky[1], Michala Gayova[2], and Peter Drotar[1]

[1] IISlab, Technical University of Kosice, Kosice, Slovakia
[2] Department of Burns and Reconstructive Surgery, PJ Safarik University and Hospital Agel, Kosice-Saca, Slovakia

Abstract. Diabetic foot syndrome is one of the chronic diabetic complications that present itself as foot ulceration, which in many cases leads to limb amputation. Moreover, it is linked to the high percentage of post-amputation mortality within the period of five years. Thus it is crucial to diagnose and plan careful treatment properly in the early stages. Diagnosis is often time-consuming and requires a skilled clinician who can differentiate between similarities between diabetes-related and other types of ulcers by evaluating their morphology and location. To mitigate diagnostics issues, we propose an improved version of nnU-net architecture with residual short skip connections in the encoder part and additional mixup augmentation as the preprocessing step. The obtained results on DFU22 challenge dataset prove that our improvements can boost overall performance for ulcer segmentation tasks, even in scenarios where targeted structures are heterogeneous and under high imbalance conditions in the evaluated dataset. With mentioned approach we achieved 9th place with Dice score equal to 0.6975.

Keywords: nnU-Net · mixup augmentation · skip connections · segmentation · diabetic ulcer

1 Introduction

Diabetes is a metabolic disease characterized by high blood sugar levels over a prolonged period of time. Diabetes is already considered a pandemic disease, and with time, its prevalence is expected only to increase (from 536,6 million in 2021 to 783,2 million in 2045) [9]. Even though the diagnosis is straightforward, it frequently happens late and already in the stage of chronic complications [7]. Diabetic foot syndrome (DFS) is one of chronic diabetic complications that presents as foot ulceration. Despite intensive treatment, the affected limb is amputated within 6-18 months after the first evaluation in up to a quarter of patients, while the 5-year post-amputation mortality is as high as 50% [4]. On the other side, early diagnosis and careful treatment can prevent limb amputation and ulcer-related mortality. Therefore, careful treatment must not be delayed due to misdiagnosis, which often happens due to similarities among diabetes-related and other cause-related ulcers (ischemic, venous, or pressure ulcers). To a certain level of confidence, a skilled clinician can differentiate between them by evaluating their morphology and location. However, patients and caretakers cannot do so, and even specialists need to employ advanced imaging methods to confirm the diagnosis.

In recent years, convolutional neural networks (CNN) penetrated almost every aspect of medical imaging. Whether it is classification, segmentation, or registration of medical images, CNN offers performance close to the human specialist. Specifically for medical image segmentation, CNN with encoder-decoder architecture, named U-net [8], was proposed and became a de facto standard solution for medical image semantical segmentation. There were several attempts to improve the U-net architecture to boost the network performance for particular task [10, 13]. However, an interesting observation was published in [5], claiming that the method configurations, such as pre-processing, have better potential to improve the network's performance than architectural modifications of the U-net. The so-called not new U-net (nnU-net) dominated in 23 different segmentation challenges.

In this paper, we utilize nnU-net architecture and propose its improved version with residual short skip connections [1], and additional mixup augmentation [12], which was specifically tailored for the foot ulcer segmentation task. The fundamental goal was to demonstrate that this slight architectural modification can further boost overall performance, even for the heavily imbalanced dataset with a wide variety of targeted structures.

The rest of the paper is organized as follows. In the next section we provide detailed data description. In the methodology section, the proposed solution is described. Finally, we present the results and discuss different aspects of our submission.

2 Data

The data used during our experiments were part of the official DFU22 challenge dataset gathered by the organizers themselves [6]. They used three different cameras to capture the data. Kodak DX4530, Nikon D3300, and Nikon COOLPIX P100 were used to acquire close-ups of the whole foot at a distance of around 30–40 cm with the parallel orientation to the plane of an ulcer. Adequate room lights were used, instead of a camera flash, to get consistent colors in the images. Additionally, to improve the performance and reduce computational costs of proposed models, the original size of images, which varied between 1600×1200 and 3648×2736 pixels, was modified to 640x480 with the preservation of the aspect ratio of the ulcers. Moreover, out-of-focus photographs and blurry ones were discarded to prevent misleading assumptions.

The ground-truth corresponds to the area of ulcer and ulcer periwounds and was produced by a podiatrist and a consultant physician specializing in the diabetic foot with more than five years of professional experience. In the case of disagreement, a third specialist podiatrist examined the photograph. Finally, the whole annotation process was completed using the polygonal shape tool within the VGG Image Annotator (VIA) application [2].

Altogether 3000 cases were collected for the training phase, where 1500 of these cases remained unlabeled. Another 500 labeled cases were used for the validation phase, and the remaining 3000 labeled cases were used for the testing phase of the challenge.

3 Methodology

Based on the previous successful applications of nnU-net architecture on various challenges, we chose its baseline version as the starting point for our experiments to segment ulcer structures. Authors of [1] introduced short skip connections as a solution for training very deep networks, which suffers from vanishing gradient problem and slow convergence. They also proved that these short skip connections could help the model to learn a better representation of targeted structures, thus improving overall performance. Based on these results, we decided to enhance the encoder part of the nnU-net with this implementation of skip connections. Furthermore, we applied mixup augmentation on the network's input layer to further boost the overall performance. This technique was already shown to be beneficial for segmentation of 3D CT scans of abdominal area [3].

3.1 Preprocessing

The original DFU 2022 challenge dataset consists of 2D photos. However, the standard nnU-Net pipeline can only work with 3D NIfTI data, so our first step was to transform the original images and masks to the required format. Here we applied two different approaches , one for the image data and different for masks. In case of foot photos we added two additional dimensions, where the first one represented color channel of the image and the second one was used to create dummy 3D image with value one along x-axis. Due to network requirements we additionally reordered the dimensions to be in the channel first format. The slightly different process was applied on masks. By default, the nnU-net treats all pixels with value of 0 as background and everything with higher value as label. In order to fulfil this requirement we applied our custom thresholding method. Every pixel with value below 255 was considered as background and the remaining pixels represented label.

Afterward, we performed standard preprocessing steps such as normalization, scaling, and other spatial transformations. In the case of the nnU-Net, these steps were automatically handled by nnU-Net processing. Concretely, the image resampling strategy was third order spline interpolation for in-plane and nearest neighbors for out-of-plane. Global percentile clipping along with z-score with global foreground mean and standard deviation was chosen to normalize data. To clip HU values, nnU-Net pipeline applied the default settings in the 0.5 to 99.5 percentile range.

3.2 Proposed approach

We employ U-Net architecture in all experiments. As the first modification, we implemented short skip connections in the encoder part to speed up the convergence process and boost the information flow through the individual intermediate layers. Primary, the encoder part of the nnU-net is responsible for feature extraction and structure recognition. Thus the application of these additional skip connections in the decoder part is not necessary and would not contribute to the overall network performance.

As the second improvement, we propose utilization of the mixup augmentation. According to authors of [12], input images with ground truth labels denoted as (x_i, y_i), (x_j, y_j) are transformed to an output tuples $(\hat{x}_i, \hat{y}_i)$ by the following transformations:

$$\hat{x} = \lambda x_i + (1 - \lambda) x_j \tag{1}$$

$$\hat{y} = \lambda y_i + (1 - \lambda) y_j \tag{2}$$

We hypothesize that this modification can prevent the issues related to imbalanced datasets, where the network has a tendency to segment the non-background portion of the image as background, thus ignoring targeted structures. We performed multiple experiments with regular mixup and its modified version named CutMix [11]. The key difference between these augmentations is that the regular mixup is applied on the whole mini-batch with a specified mixup ratio, while the CutMix version utilizes only cropped part of the image, which is later mixed with another non-cropped image. The illustration of the mixup augmentations on original photos can be seen in Fig. 1.

(a) CutMix (b) Regular mixup

Fig. 1: Illustration of mixup augmentations.

Based on the obtained results, we decided to incorporate standard mixup augmentation with a mixup ratio 50:50. This ratio was used for foot photos and also for labels. The mixup process for photos was straightforward, but not for the labels. Here we must have handled three different scenarios, based on the different meanings of pixel values. Pixels marked as background were only treated as background when they were mixed with background values from second label. In the remaining two scenarios the pixels after mixup were always considered as label. So, in the end the weakest value the label could have after mixing process was 127, therefore there was no risk of having too weak masks that are more similar to background than to label and being marked as wound instead of background. The overview of proposed method can be seen in Fig. 2.

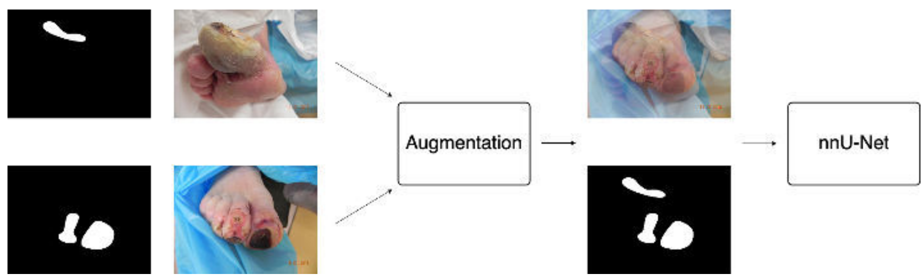

Fig. 2: Overview of proposed method

4 Experimental results

To measure the overall performance of proposed method and to evaluate the similarity between the predicted contours and the ground truth contours produced by the domain experts, Dice Similarity Coefficient (DSC) was selected as a primary statistical method.

The final version of modified nnU-Net was trained with stochastic gradient descent (SGD) optimizer, where the initial learning rate parameter was set to 0.01. We trained this model with the batch size of two for 1000 epochs, which is the default length of the training for nnU-Net pipeline. To minimize loss we used combined Dice and Cross Entropy loss function. Here we did not further modify the short skip connections implementation and kept it as described in chapter 3.2. In the case of the mixup augmentation, we set the value of the mixing hyperparameter to $\alpha = 0.5$. The measured results from the training phase and comparisons to all performed experiments are presented in Table 1.

We also noticed that individual implementation of short skip connections and mixup augmentation slightly enhanced segmentation quality. Based on this observation, we proposed a solution that is based on the combination of mixup and short skip connections. As a result, the proposed neural network significantly improved the overall performance of the nnU-Net and outperformed its baseline version. On the other hand, CutMix augmentation's utilization proved insufficient and performed even worse than baseline nnU-Net. We assume that

Network	Dice Score
Baseline nnU-Net	0.6530
nnU-Net + mixup + skip connections (final solution)	**0.6713**
nnU-Net + skip connections	0.6565
nnU-Net + mixup	0.6616
nnU-Net + CutMix	0.6365

Table 1: Experimental results during training phase on DFU22 public data

this degradation was caused by extensive modification of the input data, which negatively impacted the learning process and thus resulted in misleading segmentations.

To demonstrate the capability of trained model, we randomly chose case 100630 from publicly available images. Here, we were capable to precisely segment ulcer structure with Dice score equal to 0.9690. This example segmentation depicted in figure 3, which shows a diabetic foot ulcer in a common position. The picture was taken from above of the well-cleaned ulcer in good lighting conditions. The Dice score is high even though the ulcer was not healing at the time and despite the fact that the picture was taken before the necrectomy (removal of the dead tissue) was performed.

(a) Segmented photo

(b) Original photo

Fig. 3: Example segmentation for the case 100630. Green color denotes ground truth label, red our segmentation and orange overlap of these two labels.

Additionally, our model was evaluated on public test data during testing phase of DFU22 challenge, where we achieved 9th place with Dice score equal to 0.6975, which was not significantly different from 1st place with Dice score equal to 0.7287.

4.1 Fault-case analysis

In order to fully understand the capabilities of our proposed method, we have also performed fault-case analysis with a skilled clinician, who identified several sources of error in segmentations. Firstly, inappropriate wound toilet, which led to capturing blood stains in the proximity of the ground truth wound as illustrated in Fig. 4a. Second, Fig. 4b illustrates atypical angles from which the pictures were taken. These resulted in wound morphology deformation. Third, an incorrect ground truth segmentation (some wounds are partially or entirely left out from the ground truth (Fig. 4c). On the other hand, there are also opposite examples, where the ground truth includes healthy tissue as in Fig. 4d.

(a) Case 100659

(b) Case 101971

(c) Case 100215

(d) Case 100835

Fig. 4: Examples of faulty segmentations. Segmented images are located on the left side and original images on the right side. Green color denotes ground truth label, red our segmentation and orange overlap of these two labels.

5 Conclusions

In this paper, we proposed nnU-net-based fully convolutional neural network for diabetic foot ulcer segmentation. The proposed approach minimizes the probability of over-fitting by using mixup augmentation. Additionally, we implemented a short skip connection into nnU-net architecture to improve training. The proposed solution achieved 9th place with Dice segmentation score equal to 0.6975 on the test dataset of DFU2022 challenge.

Acknowledgements

This work was supported by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences under contract VEGA 1/0327/20.

References

1. Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C.: The importance of skip connections in biomedical image segmentation. In: Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R.S., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z., Cardoso, J.S., Cornebise, J. (eds.) Deep Learning and Data Labeling for Medical Applications. pp. 179–187. Springer International Publishing, Cham (2016)
2. Dutta, A., Zisserman, A.: The via annotation software for images, audio and video. Proceedings of the 27th ACM International Conference on Multimedia (2019)
3. Gazda, M., Bugata, P., Gazda, J., Hubacek, D., Hresko, D.J., Drotar, P.: Mixup augmentation for kidney and kidney tumor segmentation. In: Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Papanikolopoulos, N., Weight, C. (eds.) Kidney and Kidney Tumor Segmentation. pp. 90–97. Springer International Publishing, Cham (2022)
4. Guo, J., Dardik, A., Fang, K., Huang, R., Gu, Y.: Meta-analysis on the treatment of diabetic foot ulcers with autologous stem cells. Stem cell research & therapy **8**(1), 1–8 (2017)
5. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (Feb 2021). https://doi.org/10.1038/s41592-020-01008-z, https://doi.org/10.1038/s41592-020-01008-z
6. Kendrick, C., Cassidy, B., Pappachan, J.M., O'Shea, C., Fernandez, C.J., Chacko, E., Jacob, K., Reeves, N.D., Yap, M.H.: Translating clinical delineation of diabetic foot ulcers into machine interpretable segmentation (2022). https://doi.org/10.48550/ARXIV.2204.11618, https://arxiv.org/abs/2204.11618
7. Roche, M.M., Wang, P.P.: Factors associated with a diabetes diagnosis and late diabetes diagnosis for males and females. Journal of clinical & translational endocrinology **1**(3), 77–84 (2014)
8. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015)
9. Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B.B., Stein, C., Basit, A., Chan, J.C., Mbanya, J.C., et al.: Idf diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes research and clinical practice **183**, 109119 (2022)
10. Tsai, T.H., Huang, S.A.: Refined u-net: A new semantic technique on hand segmentation. Neurocomputing **495**, 1–10 (2022). https://doi.org/https://doi.org/10.1016/j.neucom.2022.04.079, https://www.sciencedirect.com/science/article/pii/S0925231222004696
11. Yun, S., Han, D., Chun, S., Oh, S., Yoo, Y., Choe, J.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 6022–6031. IEEE Computer Society, Los Alamitos, CA, USA (nov 2019). https://doi.org/10.1109/ICCV.2019.00612, https://doi.ieeecomputersociety.org/10.1109/ICCV.2019.00612
12. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization (2017). https://doi.org/10.48550/ARXIV.1710.09412, https://arxiv.org/abs/1710.09412

13. Zhou, J., Lu, Y., Tao, S., Cheng, X., Huang, C.: E-res u-net: An improved u-net model for segmentation of muscle images. Expert Systems with Applications **185**, 115625 (2021). https://doi.org/https://doi.org/10.1016/j.eswa.2021.115625, https://www.sciencedirect.com/science/article/pii/S0957417421010198