Deep Learning Based Wound segmentation using Local-Global Architecture

Atheeth Naik
Machine Learning and Data Analytics Lab (MaD Lab)
Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg
Erlangen, Germany

***Abstract*—Wound image segmentation is important for improving how wounds are assessed and treated in clinical settings. In this project, we implemented the paper WSNet, a model developed specifically for wound segmentation, based on a recent research paper. Our goal was to replicate the model and test how well it works with our data and setup. The model was evaluated on a dataset of wound images, and the experimental results demonstrate significant improvement over traditional methods in terms of accuracy and segmentation quality. Future work could focus on fine-tuning the model and testing it on different types of wounds.**

***Index Terms*—wound segmentation, deep learning, WSNet, medical image analysis, wound assessment**

I. Introduction

Proper wound assessment is essential for tracking healing, planning treatments, and reducing infection risks.Segmenting wounds from images can be challenging because of the wide variety of wound shapes, colors, and the textures around the wound area. A good segmentation model needs to distinguish the wound from healthy skin and other background elements. This is necessary to get accurate wound measurements, which are key for proper assessment, but it requires a model that can handle these complexities.

Recently, WSNet was introduced as a model designed specifically for wound segmentation, making it better suited for this task than general-purpose models. Most earlier approaches either used traditional image processing techniques or general deep learning models, which often had trouble adapting to the specific challenges of wound images. By focusing on WSNet, this study aims to see how well it can handle real-world wound segmentation tasks using global-local architecture.

This work involves implementing the WSNet model and evaluating its effectiveness for wound image segmentation. Our main objective was to test whether the global-local architecture could have better wounds segmentation accuracy. This project gives insights into WSNet's potential for clinical use, and we evaluate its reliability and effectiveness as a wound segmentation tool.

The rest of this report is organized as follows: Section II describes the methods we used, including data preparation, model setup, and training. Section III presents the evaluation metrics and results. Finally, Section IV concludes with a discussion of the model's strengths and potential improvements for future work.

II. Methodology

This section details the steps taken to implement and evaluate various segmentation models and configurations for wound image segmentation. We experimented with different image sizes, architectures, and loss functions to identify the best setup for our task. Below, we describe these experiments and the final configuration selected.

The methodology is organized as follows: Section A discusses Image Resolutions Selection, Section B describes Global and Local Training with Pretrained Weights for Loss Selection, and Section C elaborates on the Final Model Selection and Combined Training.

A. Image Resolutions Selection

To determine the optimal input image resolution for wound segmentation, we experimented with three different image sizes: 256x256, 512x512, and 1024x1024, using a global architecture for the Main wound dataset. The actual image and mask sizes were approximately 4000x4000 pixels and were resized to 256x256, 512x512, and 1024x1024 while preserving aspect ratios. Each resolution was tested using ResNet34 as the backbone for four commonly used architectures in image segmentation: UNet, LinkNet, FPN (Feature Pyramid Network), and PNP. The models were trained using the Adam optimizer with a learning rate of 0.001 for 30 epochs.

Objective: Evaluate model performance across varying resolutions and architectures to identify the best combination.

Results: Using a resolution of 1024x1024 yielded the most accurate and consistent results across all models, likely due to the higher detail and improved boundary definition achievable at this resolution. These experiments demonstrated that the 1024x1024 resolution was the most effective.

B. Global and Local Training with Pretrained Weights for loss selection

Once the optimal resolution (1024x1024) was identified, we further optimized the model by experimenting with two training approaches: global training and local training. The global approach utilized pretrained weights from the WSEG dataset, which had been trained on 256x256 images, and adapted these pretrained weights for segmentation tasks at a higher resolution.

Global Training: In the global training configuration, the entire 1024x1024 image was used for training. Pretrained weights from the WSEG dataset were employed to initialize all parts of the models across different architectures (U-Net, FPN, LinkNet, and PSPNet). These pretrained weights were specifically designed for each architecture—one set for U-Net, one for LinkNet, one for PSPNet, and one for FPN. Each architecture's weights were pretrained on 256x256 images from the WSEG dataset.

When adapting these pretrained weights for 1024x1024 images, the entire set of pretrained weights was used to initialize the models. This adaptation allowed us to retain the pretrained knowledge and feature extraction capabilities without altering the architecture itself. The pretrained weights, which had been fine-tuned for smaller images, were effectively transferred to larger images, enabling the model to process 1024x1024 resolution inputs while still being capable of reconstructing accurate segmentation masks.However, further trainings pretraining was not pursued due to the increased computational complexity involved when training combined models.

Local Training: In the local training configuration, each 1024x1024 image was divided into 16 smaller patches of 256x256 pixels. This partitioning allowed the model to focus on localized areas, which is particularly useful for capturing detailed wound boundaries and other intricate structures. By processing each patch individually, the model was able to better capture fine details that might be overlooked when processing the whole image at once. This localized approach enhanced model performance in identifying smaller wound areas, improving segmentation accuracy.

Training Details: Both global and local training approaches utilized the Adam optimizer with a learning rate of 0.001 for 50 epochs. During training, we tested a learning rate scheduler and incorporated dropout layers to optimize performance further. However, these modifications did not result in significant improvements in the model's performance and were subsequently excluded from the final configurations.

Additionally, three different loss functions were tested during training to optimize model accuracy and improve segmentation performance. Three different loss functions were tested during training to optimize model accuracy:

1. **Recall Cross-Entropy (RCE): Description:** Recall Cross-Entropy (RCE) is a loss function that focuses on maximizing recall by emphasizing the correct prediction of positive samples (foreground). It is particularly useful in tasks with class imbalance, as it penalizes false negatives more heavily than false positives.

Formula:

$$\text{RCE}(p, y) = -\frac{1}{|P|} \sum_{i=1}^{N} y_i \cdot \log(p_i) \tag{1}$$

where:

- p_i is the predicted probability of a pixel being in the positive class (foreground).
- y_i is the ground truth label for pixel i (0 or 1).
- $|P|$ is the total number of positive samples, defined as $|P| = \sum_{i=1}^{N} y_i$.

2. **Focal Loss:**

Description: Focal Loss is designed to address class imbalance by giving more weight to hard-to-classify pixels. In segmentation, this means focusing on areas where the model is making errors, such as the boundaries between classes. It is particularly effective when the positive (foreground) pixels are sparse.

Formula:

$$\text{Focal Loss}(p, y) = -\alpha \cdot (1-p)^{\gamma} \cdot y \cdot \log(p) - \alpha \cdot p^{\gamma} \cdot (1-y) \cdot \log(1-p) \tag{2}$$

where:

- α is a balancing factor for class imbalance.
- γ is a focusing parameter that controls the weight assigned to hard-to-classify pixels (typically $\gamma \geq 1$).
- p and y are as defined above.

3. **Dice Loss: Description:** Dice Loss is based on the Dice Similarity Coefficient (DSC), which measures the overlap between predicted and actual segments. It's well-suited for handling imbalanced data by directly optimizing for the overlap, rather than each individual pixel. Dice Loss is particularly popular for medical and binary segmentation tasks.

Formula:

$$\text{Dice Loss} = 1 - \frac{2 \sum_{i=1}^{N} p_i \cdot y_i}{\sum_{i=1}^{N} p_i + \sum_{i=1}^{N} y_i} \tag{3}$$

where:

- N is the total number of pixels.
- p_i is the predicted probability for pixel i.
- y_i is the ground truth label for pixel i.

Findings: From this experiment, we found that FPN and UNet architectures using focal loss delivered the best results, showing improved segmentation accuracy, particularly for wound boundaries and small details.

C. Final Model Selection and Combined Training

Based on the previous experiments, we selected the Feature Pyramid Network (FPN) as the final model architecture due to its consistent performance. To enhance model robustness and generalization, we applied several data augmentation techniques during training. These included:Horizontal Flipping,Rotation,Optical Distortion,Grid Distortion,Motion Blur. We then refined the training process as follows (see Fig 1 for an overview of the combined model architecture):

Global Training Without Pretraining: The FPN model was trained on the 1024x1024 images without using pretrained weights, allowing it to learn wound-specific features directly from our dataset.

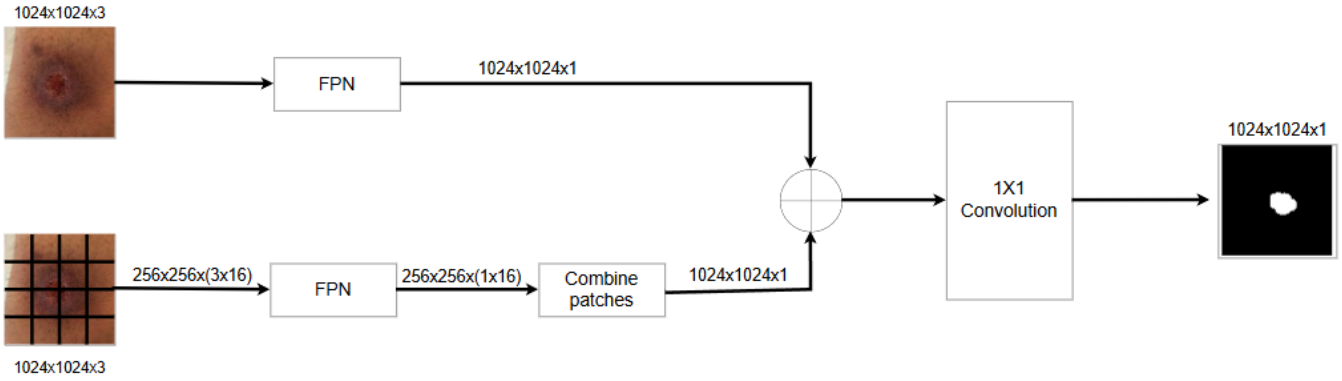

Fig. 1: Architecture of global-local image segmentation method using FPN

Local Training with Patch-Based Approach: Local Training with Patch-Based Approach: In the local training configuration, each 1024x1024 image was divided into 16 smaller patches of 256x256 pixels. This patch-based approach allowed the model to focus on localized areas within the larger image. By processing each 256x256 patch individually, the model could capture finer details and focus on specific features such as wound boundaries, which might be missed when processing the whole image at once.

Combining Global and Local Models: Finally, we combined the predictions from both global and local models using a convolution layer with a reduced learning rate to create a comprehensive segmentation that leveraged both overall context (from global training) and detailed wound boundary information (from local patch-based training). The combined model was trained with the Adam optimizer and a learning rate of 0.0001 for 75 epochs using Focal Loss.

III. Evaluation Metrics and Results

To assess the performance of different models and training configurations, we used several common evaluation metrics for segmentation tasks: Focal Loss, Precision, Recall, Intersection over Union (IoU), and Dice Score. Below, we explain each metric and present the results obtained for the FPN network under different configurations.

A. Evaluation Metrics

1. **Precision:** Precision represents the proportion of pixels correctly classified as part of the wound region (true positives) out of all pixels that the model predicted as wound (true positives + false positives). A high precision indicates that the model effectively avoids false positives.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{4}$$

2. **Recall:** Recall is the proportion of pixels correctly classified as wound (true positives) out of all actual wound pixels in the ground truth (true positives + false negatives). High recall indicates that the model successfully identifies most of the wound region.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{5}$$

3. **Intersection over Union (IoU):** IoU, also known as the Jaccard Index, measures the overlap between the predicted wound area and the actual wound area.

$$\text{IoU} = \frac{\text{Intersection of Predicted and Ground Truth}}{\text{Union of Predicted and Ground Truth}} \tag{6}$$

4. **Dice Score:** The Dice Score, or Dice Coefficient, is another measure of overlap between the predicted and ground truth segments. It is calculated as:

$$\text{Dice Score} = \frac{2 \times |\text{Predicted} \cap \text{Ground Truth}|}{|\text{Predicted}| + |\text{Ground Truth}|} \tag{7}$$

B. Results

1) Architecture Selection Results: The performance results for global models using UNet, LinkNet, FPN, and PSPNet are presented at three resolutions. Table I shows the results for 256x256, Table II for 512x512, and Table III for 1024x1024.

Model	Precision	Recall	IoU	Dice Score
UNet	0.8559	0.4613	0.4201	0.5585
LinkNet	0.8052	0.4741	0.4245	0.5828
PSPNet	**0.8614**	0.5631	0.5181	0.6706
FPN	0.6381	**0.8752**	**0.5857**	**0.7312**

TABLE I: Performance Results for 256x256 Resolution

Model	Precision	Recall	IoU	Dice Score
UNet	**0.7480**	0.7234	0.5819	0.7291
LinkNet	0.6694	0.7727	0.5597	0.7127
PSPNet	0.7164	0.7469	0.5767	0.7248
FPN	0.6900	**0.8051**	**0.5903**	**0.7380**

TABLE II: Performance Results for 512x512 Resolution

Model	Precision	Recall	IoU	Dice Score
UNet	0.6784	**0.8645**	0.6109	0.7534
LinkNet	0.7807	0.7974	**0.6513**	**0.7853**
PSPNet	0.7904	0.7878	0.6512	0.7850
FPN	**0.8630**	0.7256	0.6464	0.7805

TABLE III: Performance Results for 1024x1024 Resolution

2) Loss Selection Results: The performance results for global and local models using UNet with Dice loss, focal loss, and recall cross-entropy loss are presented in Table IV. The performance results for global and local models using LinkNet with the same loss functions are shown in Table V. Additionally, the performance results for global and local models using PSPNet with Dice loss, focal loss, and recall cross-entropy loss can be found in Table VI. Finally, the performance results for global and local models using FPN with these loss functions are summarized in Table VII.

Model	Precision	Recall	IoU	Dice Score
Global – Dice Loss	0.7457	0.6347	0.6779	0.6779
Global – Focal Loss	**0.8495**	0.7883	0.6888	0.8138
Global – RCE Loss	0.8446	**0.8455**	**0.7303**	**0.8416**
Local – Dice Loss	0.6784	**0.8645**	0.6109	0.7534
Local – Focal Loss	**0.8517**	0.8047	**0.7016**	**0.8229**
Local – RCE Loss	0.8102	0.8315	0.6921	0.8162

TABLE IV: Performance Results for Different Losses for UNet - 1024x1024

Model	Precision	Recall	IoU	Dice Score
Global – Dice Loss	0.8254	**0.8504**	**0.7180**	0.8338
Global – Focal Loss	**0.9010**	0.7607	0.6987	**0.8215**
Global – RCE Loss	0.8630	0.7749	0.6887	0.8128
Local – Dice Loss	0.7807	0.7974	0.6513	0.7853
Local – Focal Loss	0.7657	0.6209	0.5170	0.6713
Local – RCE Loss	**0.7907**	**0.8402**	**0.6879**	**0.8122**

TABLE V: Performance Results for Different Losses for LinkNet - 1024x1024

Model	Precision	Recall	IoU	Dice Score
Global – Dice Loss	0.8100	0.7583	0.6396	0.7771
Global – Focal Loss	**0.8863**	0.6388	0.5797	0.7244
Global – RCE Loss	0.7475	**0.8652**	**0.6633**	**0.7954**
Local – Dice Loss	0.7904	0.7878	0.6512	0.7850
Local – Focal Loss	**0.8602**	0.7303	0.6512	0.7845
Local – RCE Loss	0.8150	**0.8140**	**0.6839**	**0.8098**

TABLE VI: Performance Results for Different Losses for PSPNet - 1024x1024

Model	Precision	Recall	IoU	Dice Score
Global – Dice Loss	0.7460	0.8071	0.6215	0.7638
Global – Focal Loss	**0.9084**	**0.8210**	**0.7607**	**0.8622**
Global – RCE Loss	0.8689	0.8118	0.7186	0.8349
Local – Dice Loss	0.8630	0.7256	0.6464	0.7805
Local – Focal Loss	**0.8846**	0.7154	0.6536	0.7852
Local – RCE Loss	0.8475	**0.7842**	**0.6907**	**0.8117**

TABLE VII: Performance Results for Different Losses for FPN - 1024x1024

3) Combined Training Results: The performance results for the global FPN model, local FPN model, and combined model are presented in Table VIII. The global and local models were each trained with a learning rate of 0.001 using the Adam optimizer for 50 epochs. For the combined model, end-to-end training was performed by adding a convolutional layer after the global and local outputs, allowing both to pass through this layer. This model was trained for 75 epochs with the learning rate reduced to 0.0001.

Model	Precision	Recall	IoU	Dice Score
Global FPN	0.8155	0.7349	0.6215	0.7607
Local FPN	0.8162	0.7373	0.6129	0.7501
Combined Model	**0.8381**	**0.8105**	**0.6881**	**0.8050**

TABLE VIII: Performance Results for Global, Local, and Combined FPN Models

C. Model Analysis and Future Enhancements

The combined model achieved a higher Dice Score compared to the individual global and local models, demonstrating the effectiveness of integrating global and local perspectives for wound segmentation.

A notable observation was that model performance dropped significantly when evaluated with a batch size different from the one used during training. Specifically, when the evaluation batch size was reduced from 8 (used in training) to 4, segmentation accuracy and consistency noticeably declined. This finding underscores the importance of maintaining a uniform batch size across training and testing phases, as the model may rely on batch-specific patterns or normalization statistics that shift with batch size.

While these results are promising, there remain several avenues for further enhancement. For instance, increasing the number of training epochs or incorporating batch normalization could help improve model stability and accuracy. Such adjustments may reduce the model's sensitivity to batch size variations, contributing to more reliable performance under diverse testing conditions.

Additionally, implementing a learning rate scheduler during the training of the combined model could further optimize performance. A scheduler can adapt the learning rate throughout training, guiding the convolutional layer to effectively fuse global and local features and potentially achieving a more effective global minimum. These enhancements offer promising directions for future work to improve segmentation accuracy and robustness.

REFERENCES

[1] S. R. Oota, V. Rowtula, S. Mohammed, M. Liu and M. Gupta, "WSNet: Towards An Effective Method for Wound Image Segmentation," in *Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2023, pp. 3233-3242, doi: 10.1109/WACV56688.2023.00325.