Study of Chronic Wound Image Segmentation: Impact of Tissue Type and Color Data Augmentation

Chronic wound segmentation is an essential task for evaluating wound and its recovery progress. A physician usually measures a wound area to choose proper treatment according to wound conditions. However, precise measurement needs accurate image-region segmentation. With the advent of deep learning for semantic image segmentation, accuracy of region segmentation is dramatically higher than traditional methods. Unfortunately, semantic segmentation in prior work did not produce satisfactory outputs in wound image segmentation, even with a large training dataset. This work, therefore, rethinks about the challenge and aims at not only improving segmentation accuracy, but also studying the impact of wound tissue types and color on accuracy. Since an end-to-end approach of semantic segmentation in prior work performed relatively poorly, the proposed method employs both image processing and deep learning techniques. The experiments indicated that slough was the most challenging tissue to be segmented. Also, properly increasing color variety of wound images significantly improved segmentation performance. The accuracy of the proposed method was 72%, 40%, and 53% in terms of intersection over union for granulation, necrosis, and slough wound tissue types, respectively. The proposed method outperformed a prior end-to-end approach, even though this method employed particularly simpler neural network models and much smaller number of training images.

Keywords— wound measurement; deep learning; image segmentation;

# Introduction

Chronic wound segmentation is an essential task for evaluating wound and its recovery progress. A physician usually measures a wound area to choose proper treatment according to wound conditions. Current practice tends to use a scale to measure width and length of the wound. Nevertheless, wound shape is highly arbitrary, thereby making it difficult to set a standard that has minor inter- and intra-observer variability in measurement. This urges the use of imaging techniques directly measuring the actual wound area.

Nonetheless, precise measurement needs accurate image region segmentation. Furthermore, an automatic method is preferred, as it can minimize the observer variability. This leads to many research on wound image segmentation. Earlier techniques, however, relied on certain assumed conditions, such as a controlled flash and a use of expensive cameras [1][2]. Although the accuracy seemed to be high, this renders these techniques impractical for use in common clinical environments.

The above issues require a more sophisticated methods, such as deep learning, to deal with. Since the introduction of AlexNet [3], deep learning has shown its performance advantages in many fields, especially in object recognition. Since then, its applications in image processing have become prevalent and many techniques were proposed. One of the major milestones for image segmentation is the introduction of an end-to-end method for semantic segmentation. This includes a fully convolutional neural network and U-Net, which is well received in a research community [4][5].

Unfortunately, those techniques may not be successful in certain applications when compared with a more traditional techniques. For example, Wang et al. proposed an end-to-end method based on convolutional neural network and trained it with, to the best of our knowledge, the largest wound image dataset [6], but its accuracy in terms of intersection over union (IoU) was merely 47.3%.

This urged the authors to rethink about the challenge in chronic wound segmentation. We looked into the nature of the image data in a Medetec image database [7] and found that even a wound tissue of the same type could greatly vary in shape, size, and color. Also, some slough tissues have similar color tone to the patient’s skin.

Therefore, the hypotheses that drive this research are: (1) a type of wound tissues plays an important role in segmentation accuracy and (2) utilization of better color data can significantly improve the accuracy.

To validate the hypotheses, three groups of images were created: granulation, necrosis (eschar), and slough. Criteria for grouping are based on portion of wound tissues in an image. One image may fall into two groups if there are significant amount of two tissue types (approximately > 30% of wound region). Figure 1 depicts an example of image and its corresponding types.

Then, three convolutional neural network models were trained to specialize in a single wound type. These models were designed to perform pixel classification and to be simple for speed and study of impact of each tissue type. Once pixels are classified, additional morphological operations and connected component analysis are executed to obtain the final wound region. Next, the same neural network architecture was trained with all wound types to find a baseline accuracy for comparative study.

In our study, slough was the most challenging tissue because of its resemblance to a patient’s skin color. The baseline accuracy was the worst. Our experiments indicated that models specialized in a certain tissue type performed significantly better, especially for granulation. Therefore, employing multiple models together, instead of one model as done in prior work, to form an ensemble could possibly improve overall performance.

Figure 1. Example of wound tissue types.



a) Granulation

b) Slough

c) Necrosis

Since the images in the dataset may be acquired by several cameras and lighting conditions, the brightness and color tone are not consistent throughout the dataset. Consequently, we also explored whether color data augmentation could improve accuracy. This experiment is important if we want the method being applicable to a typical clinical environment.

Our experiments showed that training models to target a specific tissue type has high impact on accuracy improvement. When trained by color data having higher variation, a rather simple technique could significantly outperform a complex model trained with a much larger training images.

The rest of this paper discusses related work in Section II. Details on our methodology is provided in Section III, while experiments and results are discussed in Section IV. Finally, conclusion and other comments are given in Section V.

# Related work

There have been a large amount of research effort on wound image segmentation and tissue classification. Wannous et al. used multiple methods including mean shift and a support vector machine. The IoU index of their method was high, especially for granulation (80.2%). Nonetheless, their images were acquired in a controlled condition and most of background was manually removed before final segmentation.

Fauzi et al. did not employ a machine learning technique, but instead employed a region growing based on their specialized probability map and a modified version of HSV color space [9]. Their method then performed optimal thresholding to classify pixels. The accuracy of their method was relatively high with 74.0% in the IoU index. Their images were captured by several imaging devices and capture conditions. Their dataset, however, contained only 80 images and these images may involve in their parameter tuning to achieve this accuracy.

Wang et al. employed a deep convolutional neural network [6]. Their end-to-end method uses an encoder and decoder for wound segmentation. Their training set contained up to 2,700 images. The IoU index of their method, however, was comparatively low (47.3%). In their experiments, there was no study about tissue types and all training images were used together to train a single model.

It seems that machine learning methods fall behind an image processing techniques in terms of accuracy, even when a large training dataset is available and a somewhat complex neural network architecture was utilized.

# Methodology

The proposed method is a combination of machine learning and traditional image processing techniques. It starts with pixel classification to detect wound pixels in an input image. Typically, a cluster of wound pixels in an image is detected in this process. Then, morphological operations and connected component analysis are applied to segment a target wound region which is the largest connected component obtained.

Section III.A describes our neural network model employed for pixel classification. Section III.B provides details on image processing techniques for final segmentation and Section III.C gives overview on color data augmentation in this work.

## Architecture of Neural Network Model

The proposed neural network model is based on a convolutional neural network. We designed a simple, yet effective, architecture so that classification in a relatively large number of regions is not time consuming and produces comparatively good results. It is interesting to note that in this application, more complex models such as AlexNet and ResNet did not outperform the proposed model in terms of accuracy and they are relatively slow [3][10] Therefore, the simple architecture is a reasonable choice in this case.

Figure 2 depicts the proposed architecture of a neural network. The network accepts an image patch of size 31 x 31 pixels as an input in an RGB format. Then, spatial convolution is executed on the image patch and data in the next two layers. The kernel size of convolution filters is 5 x 5 and the stride is 2 for both *x*- and *y*-axes. This effectively reduces the image size (feature maps) by 2 for both axes without using max pooling. The number of feature maps generated by each convolution layer is 16, 32, and 64, respectively. Outputs from convolutional layers are fed to ReLU activation function before sending to the next layer. The last two are fully connected layers consisting of nine and three nodes. These three nodes correspond to the number of classes to be classified: wound bed, wound boundary, and background.

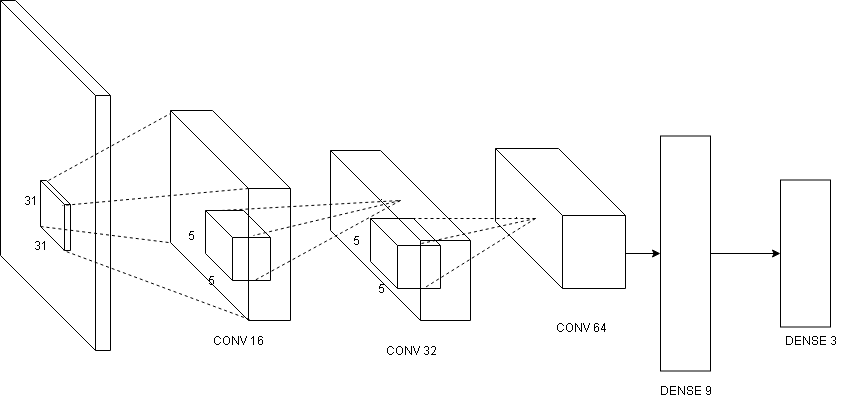


Figure 2. Model architecture.

Prediction results are employed for final segmentation described in the next section.

## Image Segmentation

To segment an image, we utilize neural network for pixel classification. Pixels classified as wound bed or boundary are included for further analysis. Due to computational cost of pixel classification, however, classification is not executed on every image pixel. The stride of 2 was set to reduce the cost by approximately 75%. This results in unconnected wound regions shown in Figure 3c.

To reconstruct a wound region, a morphological close operation is executed. The disc structuring element with 5-pixel diameter is utilized. Since the obtained region tends to be a little smaller than its actual size, a morphological dilation with the disc structuring element whose diameter size is 3 is employed. This tendency arises because we do not classify every pixels and misclassification bias toward the background class.

Once morphological operations are finished, the largest connected component is selected as a wound region. It is worth noting that there may be more than one wound in an image. In that case, a user may manually pick the wound he or she want to measure the area. We are, however, interested in only the largest wound in our settings.

a. Input images

b. Ground truth

c. Predicted region

d. Segmented region

e. Comparison

f. Region overlay

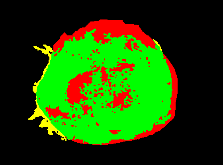
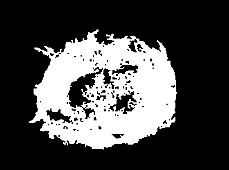
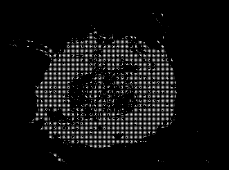
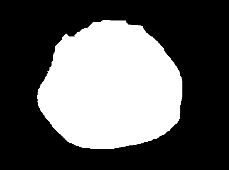
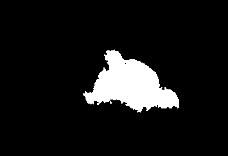


Figure 3. Examples of inputs and segmentation results. Top row: example of a granulation wound and its segmentation. Bottom row: example of a necrosis wound and its segmentation. Column b is a ground-truth region provided by human experts. Column c is prediction outputs from the proposed neural network model. Column d is segmented wound region. Column e is comparison between ground-truth and segmented regions. Column f overlays segmented regions over input images.

Example outputs are shown in Figure 3d. It is important to note that cavity filling is not performed in our method (bottom row) because a wound may have a cavity inside. Filling a cavity without careful analysis may cause an issue in some cases.

## Color Data Augmentation

One of the biggest challenges in wound segmentation is color variation due to environment and image capturing devices. Since the method relies on color data, we need to provide a proper training method that can handle such color variation.

A color data augmentation method proposed in [8] attempt to learn color mapping parameters between cameras to study color variation in actual scenarios. Then, these color mapping parameters are employed to augment data. This technique can reflect possible color difference among images and is applicable even when image capturing devices are not what we used to learn color mapping parameters. In our settings, we utilized color mapping parameters from two camera pairs and triple our training data size.

This color data augmentation technique significantly improves segmentation accuracy in wound segmentation, as color is the primary information source for region classification in our neural network model. Details of performance improvements related to color data augmentation are provided in Section V.B.

a) Granulation

b) Slough

c) Necrosis

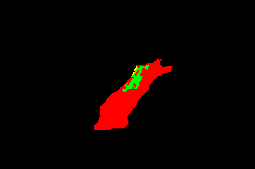
Input

Comparison

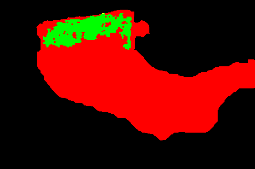
Region overlay

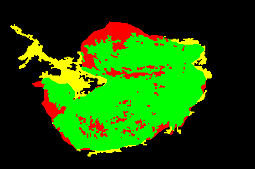
Figure 4. Best and worst outputs for each wound type when models were trained with augmented dataset. The top row of each wound type is the best result and the bottom is the worst.

# Experiments

In our experiments, wound images were taken from Medetec medical image database [7]. It is worth noting that this public database contains only a few hundreds of wound images, while Wang et al.’s private training dataset had 2,700 images. A virtual machine equipped with Intel Xenon@2.6 GHz CPU and NVIDIA Tesla M60 GPU was employed in this study. Sections IV.A through IV.C provide details on data preparation, model training, and evaluation metrics.

## C:\Users\MannyComputer\Downloads\p2_57_blending (1).pngC:\Users\MannyComputer\Downloads\p2_57_compare (1).pngD:\_WORK\_Thesis\image wound care\image\testing_image\slough\p2_57.pngData Preparation

To explore whether all wound types are equally challenging or not, we first divided wound images into 3 types according to the amount of each wound-tissue type in an image. As mentioned earlier, these 3 types were granulation, slough, and necrosis. It is important to note that one image may fall into two groups if there are significant amounts of two wound-tissue types in that image.

We separated data to training dataset and testing dataset. The training dataset consists 3 sets: granulation (77 images), slough (85 images), and necrosis (51 images). There were 180 unique images in the training set. The testing dataset is separated in the same way as the training set and it has 10, 10, and 7 images for granulation, slough, and necrosis, respectively.

Ground truth data were created by manually tracing the wound boundary. Pixels were denoted as wound bed, wound boundary and background. We counted both wound bed and boundary as wound region.

Since a model is designed to work with an image patch of size 31 x 31 pixels, we need to sample such image patches throughout the training images. Since the number of samples for wound boundary is smallest, to balance the training set, wound-bed and background regions were randomly selected by the number of wound-boundary samples.

## Model Training

Model training was executed with Microsoft CNTK 2.3. We employed adaptive learning rate. Initially, the learning rate per sample was 0.2 and decreased by half for each epoch. Once epoch 6 was reach, however, the learning rate was set to a constant, which is 0.2/32. In addition, the momentum was set to 0.9 with 0.0001 weight decay for every epoch.

In this work, four models were trained where three were trained according to their target wound types and the other was trained for all wound types. In other words, all training samples were applied to train the last model. Lastly, to evaluate the impact of color data augmentation, color data augmentation discussed in Section III.C was utilized to train other four corresponding models.

There were 130,839 samples, 165,987 samples, and 103,608 samples for training granulation, slough, and necrosis models, respectively. There were 332,061 samples for training the combined model. Also, color data augmentation tripled the size of training samples for each model. For the testing set, there were 99,093 samples, 131,985 samples, 60,408 samples, and 196,758 samples for testing granulation, slough, necrosis, and combined models, respectively.

## Evaluation

Precision, recall, and region overlap were performance metrics in this study. Precision and recalled were measured from pixel prediction of trained models, while region overlap was computed by intersection over union (IoU) of final segmentation output, which is

where *GT* refers to a set of ground truth pixels provided by experts, and *CS* refers to a set of pixels in a computer segmented region.

# Results

This section provides results of the proposed method. The results are divided into two sub-sections based on our hypothesis about impacts of wound tissue types and color data.

## Impact of Tissue Types

The first experiment focuses on the learning capabilities of neural network models on wound tissues. We started with the combined model, which employed all wound tissues for training. We also investigated how this model performed on each tissue type. As expected, this method performed relatively poor. The mean IoU was merely 32.28% (Table I), while the encoder-decoder model proposed by Want et al. obtained 47.3% [6]. Results on specific tissue types were similar to what other image processing techniques obtained in that slough wound appears to be the most difficult while granulation appears to be easiest [1].

Table I

PERFORMANCE ON PRECISION, RECALL, AND IOU INDICES OF THE COMBINED MODEL.

|  |  |  |  |
| --- | --- | --- | --- |
| **Testing dataset** | **Precision (%)** | **Recall (%)** | **IoU (%)** |
| Combined | 50.31 | 47.61 | 32.28 |
| Granulation | 44.28 | 63.31 | 42.68 |
| Slough | 62.11 | 32.60 | 22.49 |
| Necrosis | 50.58 | 51.66 | 31.85 |

As mentioned earlier, we hypothesized that a model specifically trained for a wound tissue type could possibly outperform a much more complex model. Therefore, another three models were trained as such and their performance was shown in Table II. Performance on the three metrics significantly improved. More importantly, our proposed method outperformed Wang et al.’s method, which utilized an encoder-decoder convolutional neural network model [6].

Table II

PERFORMANCE ON PRECISION, RECALL, AND IOU INDICES OF TISSUE-SPECIFIC MODELS. The average IoU was weighted by the number of images in each wound-tissue type.

|  |  |  |  |
| --- | --- | --- | --- |
| **Models** | **Precision (%)** | **Recall (%)** | **IoU (%)** |
| Granulation | 92.44 | 77.18 | 67.65 |
| Slough | 48.59 | 61.52 | 34.16 |
| Necrosis | 61.29 | 75.57 | 48.69 |
|  |  | Average | 50.33 |

Unfortunately, accuracy on slough and necrosis wounds was not satisfactory. Since distinction of color in these wounds appeared to be highly difficult, a novel color data augmentation method was applied [8]. Its impact is discussed in the next section.

## Impact of Color Data Augmentation

We augmented the training dataset that simulated colors achieved by other two cameras. This effectively tripled the size of the training set. Tables III and IV shows results of corresponding models from the previous section. It is clear that color data augmentation improved the performance of all models and, in turn, greatly increased accuracy in terms of IoU index.

Although the overall accuracy was still unsatisfactory, the proposed method demonstrated a significant improvement in wound segmentation. Color data is one of the key factors in wound segmentation. Properly increasing the variety of color data can enhance overall performance of the proposed method.

Regarding computation time, training took approximately 1 to 3 hours depending on the model type. The whole segmentation took roughly 10 seconds per image.

Table III

PERFORMANCE ON PRECISION, RECALL, AND IOU INDICES OF THE COMBINED MODEL TRAINED WITH AUGMENTED DATA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Testing dataset** | **Precision (%)** | **Recall (%)** | **IoU (%)** |
| Combined | 58.23 | 63.61 | 42.19 |
| Granulation | 67.07 | 88.05 | 59.74 |
| Slough | 44.21 | 41.27 | 22.04 |
| Necrosis | 63.49 | 68.53 | 45.88 |

Table IV

PERFORMANCE OF TISSUE-SPECIFIC MODELS TRAINED WITH AUGMENTED DATA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Models** | **Precision (%)** | **Recall (%)** | **IoU (%)** |
| Granulation | 92.56 | 81.01 | 72.08 |
| Slough | 65.64 | 53.95 | 40.40 |
| Necrosis | 79.28 | 70.05 | 53.52 |
|  |  | Average | 55.53 |

# Conclusion, Discussion and Future Work

Based on the results of Section V, a tissue type and color data augmentation highly affect both classification and segmentation of wound data. With minor user effort to specify the dominant wound tissue type in the image, a somewhat simple technique could outperform a complex end-to-end neural network model.

More interestingly, prior work trained a model with 2,700 images and obtained only 47% in the IoU index. This work trained with 180 images, but separation of tissue types and color data augmentation allowed the method to achieved up to 55% in the IoU index.

It seems that dealing with the wound segmentation problem with a large number of training set and somewhat complex neural network model was less effective than dealing with tissue identification and better color data augmentation.

In the future, we will explore how an end-to-end, fully automatic approach can benefit from this study.

# References

1. H. Wannous, S. Treuillet, Y. Lucas, “Supervised tissue classification from color images for a complete wound assessment tool,” in Proc. IEEE EMBS Int. Conf., pp. 6031–6034, 2007.
2. N.D.J. Hettiarachchi, R.B.H. Mahindaratne, G.D.C. Mendis, H.T. Nanayakkara, N.D. Nanayakkara, “Mobile-based wound measurement,” in Proc. IEEE Point-of-Care Healthcare Technol., pp. 298–301, 2013.
3. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ‘ImageNet classification with deep convolutional neural networks’. Neural Information Processing Systems (NIPS 2012), pp. 1106–1114, 2012.
4. Long, J., Shelhamer, E., Darrell, T.: ‘Fully convolutional networks for semantic segmentation’, 2015.
5. Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham.
6. C. Wang et al., "A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks," 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, pp. 2415-2418, 2015.
7. S. Thomas, “Wound dressings,” in The Epidermis in Wound Healing, D. T. Rovee and H. I. Maibach, Eds., CRC Press, Boca Raton, Fla, USA, 2004. Website: http://www.medetec.co.uk/files/medetec-image-databases.html
8. [disclosed to conceal author identity during review process], "Estimation of Color-Mapping Parameters among Multiple Cameras for Data Augmentation and Color Calibration", submitted to 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), Nakhon Pathom, 2018.
9. M. F. A. Fauzi, I. Khansa, K. Catignani, G. Gordillo, C. K. Sen, and M. N. Gurcan, "Computerized segmentation and measuremen tof chronic wound images", Computers in Biology and Medicine 60, pp. 74-85, 2015
10. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 770-778.doi: 10.1109/CVPR.2016.90