

A Unified Framework for Automatic Wound Segmentation and Analysis with Deep Convolutional Neural Networks

Changhan Wang¹, Xinchun Yan¹, Max Smith¹, Kanika Kochhar², Marcie Rubin³
Stephen M. Warren³, James Wrobel², and Honglak Lee^{1,*}

Abstract—Wound surface area changes over multiple weeks are highly predictive of the wound healing process. Furthermore, the quality and quantity of the tissue in the wound bed also offer important prognostic information. Unfortunately, accurate measurements of wound surface area changes are out of reach in the busy wound practice setting. Currently, clinicians estimate wound size by estimating wound width and length using a scalpel after wound treatment, which is highly inaccurate. To address this problem, we propose an integrated system to automatically segment wound regions and analyze wound conditions in wound images. Different from previous segmentation techniques which rely on hand-crafted features or unsupervised approaches, our proposed deep learning method jointly learns task-relevant visual features and performs wound segmentation. Moreover, learned features are applied to further analysis of wounds in two ways: infection detection and healing progress prediction. To the best of our knowledge, this is the first attempt to automate long-term predictions of general wound healing progress. Our method is computationally efficient and takes less than 5 seconds per wound image (480 by 640 pixels) on a typical laptop computer. Our evaluations on a large-scale wound database demonstrate the effectiveness and reliability of the proposed system.

I. INTRODUCTION

Currently, millions of patients are suffering from chronic wounds, which results in billions of annual spending [20]. For diabetes-related foot ulcers alone, it has been estimated that 38 billion dollars are spent caring for these patients in the United States [6]. While the majority of costs are attributed to hospital admissions and surgery, the majority of long-term care costs are attributed to home care and social services [1]. Moreover, time-consuming quantitative wound assessments are not always feasible due to the high volume of patients. In this setting, accurate diagnosis and timely treatment will highly rely on expertise and experience.

Advances in image processing and machine learning [14], [4], [21], [19], [24] enable the analysis of wound images by computer programs. Following this trend, recent work has started to explore the fields of wound segmentation and wound condition analysis. In [11], [12], [23], wound segmentation was formulated as a foreground-background classification problem and solved by regular classifiers (e.g., SVM [19]) with hand-crafted features. Attempts were also

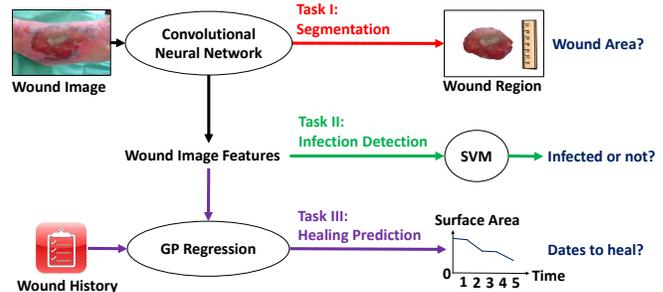


Fig. 1. **Illustration of our proposed system.** Given wound images, our deep convolutional neural network model performs feature learning and wound segmentation at the same time. Wound segments can be used to estimate actual wound areas (with the scale information from the ruler ticks). In addition, ConvNet visual features are used to assist wound condition analysis (infection detection and healing progress prediction).

made in wound surface area estimation and healing progress prediction [8], [15]. Although achieving good performance, those models are highly dependent on hand-crafted features and focus only on single tasks and small databases. To the best of our knowledge, there is no unified framework for joint wound segmentation and wound condition analysis which is verified on large-scale databases.

To this end, we propose an integrated system based on deep learning (see [2] for a survey) that simultaneously performs wound segmentation and analysis. An overview of our system is illustrated in Fig. 1. Given wound images as inputs, the system first automatically segments the wound region from the background by a novel variant of deep convolutional neural network (ConvNet) [14]. From wound region segments, actual wound surface areas are then calculated with the scale information from the ruler ticks in wound images. At the same time, learned ConvNet features are used in (1) infection detection via SVM classifiers [19] and (2) healing progress prediction via Gaussian process (GP) regression [24]. Compared to previous segmentation methods which require intermediate steps like image smoothing and hand-crafted feature extraction, our end-to-end model simply takes in raw images and produces final results. Moreover, it has lower computational cost than existing ConvNet architectures [16], [7]. The learned visual features with our proposed model result in good performance for all three tasks.

In model training and evaluation, we adopted the NYU Wound Database, a large-scale dataset with over 8000 high-resolution wound images and corresponding medical records (e.g., clinic visit dates and wound surface areas), and created additional wound segment annotations and binary infection labels. In experiments, we successfully demonstrate our

¹ Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA {wangchh, xcyan, mxsmith}@umich.edu, honglak@eecs.umich.edu

² University of Michigan Medical School, Ann Arbor, MI, USA kanikak@umich.edu, jswrobel@med.umich.edu

³ Department of Plastic Surgery, New York University, NY, USA {Marcie.Rubin, Stephen.Warren}@nyumc.org

* Corresponding author

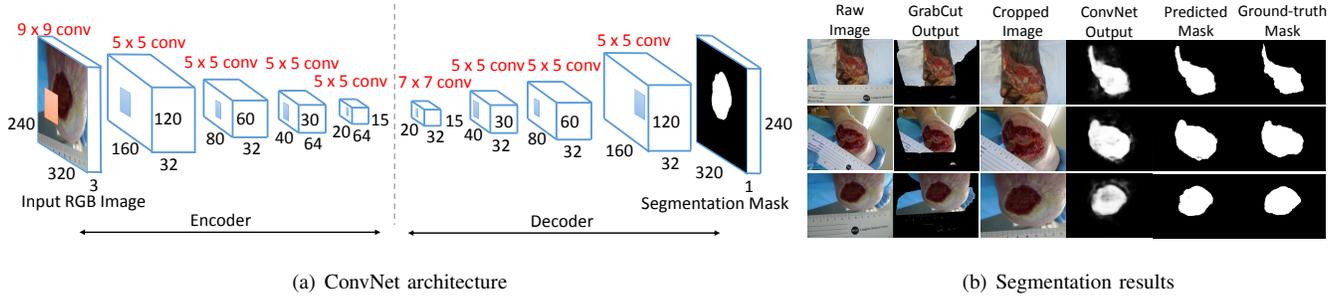


Fig. 2. The ConvNet architecture and segmentation results. (a) Our model performs wound segmentation in an end-to-end style. (b) A modified GrabCut [18] was used to roughly crop wound regions from raw images. The model takes cropped images as inputs and provides pixel-wise probabilities of wound segment masks as outputs (the lighter, the higher probability). Final masks are obtained by setting a threshold of 0.5 on every pixel.

unified framework for wound image segmentation, infection detection, and healing progress prediction.

II. METHODS

A. Wound Segmentation and Surface Area Estimation

Given an $h \times w$ RGB image $X \in [0, 1]^{h \times w \times 3}$, the objective is to learn a function f that produces a binary mask $Y \in \{0, 1\}^{h \times w}$ for the ground-truth wound region $R^{gt} \subseteq \{1, 2, \dots, h\} \times \{1, 2, \dots, w\}$ (i.e., R^{gt} is a set of pixel coordinates that correspond to the wound region):

$$f(X) \approx Y, \text{ where } Y_{i,j} = \mathbb{1}_{\{(i,j) \in R^{gt}\}}(X), \quad (1)$$

To learn the function f in an end-to-end style, we propose the convolutional encoder-decoder network, which is a variant of ConvNet. Basically, the encoder network computes a set of compact feature maps on high-resolution images, while the decoder network upsamples the feature maps to get full-size segmentation masks. More specifically, the encoder network is constructed by stacking basic computation blocks like convolution, non-linear transformation (e.g. ReLU [17]), spatial pooling, and local response normalization [13]. We group the stacked operations into ‘layers’ and define layer-wise functions $g^{(l)}$ (for computing the l -th hidden layer activation $H^{(l)}$ of the network) as follows:

$$\begin{aligned} H^{(l)} &= g^{(l)}(H^{(l-1)}; W^{(l)}, b^{(l)}), \\ &= \text{normalize}(\text{pool}(\text{relu}(W^{(l)} * H^{(l-1)} + b^{(l)}))) \end{aligned} \quad (2)$$

where $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias vector of the l -th layer, respectively, and $*$ is the convolution operator, and $H^{(l-1)}$ is either the $l-1$ th hidden layer activation for $l > 1$ or input image X for $l = 1$ (i.e., $H^{(0)} = X$).

In a naive approach, f might be constructed from the stack of layers:

$$f(X; W, b) = \text{softmax}(g^{(L)} \dots g^{(2)} g^{(1)}(X)) \quad (3)$$

In other words, to produce probability maps, the softmax layer [3] can be appended to the end of the network. However, the final probability maps suffer from the down-scaling effects of the convolution and pooling layers. Interpolation methods may be used to up-scale them to full size, but at the cost of blurry outputs. To improve this, we introduce a decoder network constructed by a stack of layers $h^{(l)}$

with upsampling operations [5] (which can be considered as reverse operations of pooling):

$$h^{(l)}(H) = \text{relu}(W^{(l)} * \text{upsample}(H) + \beta^{(l)}), \quad (4)$$

where W and β are the weight matrix and bias vector of the l -th layer. Then the function in (3) can be represented by:

$$f(X; W, \mathcal{W}, b, \beta) = \text{softmax}(h^{(1)} \dots h^{(L)} g^{(L)} \dots g^{(1)}(X)) \quad (5)$$

A threshold of 0.5 is used to produce final segmentation masks from the pixel-wise probabilities.

With the wound region R^{gt} segmented from the image and the number of wound pixels (i.e. $|R^{gt}|$) counted, we can estimate the actual wound area \mathcal{S} . More specifically, ruler ticks in wound images are used to make the conversion from pixel lengths to actual lengths. Denote r as the conversion ratio, then the actual wound area \mathcal{S} is given by

$$\mathcal{S} = \frac{|R^{gt}|}{r^2}, \quad (6)$$

We use Hough transform [9] to parametrize adjacent ruler ticks (0.5 cm in between) and measure the pixel distance.

B. Wound Infection Detection

This task is defined as a binary classification problem, i.e., assigning each wound image a binary label to indicate whether the wound at the photographing time is infected or not. Given image features \mathbf{x} for the wound image X , the goal is to learn an indicator function f^I :

$$f^I(\mathbf{x}) \approx \mathbb{1}_{\{X \text{ shows an infected wound}\}}(\mathbf{x}) \quad (7)$$

To learn function f^I , we use SVM classifiers [19] using the ConvNet features.

C. Healing Progress Prediction

With wound images $X_{t_1}, \dots, X_{t_N} \in [0, 1]^{h \times w \times 3}$ and corresponding wound surface areas $\mathcal{S}_{t_1}, \dots, \mathcal{S}_{t_N}$ in past N time frames t_1, t_2, \dots, t_N , the objective is to learn a function f^H that predicts wound areas for healing date estimation:

$$\hat{\mathcal{S}}_{t_N + \Delta t} = f^H(\mathcal{S}_{t_N + \Delta t}; X_{t_1}, \dots, X_{t_N}, \mathcal{S}_{t_1}, \dots, \mathcal{S}_{t_N}), \quad (8)$$

where $\Delta t \geq 0$ and $t_N + \Delta t$ is the future date for the prediction.

In our system, Gaussian process regression is used to model f^H . Given image features $\mathbf{x}_1, \dots, \mathbf{x}_N$ for wound images $X_{t_1}, X_{t_2}, \dots, X_{t_N}$, the prior on f^H is modeled as:

$$f^H(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \quad (9)$$

TABLE I
WOUND SEGMENTATION EVALUATION ON THE TEST SET

	Pixel Accuracy	mean IoU
SVM (RGB)	77.6%	26.4%
ConvNet	95.0%	47.3%

where the mean and covariance functions are given by $m(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + c$ and $k(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp[-\frac{1}{2} \|\Lambda(\mathbf{x} - \mathbf{x}')\|^2]$, respectively. Hyperparameters \mathbf{w} , c , σ and Λ are obtained by the maximum likelihood estimation [24].

III. EXPERIMENTAL RESULTS

Preprocessing. Images in the NYU Database are in high resolution (roughly 600 by 900 pixels) which would cause high computational costs in processing. Since limb regions cover less than 30% of the area in most images, extracting features from the entire image is unnecessary. Therefore, we cropped images to 480×640 pixels with a modified version of GrabCut [18] to concentrate on limb regions (shown in Fig. 2(b)).

Experimental Setup. We used an NVIDIA Telsa K40 GPU to speed up parameter learning and tested the performance of the learned model on a laptop computer with Intel Core i5-3317U CPU and 8GB RAM.

A. Wound Segmentation and Surface Area Estimation

As illustrated in Fig. 2(a), our proposed convolutional neural networks have 5 encoding layers followed by 4 decoding layers. Specifically, we used ReLU as nonlinearity function for both convolutional encoder and decoder. At the output of network, we used cross-entropy loss function, together with a L2 regularization term with regularization coefficient 10^{-5} . We trained the model using the mini-batch Stochastic Gradient Descent (mini-batch size is 16) with Nesterov Momentum [22].

For the baseline, we used a linear SVM classifier (with hinge loss), treating segmentation as binary classification for each pixel. In our experiments, 9×9 image patches randomly sampled from the image were used as data for training the linear SVM classifier. We took the raw RGB intensity as the feature for each pixel. For testing, we adopted a sliding-window approach to classify each pixel given a patch centered at that particular pixel.

For this task, we used 500 training images and 150 testing images. The qualitative wound segmentation results are shown in Fig. 2(b). For quantitative comparison, average pixel accuracy (PA) and mean intersection-over-union (mean IOU) are adopted as evaluation metrics. Note that IOU reflects the degree of overlapping between the predicted mask and the ground-truth mask (i.e., given the predicted binary mask R^{pred} and the ground-truth binary mask R^{gt} , the IoU is defined as $\frac{|R^{pred} \cap R^{gt}|}{|R^{pred} \cup R^{gt}|}$). As shown in Table I, our proposed ConvNet achieved significantly better results than the baseline model using raw features.

For area estimation, ruler ticks (short, parallel line segments) were automatically parametrized by Hough transform and then the average pixel distance between adjacent long tick marks (0.5 cm in between) were computed. With the foreground mask and the conversion ratio (from 0.5 cm to its

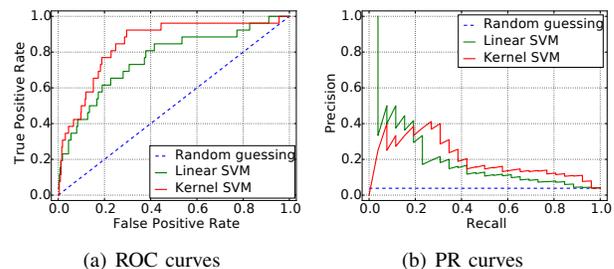


Fig. 3. ROC and PR curves for infection detection. Red curves correspond to results of the kernel SVM while green curves correspond to results of the linear SVM. Random guessing is represented by a dashed blue line.

TABLE II
INFECTON DETECTION EVALUATION ON THE TEST SET. THE CONVNET FEATURES WERE USED FOR TRAINING ALL SVM MODELS.

	p	Accuracy	Recall	Precision	F-1 Score	AUC
Random guessing	10%	86.9%	10%	3.83%	0.055	50%
	50%	50%	50%	3.83%	0.071	
	100%	3.83%	100%	3.83%	0.074	
Linear SVM		95.3%	23.1%	33.3%	0.273	76.3%
Kernel SVM		95.6%	30.8%	40.0%	0.348	84.7%

pixel length) estimated, the actual wound area was calculated via Equation (6).

B. Wound Infection Detection

For this task, 2700 images (including 120 infected cases) were used for training while we evaluated the performance on 700 images (including 35 infected cases). We adopted 5-fold cross validation for hyperparameter search. As mentioned previously, hidden layer activations calculated by the ConvNet were used as our features for infection detection (feature dimension was reduced by a factor of 25 via spatial pooling). We regarded infection as positive sample and trained SVM classifiers (e.g., linear and polynomial kernels) to handle the binary classification task.

In Table II, we report classification accuracy, recall, precision, F-1 score and the area under ROC curve (AUC) as our evaluation metrics. Our kernel SVM classifier achieves 84.7% AUC score while linear SVM classifier achieves 76.3%. Both are significantly better than random guessing (guessing positives in probability p) which achieves 50% AUC. For detailed comparison, we demonstrate the ROC curves and precision-recall (PR) curves in Fig. 3.

C. Healing Progress Prediction

For this task, we used 192 wound sequences with their corresponding medical records (e.g., actual area, wound image and patient information), splitting into 160 training sequences and 32 testing sequences. We initialized our Gaussian process regression (GPR) model using the first half sequence, and then predicted healing progress in the future (the second half). In general, our model is assumed to make long-term predictions, since the wound sequences span a period ranging from 40 to 150 weeks. We provide quantitative analysis of the time (weeks) to take until the wound size become 10%, 5%, and 0% of original wound area, measured by mean absolute error (MAE_{time}). We also report the average mean absolute error of predicted wound areas across all time

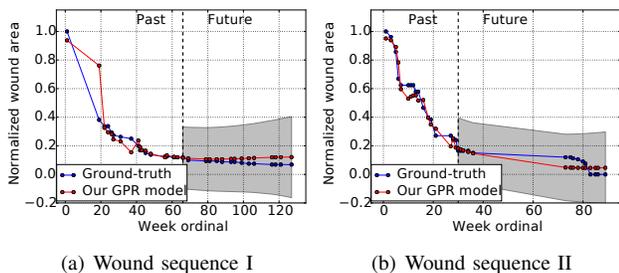


Fig. 4. Healing progress of two wound sequences. The y axis represents the relative wound area while the x axis represents the time (in weeks). The red curve corresponds to the predictions of the GP regression while the blue curve shows the ground-truth. The shaded area indicates the 95% confidence intervals. Given past records, our GP regression model can make long-term predictions of the future.

TABLE III
HEALING PROGRESS EVALUATION

	MAE _{time} (10%)	MAE _{time} (5%)	MAE _{time} (0%)	Avg. MAE _{area}
Linear	8.84	18.64	3.30	6.06%
Polynomial (3-order)	16.70	5.11	3.81	6.07%
GP	10.07	2.94	2.17	3.95%

frames (Avg. MAE_{area}), defined as $\frac{1}{T} \sum_{t=1}^T \frac{|R_{area}(t) - R_{gt}(t)|}{R_{gt}(t)}$, where T is the number of observations of wound condition.

To incorporate both visual and non-visual information, we constructed our feature vector by including 1) the current week ordinal; 2) the previous time ordinal; 3) previous wound area (relative to the original area); 4) visual features for previous time (reduced to 10 dimensions by performing PCA [10] on the ConvNet features used in the infection detection); and 5) patient age and gender. During the training period, GPR hyperparameters were selected through maximum likelihood estimation.

As shown in Fig. 4, our Gaussian process regression model is able to capture healing dynamics with the previously defined features. Specifically, combining visual and non-visual characteristics in a temporal fashion, we are able to predict the healing trends in the future.

We use linear and polynomial (3-order) regression models as baselines to compare with our GP regression model (Table III). Considering the average healing duration is 95.67 weeks in our testing set, we find all the models are accurate enough to predict the healing dates with our features. It is worth noting that the GP model has the least errors when wound areas approach 0 (entirely healed). This result suggests reliability of our proposed system in making long-term predictions, which can benefit both patients and clinicians.

IV. CONCLUSIONS

We introduced a unified framework for wound segmentation and wound condition analysis. With our proposed ConvNet model, our system can segment wound image in an end-to-end style. At the same time, learned ConvNet features can benefit other tasks, such as wound infection detection and healing progress prediction. Overall, the system is efficient enough to process the wound image within 5 seconds on

a typical laptop computer. Finally, we demonstrated the strength of our system in performing all three tasks.

ACKNOWLEDGMENT

The experiments involving human subjects were approved by the Institutional Review Board. This work was partly supported by the MCubed initiative and Google Faculty Research Award. We are grateful to NVIDIA Corporation for the donation of Tesla K40 GPU. We thank Kihyuk Sohn, Ruben Villegas, Junhyuk Oh, and Ye Liu for helpful comments and discussions.

REFERENCES

- [1] J. Apelqvist, G. Ragnarson-Tennvall, J. Larsson, and U. Persson. Long-term costs for foot ulcers in diabetic patients in a multidisciplinary setting. *Foot & Ankle International*, 16(7):388–394, 1995.
- [2] Y. Bengio. Learning deep architectures for ai. *Foundations and trends in Machine Learning*, 2(1):1–127, 2009.
- [3] C. M. Bishop. *Pattern recognition and machine learning*. Springer New York, 2006.
- [4] D. Comaniciu and P. Meer. Robust analysis of feature spaces: color image segmentation. In *CVPR*, 1997.
- [5] A. Dosovitskiy, J. T. Springenberg, and T. Brox. Learning to generate chairs with convolutional neural networks. In *CVPR*, 2015.
- [6] V. R. Driver, M. Fabbi, L. A. Lavery, and G. Gibbons. The costs of diabetic foot: the economic case for the limb salvage team. *Journal of vascular surgery*, 52(3):17S–22S, 2010.
- [7] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In *NIPS*, 2014.
- [8] G. C. Gurtner, S. Werner, Y. Barrandon, and M. T. Longaker. Wound repair and regeneration. *Nature*, 453(7193):314–321, 2008.
- [9] P. V. Hough. Method and means for recognizing complex patterns. Technical report, 1962.
- [10] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.
- [11] M. Kolesnik and A. Fexa. Multi-dimensional color histograms for segmentation of wounds in images. In *Image Analysis and Recognition*, pages 1014–1022. Springer Berlin Heidelberg, 2005.
- [12] M. Kolesnik and A. Fexa. How robust is the SVM wound segmentation? In *IEEE Signal Processing Symposium*, pages 50–53, 2006.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.
- [14] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.
- [15] C. P. Loizou, T. Kasparis, O. Mitsi, and M. Polyviou. Evaluation of wound healing process based on texture analysis. In *IEEE 12th International Conference on Bioinformatics & Bioengineering*, 2012.
- [16] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. *arXiv preprint arXiv:1411.4038*, 2014.
- [17] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, 2010.
- [18] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309–314, 2004.
- [19] B. Schölkopf and A. J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [20] C. K. Sen, G. M. Gordillo, S. Roy, R. Kirsner, L. Lambert, T. K. Hunt, F. Gottrup, G. C. Gurtner, and M. T. Longaker. Human skin wounds: a major and snowballing threat to public health and the economy. *Wound Repair and Regeneration*, 17(6):763–771, 2009.
- [21] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [22] I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In *ICML*, 2013.
- [23] F. Veredas, H. Mesa, and L. Morente. Binary tissue classification on wound images with neural networks and bayesian classifiers. *IEEE Transactions on Medical Imaging*, 29(2):410–427, 2010.
- [24] C. K. Williams and C. E. Rasmussen. Gaussian processes for machine learning. *MIT Press*, 2006.