

딥러닝을 이용한 창상 분할 알고리즘

강현영¹ · 허연우² · 전재준² · 정승원² · 김지예³ · 박성빈^{4*}

¹연세대학교 의공학과, ²연세대학교 원주의과대학 피부과학교실,
³연세대학교 원주의과대학 성형외과교실, ⁴연세대학교 원주의과대학 정밀의학과

Development of wound segmentation deep learning algorithm

Hyunyoung Kang¹, Yeon-Woo Heo², Jae Joon Jeon², Seung-Won Jung²,
Jiye Kim³ and Sung Bin Park^{4*}

¹Department of Biomedical Engineering, Yonsei University, Korea

²Department of Dermatology, Yonsei University Wonju College of Medicine, Korea

³Department of Plastic and Reconstructive Surgery, Yonsei University Wonju College of Medicine, Korea

⁴Department of Precision Medicine, Yonsei University Wonju College of Medicine, Korea

(Manuscript received 10 April 2024 ; revised 23 April 2024 ; accepted 23 April 2024)

Abstract: Diagnosing wounds presents a significant challenge in clinical settings due to its complexity and the subjective assessments by clinicians. Wound deep learning algorithms quantitatively assess wounds, overcoming these challenges. However, a limitation in existing research is reliance on specific datasets. To address this limitation, we created a comprehensive dataset by combining open dataset with self-produced dataset to enhance clinical applicability. In the annotation process, machine learning based on Gradient Vector Flow (GVF) was utilized to improve objectivity and efficiency over time. Furthermore, the deep learning model was equipped U-net with residual blocks. Significant improvements were observed using the input dataset with images cropped to contain only the wound region of interest (ROI), as opposed to original sized dataset. As a result, the Dice score remarkably increased from 0.80 using the original dataset to 0.89 using the wound ROI crop dataset. This study highlights the need for diverse research using comprehensive datasets. In future study, we aim to further enhance and diversify our dataset to encompass different environments and ethnicities.

Key words: Deep learning, Wound segmentation, Machine learning labeling, Medical image analysis

1. 서 론

만성창상(chronic wound) 세계 인구의 상당 부분에 영향을 미치며, 선진국에서는 인구의 1-2%가 영향을 받는 것으로 추정된다[1]. 이는 해당 국가의 보건의료시스템에 중대한 재정적 부담을 주며, 만성 창상을 겪는 환자들의 삶의 질을 저하시킨다. 6.5백만 명 이상의 사람들이 만성 창상을 앓고 있

으며, 이로 인해 미국에서는 매년 2500억 달러 이상의 비용이 발생하는 것으로 추정된다. 미국 내에서만 창상 관리 및 치료에 매년 약 250억 달러가 소요된다[2].

이러한 엄청난 부담에도 불구하고, 효율적인 창상 치료는 경험이 풍부한 임상 의사들에게도 도전적인 과제로 남아있다. 창상 진단은 주로 전문가의 육안에 의해 이루어지며, 이는 전문가의 주관성과 전문성에 따라 결과가 크게 달라질 수 있다는 한계를 내포한다[3]. 창상 면적과 조직 면적을 정량화한다면, 이는 창상 진단의 보조 수단으로 활용될 수 있다. 컴퓨터 기술의 발전으로 창상 및 조직 분할을 위한 알고리즘 개발이 진행되어 왔다. 창상 영역의 정량화는 다양한 방법으로 시도되었으며, 가장 일반적인 방법은 자를 이용하여 창상의 너비 및 길이를 측정하는 것이다[4,5]. 이러한 측정은 창상이 일반적으로 직사각형 형태가 아니기 때문에 정확한 면적 측

*Corresponding Author : Sung Bin Park
Yonsei Wonju College of Medicine, 20, IIsan-ro, Wonju-si, Republic of Korea
Tel: +82-033-741-0456
E-mail: sung.b.park@gmail.com

본 연구는 National Research Foundation of Korea(NRF) 과제의 지원을 받아 수행하였음. (NRF-2022R1A2C2091160).
본 연구는 산업통상자원부와 한국산업기술진흥원의 “지역혁신클러스터육성(R&D, P0025442)”사업의 지원을 받아 수행된 연구결과임.

정에 어려움이 있다. 컴퓨터 지원 창상 분할 방법에 대한 연구도 활발히 진행 중이지만, 초기 연구는 영상 처리 기법을 활용해 소규모 데이터셋에 적용되었으며, 대규모 데이터셋에 대한 견고성이 입증되지 않았다[6,7].

인공지능의 발전으로 다양한 환경에서 수집된 데이터셋을 활용하여 만성 창상의 자동 분할을 위한 딥 러닝 모델을 학습시키는 것이 가능해졌다. 예를 들어, 한 연구팀은 밝기 보정을 사용하여 데이터셋을 딥러닝에 적용하였고, 다른 연구에서는 관심 영역을 강조하는 창상 감지 데이터셋을 딥러닝에 적용하였다[7,8]. 그럼에도 불구하고, 현재까지의 딥러닝 창상 연구에서 각 딥러닝 모델이 특정 데이터셋에 의존한다는 한계가 여전히 존재한다.

본 연구의 목적은 어떠한 데이터셋에도 적합한 수준의 분할 성능을 달성하는 것이다. 이를 위하여 직접 새로운 데이터셋을 구축하였으며, 특정 데이터셋에 대한 의존도를 줄이기 위해 공개 데이터셋과의 결합을 통해 연구를 수행하였다. 데이터셋 구축 과정에서 필요한 시간 및 정답 도출을 위한 명확한 기준을 설정하기 위하여, Gradient Vector Flow (GVF) 기반의 기계 학습 알고리즘을 적용하였다. 추가적으로, 창상 기준에 따른 이미지 데이터의 전처리 과정을 거친 후 학습을 실시함으로써 모델의 정확도를 향상시키는 방법을 제시하였다[9].

II. 연구 방법

1. 데이터셋

본 연구에서 데이터셋은 자체 제작한 데이터셋과 공개 데

이터셋을 결합하여 창상 데이터셋을 제작하였으며, 이는 그림 1에 표시하였다. 그림 1(a)는 신촌 세브란스 병원 피부과와 협력하여 2016~2018년에 모집된 30명의 환자들의 창상 추이를 칼라패치와 함께 촬영한 데이터셋이며, 이 중 정도에 따라 260장의 이미지를 선별하여 데이터셋을 구성했다. 칼라패치는 각 창상 이미지의 색정보 조정 및 이미지 내에서의 창상 크기의 지표로 활용하였다[10]. 그림 1(b)는 Medical Image Computing and Computer Assisted Invention Society (MICCAI)에서 Diabetic Foot Ulcer Challenge를 위해 공개된 데이터셋을 나타낸다. 이 데이터셋은 당뇨병 환자의 발에 발생하는 궤양 또는 열상의 이미지를 포함하며, 총 952장이다. 새롭게 제작한 전체 데이터셋 1212장 중 914장을 훈련데이터로, 298장을 검증데이터로 사용하였다.

2. GVF 기반 기계 학습을 이용한 레이블링

딥러닝 모델을 훈련시키기 위해서는 정답이 필요하다. 정답에 객관성을 높이며 정답을 제작하는 데 소요되는 시간을 줄이기 위해 GVF 기반 기계 학습 알고리즘을 방법을 도입했으며, 전문가의 검토 과정을 거쳐 레이블링이 완성되었다.

그림 2는 GVF 방법을 이용해 창상 이미지의 정답을 생성하는 방법의 예시이다. 그림 3에서는 구체적인 GVF 적용 단계가 명시 되어 있다. GVF를 사용하기 위해서는 그림 3(e)와 같은 초기 경계 영역이 필요하다. 창상의 대략적인 초기 경계를 구하기 위한 영상처리 알고리즘이 그림 3(a)~(d)을 통해 도식화 되어 있다. 먼저 K-means clustering 알고리즘을 통해 창상, 피부, 배경 세 가지로 분류화를 진행하며, 데이터의



그림 1. 데이터셋 (a) 자체 제작 데이터셋 (b) 공개 데이터셋
Fig. 1. Datasets (a) our dataset (b) open dataset by MICCAI

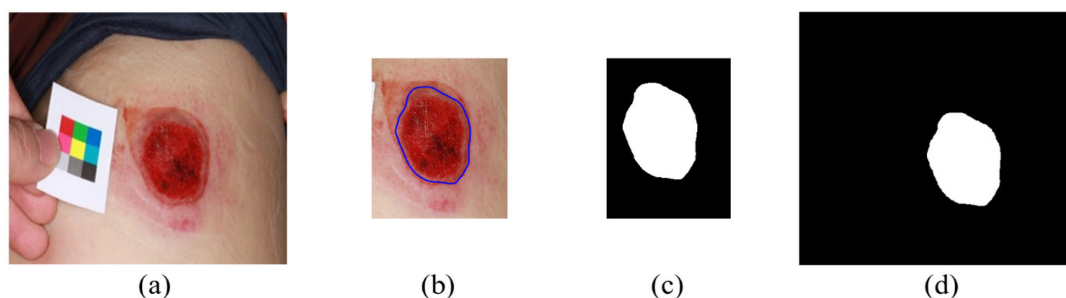


그림 2. GVF 기반 머신러닝을 이용한 레이블링 (a) 원본영상 (b) GVF 적용된 영상 (c) 이진화 영상 (d) 레이블 완료된 영상
Fig. 2. Labeling using GVF machine learning algorithms (a) original image (b) GVF applied image (c) binarized image (d) labeled image

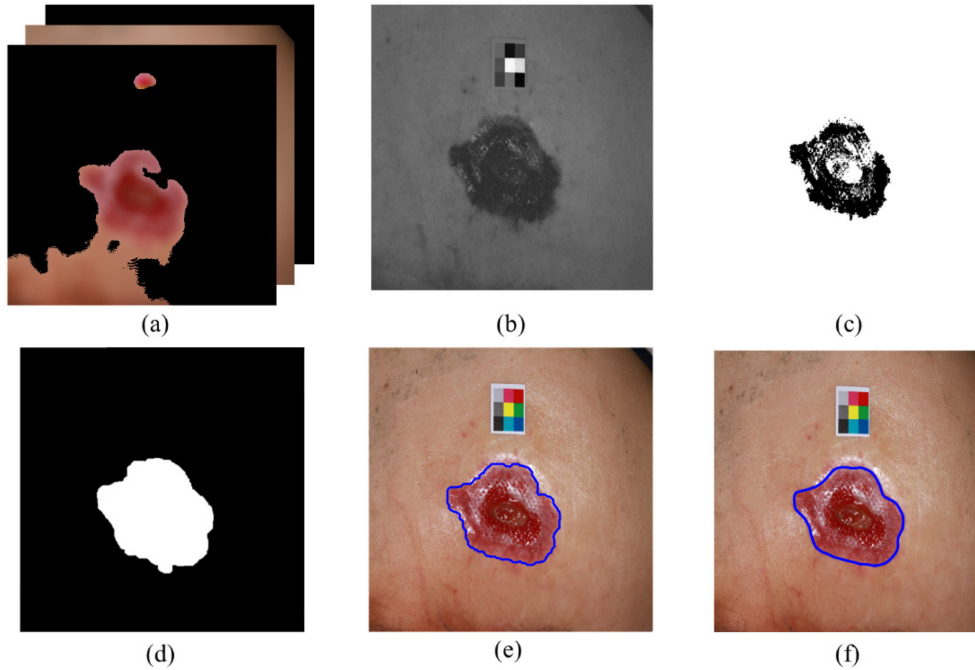


그림 3. 이미지 및 GVF 기계학습 알고리즘 (a) K-means clustering 결과 (b) LDA 특징 추출 결과 (c) Intensity 조정 결과 (d) 이진화 결과 (e) 초기 경계 영역 (f) GVF 결과

Fig. 3. Labeling using GVF machine learning algorithms (a) K-means clustering result (b) LDA feature result (c) intensity changed image (d) binarized result (e) initial boundary (f) GVF result

92

클래스 분류에서 최적의 선형 조합을 찾으려 하는 통계 방법인 Linear Discriminant Analysis (LDA)를 통해 이미지의 관심영역 및 창상 영역을 추출하였다[11,12]. 이후 이진화 과정을 통해 정확한 창상 경계를 분할하였다.

3. 창상 분할 딥러닝 알고리즘 구축

훈련에 사용된 모델은 U-net 모델에 잔차블록을 추가한 모델이다. U-net 모델은 U자 모양의 구조의 합성곱 신경망 구조로, 컨볼루션 블록 기반의 인코더와 디코더 부분으로 이루어져 있는 U자 구조이다. 영상 전체를 한 픽셀 단위로 학습하는 기존의 방법을 극복해 패치 단위로 잘라 학습해 시간 효율을 높였으며, 지역 정보 학습과 전체적인 학습 정보를 공유해 정확도를 높였다. 이에 본 연구에서는 잔차블록을 추가하

여 이전의 학습량을 제외한 잔차 정보만을 학습해 학습 효율을 높였다[13,14].

손실함수는 딥러닝 모델의 실제 정답과 예측 간의 차이를 계산하기 위해 사용되는 함수이며, 네트워크 역전파를 이용해 오류를 계산할 때 필수적이다. 위 딥러닝 모델을 위한 손실함수로 Binary Cross Entropy loss, Dice loss 더해 사용했으며, 이는 식 (1)에 표현되어 있다.

$$L_{total} = L_{Dice} + L_{Binary\ Cross\ Entropy} \quad (1)$$

$$L_{Dice} = 1 - 2 \times \frac{|G \cap P|}{|G| + |P|} \quad (2)$$

$$L_{Binary\ Cross\ Entropy} = -\frac{1}{n} \sum_{i=1}^n G_i \times \log P_i + (1 - G_i) \log(1 - P_i) \quad (3)$$

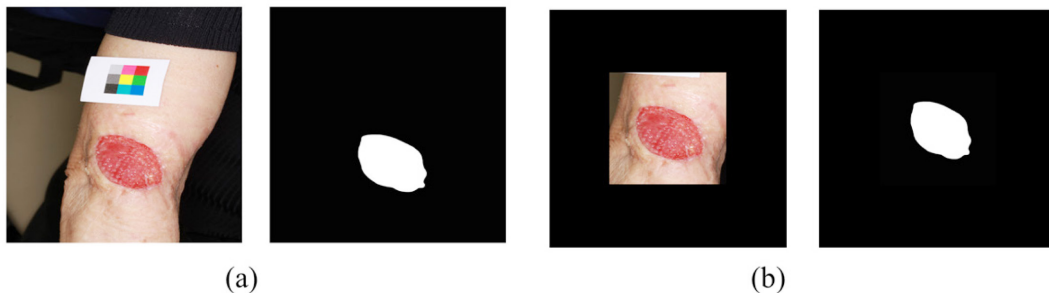


그림 4. 데이터셋 종류 (a) 원본 크기 데이터셋 (b) 창상 기반으로 자른 데이터셋
Fig. 4. Dataset Type (a) original dataset (b) wound ROI cropped dataset

Binary Cross Entropy loss와 Dice loss는 각 식 (2), (3)과 같이 기술되며, 이 때 G 는 정답, P 는 딥러닝 모델의 예측 결과를 의미한다. 최적화 함수는 학습방향(momentum)과 학습률(learning rate)을 모두 고려한 Adam에 가중치 감소를 추가한 AdamW를 사용하였다.

모델에 학습된 이미지 크기는 512,512로 두 가지 유형으로 딥러닝 학습에 훈련시켰으며, 데이터 유형의 예시는 그림 4를 통해 나타내었다. 학습에 사용한 데이터셋은 두가지 유형으로 원본 크기 데이터셋(그림 4(a))과 관심영역인 창상 기반으로 자른 후 0으로 유형 패딩한 유형(그림 4(b))이며, 이는 창상이 있는 영역을 적절한 여백(왼쪽, 오른쪽, 위, 아래 각각 10픽셀)을 두고 라벨을 기반으로 수동으로 잘라서 제작하였다.

III. 연구 결과

1. 평가지표

$$Dice\ score = 2 \times \frac{|G \cap P|}{|G| + |P|} \quad (4)$$

$$IOU = 2 \times \frac{|G \cap P|}{|G \cup P|} \quad (5)$$

딥러닝을 이용한 창상 분할 결과를 정량화 하기 위해 사용한 평가 지표는 Dice score, IOU, 손실함수의 최솟값이다. Dice score는 두 값의 유사성을 측정하는 수식으로, 두 객체의 교집합을 두 영역의 크기로 정규화한다. Dice score는 (4)와 같이 계산되며, G 는 정답, P 는 딥러닝 모델의 예측 결과이다. IOU는 두 영역의 교집합을 그들의 합집합으로 나눈 값으로, 식 (5)로 표현된다. 손실함수의 최솟값은 딥러닝의 검증 데이터셋에서의 식 (1)의 결과가 가장 작은 구간에서의 손실 값이며, 검증이 얼마나 낮은 오차율로 진행되어 있는지를 보여주는 지표이다.

2. 결과

두가지 데이터셋 유형에 따른 딥러닝 모델의 학습 결과가 표 1에 요약되어 있다. 두 Dice score와 IOU 평가지표에서 모두 'Wound cropped' 데이터셋에서 더 높은 값을 기록했다. Dice score는 'Original' 데이터셋에서 0.80에서 'Wound cropped'에서 0.89로 상승했으며, IOU는 'Original'에서 0.74에서 'Wound cropped'으로 0.80으로 증가했다. 이는 창상이 있는 부분을 중점적으로 자르고 분석하는 'Wound cropped'

방식이 창상의 위치와 형태를 더 정확하게 예측하는 것에 도움이 됨을 시사한다. 특히, Dice Score의 상대적으로 큰 증가는 작은 창상 영역에 대한 예측이 크게 개선되었음을 나타낼 수 있으며, 이는 이전 설명에서 언급된 Dice score의 특성과 일치한다. 손실 함수 값도 'Original' 데이터셋에서의 0.05에서 'Wound cropped' 데이터셋으로의 실험에서 0.04로 감소했다. 이는 'Wound cropped' 방식이 모델이 학습하는 데 더 유리하며, 이로 인해 더 정확한 예측과 낮은 오류율을 달성할 수 있음을 의미한다. 손실 값의 감소는 모델이 데이터의 패턴을 더 잘 이해하고 있으며, 결과적으로 더 정확한 예측을 제공한다는 것을 의미한다.

IV. 고찰 및 결론

본 연구에서 집중하고자 했던 문제는 일부 딥러닝 알고리즘이 특정 데이터셋에만 의존한다는 것이다. 따라서 이러한 한계를 극복하기 위해 기존의 공개 데이터셋과 자체적으로 개발한 데이터셋으로 새롭게 생성된 데이터셋으로 딥러닝 모델을 훈련하는 접근 방식을 사용했다. 데이터셋 제작과정에서 GVF 기반 머신러닝 알고리즘을 사용해 객관성과 시간 효율성을 높였다.

두가지 데이터 유형으로 딥러닝을 훈련해 데이터 전처리에 따른 딥러닝 성능을 보여준다. 창상 부위 위주로 데이터를 잘라 훈련을 시킴으로써, 모델이 불필요한 배경 정보에 의해 방해받지 않고 핵심적인 특징을 할 수 있었던 것으로 판단한다. 이러한 접근 방식은 Dice score의 상대적으로 큰 성능 향상을 확인했다. Dice score는 교집합 영역을 실제 영역 크기 합으로 정규화 하여 계산하는데, 이는 작은 객체인 창상의 크기가 작을 경우 분모가 상대적으로 작아지기 때문에 작은 교집합 영역만으로도 높은 점수를 얻을 수 있음을 의미한다. 따라서, Dice score의 상대적으로 높은 값은 창상 분할 과제에서 창상이 작더라도 훈련 결과가 정확할 수 있음을 시사한다.

결합된 데이터셋을 활용하여 창상 분할 알고리즘을 구축해 기존 문제점을 극복하고자 하였으며, 이를 통해 연구의 필요성을 입증하였다. 추후 연구에서는 데이터셋을 더욱 확대해 임상 적용성을 향상하여 연구 범위와 영향력을 확대하고자 한다. 다양한 환경과 인종에 대한 추가 데이터셋을 제작해 알고리즘을 추가 검증 및 보완시킬 계획이다.

References

- [1] Sen CK, Gordillo GM, Roy S, Kirsner R, Lambert L, Hunt TK, Gottrup F, Gurtner GC, Longaker MT. Human skin wounds: a major and snowballing threat to public health and the economy. *Wound Repair Regen.* 2009;17(6):763-71.
- [2] Marijanovic D, Filko D. A Systematic Overview of Recent

표 1. 창상 분할 결과

Table 1. Wound segmentation results

| Dataset type | Dice | IOU | L_{total} |
|---------------|------|------|-------------|
| Original | 0.80 | 0.74 | 0.05 |
| Wound cropped | 0.89 | 0.80 | 0.04 |

- Methods for Non-Contact Chronic Wound Analysis. *Appl Sci Basel*. 2020;10(21).
- [3] Gurkan A, Kirtil I, Aydin YD, Kutuk G. Pressure injuries in surgical patients: a comparison of Norton, Braden and Waterlow risk assessment scales. *J Wound Care*. 2022;31(2):170-7.
- [4] Khoo R, Jansen S. The Evolving Field of Wound Measurement Techniques: A Literature Review. *Wounds*. 2016; 28(6):175-81.
- [5] Shah A, Wollak C, Shah JB. Wound Measurement Techniques: Comparing the Use of Ruler Method, 2D Imaging and 3D Scanner. *J Am Coll Clin Wound Spec*. 2013;5(3):52-7.
- [6] Lu H, Li B, Zhu J, Li Y, Li Y, Xu X, He L, Li X, Li J, Serikawa S. Wound intensity correction and segmentation with convolutional neural networks. *Concurr Comput Pract Exp*. 2017;29(6).
- [7] Wang C, Yan X, Smith M, Kochhar K, Rubin M, Warren SM, Wrobel J, Lee HL. A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. *Annu Int Conf IEEE Eng Med Biol Soc*. 2015;2015:2415-8.
- [8] Scebbba G, Zhang J, Catanzaro S, Mihai C, Distler O, Berli M, Karlen W. Detect-and-segment: A deep learning approach to automate wound image segmentation. *Inform Med Unlock*. 2022;29:100884.
- [9] Xu T, Mandal M, Long R, Basu A. Gradient vector flow based active shape model for lung field segmentation in chest radiographs. *Annu Int Conf IEEE Eng Med Biol Soc*. 2009;2009:3561-4.
- [10] Yang SJ, Park JH, Lee HN, Kim SH, Lee BU, Chung KY, Oh BH. Sequential change of wound calculated by image analysis using a color patch method during a secondary intention healing. *PLoS One*. 2016;11(9):e0163092.
- [11] Xanthopoulos P, Pardalos PM, Trafalis TB. *Robust data mining*. New York: Springer Science & Business Media; 2012.
- [12] Cho IH, Oh J, Om KS, Song I, Chang KH, Jeong DS. Preprocessing effect by using k-means clustering and merging algorithms in MR cardiac left ventricle segmentation. *J Biomed Eng Res*. 2003;24.
- [13] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Cham: Springer International Publishing; 2015.
- [14] Koonce B, Koonce B. ResNet 50. In: *Convolutional Neural Networks with Swift for TensorFlow: Image Recognition and Dataset Categorization*. 2021;63-72.