

# Deep Learning based Brachial Plexus Ultrasound Images Segmentation by Leveraging an Object Detection Algorithm

Kukhyun Cho<sup>1</sup>, Hyunseung Ryu<sup>1</sup>, Myeongjin Lee<sup>1</sup>, Suhyung Park<sup>2,\*</sup>

<sup>1</sup>Department of Computer Engineering, Chonnam National University

<sup>2</sup>Department of Electronics and Computer Engineering, Chonnam National University

Received: July 22, 2024. Revised: August 28, 2024. Accepted: October 31, 2024.

## ABSTRACT

Ultrasound-guided regional anesthesia is one of the most common techniques used in peripheral nerve blockade by enhancing pain control and recovery time. However, accurate Brachial Plexus (BP) nerve detection and identification remains a challenging task due to the difficulty in data acquisition such as speckle and Doppler artifacts even for experienced anesthesiologists. To mitigate the issue, we introduce a BP nerve small target segmentation network by incorporating BP object detection and U-Net based semantic segmentation into a single deep learning framework based on the multi-scale approach. To this end, the current BP detection and identification was estimated: 1) A RetinaNet model was used to roughly locate the BP nerve region using multi-scale based feature representations, and 2) U-Net was then used by feeding plural BP nerve features for each scale. The experimental results demonstrate that our proposed model produces high quality BP segmentation by increasing the accuracies of the BP nerve identification with the assistance of roughly locating the BP nerve area compared to competing methods such as segmentation-only models.

Keywords: Brachial Plexus, Ultrasound Image, Object Detection, Image Segmentation

## I. INTRODUCTION

상완 신경총(BP; Brachial Plexus)은 척수에서 생겨나 일부는 목에, 일부는 겨드랑이에 위치하는 경추 신경의 일부이다. 상완 신경총 영역의 분할(Segmentation)은 폐암이나 두경부암 환자의 치료 계획 수립에서 중요한 고려 사항이다<sup>[1]</sup>. 암 환자를 치료하기 위해 사용되는 방사선 요법이 BP에 과도한 방사선 에너지를 가하면, 상완 신경총 병증(Brachial Plexopathy)을 유발하여 심각하고 되돌릴 수 없는 영향을 줄 수 있기 때문이다. 또한, BP는 쇄골 수술에 있어 지역 마취(RA; Regional Anesthesia)에 자주 사용되며, 이는 일반 마취를 대체하여 통증 조절을 개선하고, 아편유사제(Opioid) 소비를 줄이며, 회복 시간을 단축할 수 있다. 고전적인 RA

절차는 대상 신경 영역에 마취약을 정밀한 위치 확인 없이 주입하는 방식이지만, 이 방법은 차단 실패, 신경 손상, 국소 마취약 독성과 같은 위험을 내포한다. 초음파를 이용한 지역 마취(UGRA; Ultrasound-Guided RA)는 이 절차를 시각화할 수 있는 수요가 많은 방법으로, 더 빠르고 시도 횟수가 적으며, 일부 경우에는 다른 RA 기술에 비해 더 나은 감각 차단을 가능하게 한다<sup>[2]</sup>.

초음파(US; Ultrasound) 영상에서의 조직 분할은 자기공명영상(MRI; Magnetic Resonance Imaging)이나 컴퓨터단층촬영(CT; Computed Tomography)과 같은 다른 이미징 모달리티들에 비해 영상에 나타나는 조직들 사이의 신호대 잡음비(CNR; Contrast-to-Noise Ratio)가 낮아서 여전히 어려운 작업으로 남아있다<sup>[3]</sup>. 특히 신경 조직의 분할은 초음파 기법 특유의 스펙

\* Corresponding Author: Suhyung Park

E-mail: suhyung@jnu.ac.kr

Tel: +82-62-530-1797

클 노이즈(Speckled Noise)와 신경 영역이 영상에서 두드러지는 구조를 형성하지 않기 때문에 쉽지 않은 작업이다. 이러한 어려움으로 인해 관심 영역(ROI; Region Of Interest)을 인식할 수 있는 능숙한 전문가가 필요하며, 이는 UGRA 절차의 적용 가능성을 낮게 제한하고 있다. 하지만, 딥러닝 알고리즘의 등장은 임상 전문가를 대신하여 BP의 ROI를 자동으로 인식 및 구별할 수 있는 가능성을 제시함으로써 UGRA를 더 광범위하게 사용할 수 있는 가능성을 높일 수 있다.

딥러닝 알고리즘들 중에서, U-Net은 의료 영상 분할을 위한 주요 모델 구조로서, 인코더와 디코더 부분으로 구성되어 있으며, 1) 인코더는 일반 합성곱 신경망(CNN; Convolutional Neural Network)과 유사하며, 연속적인 합성곱, 활성화 및 풀링 계층으로 이루어져 있고, 2) 디코더는 인코더의 대칭이며, 업샘플링 계층과 인코더 계층의 특징 맵을 추가하는 연결 계층, 그리고 이어지는 합성곱 계층으로 구성되어 있다<sup>[4]</sup>. U-Net 구조는 비교적 작은 학습 세트로 훈련되었을 때도 관심 영역(ROI)을 정확하게 지역화할 수 있는 능력이 입증되어왔다. 또한, Multi-scale 기반의 U-Net (M-Net)은 다양한 계층에서 나오는 다중 스케일 입력과 다중 출력을 사용하여 추출된 특징 맵을 밀도 높게 처리하는 방식으로 U-Net의 성능을 개선하려는 시도 중 하나로써 사용되어 왔다<sup>[5]</sup>.

BP의 위치를 정확하게 특정하기 위해 객체 탐지 기술들이 활용될 수 있다. 그 중에서, RetinaNet은 다른 경쟁 방법에 비해 여러 가지 이점을 제공한다<sup>[6-12]</sup>. 첫째, Focal Loss 함수를 도입해 어려운 객체 감지 문제에 효과적으로 대응하며, 초음파 이미지에서 배경 대비 객체의 비율이 낮은 불균형한 데이터 상황에서도 특정 부위를 정확하게 탐지할 수 있다. 둘째, 이 모델은 정확도와 속도 사이에 우수한 균형을 제공하여, 실시간 또는 거의 실시간으로 BP를 감지해야 하는 의료 환경에서 큰 유용성을 가진다. 셋째, FPN(Feature Pyramid Network)을 활용해 다양한 크기의 객체를 효과적으로 감지할 수 있어, 환자마다 다른 BP의 크기나 모양을 잘 처리할 수 있는 장점이 있다. 마지막으로, RetinaNet은 기존의

많은 객체 탐지 모델들에 비해 상대적으로 적은 학습 시도로도 높은 성능을 달성할 수 있으며, 이는 시간과 자원이 제한적인 의료 분야에서 특히 중요한 이점을 가진다.

하지만, 위에선 제안된 기존의 연구들은 U-Net / M-Net과 같은 영상 분할 알고리즘이나 혹은 YOLO와 RetinaNet과 같은 객체 검출 알고리즘의 단일 특성을 활용하여 관찰하고자 하는 조직을 특정하였다. 이런 단일 알고리즘은 영상의 전역 정보로부터 조직의 개별적인 특성을 추출함으로써 정확한 목표 영역의 특성을 시각적으로 표현하는 데 한계가 존재하였다. 또한, 이를 극복하고자 Squeeze and Excitation Network (SE-Net) 혹은 Convolutional Block Attention Module (CBAM)과 같은 Attention 기반의 알고리즘들이 기존의 객체 검출이나 영상 분할 알고리즘에 통합되어 전역 정보로부터 지역적인 특성을 추출함으로써 한계를 극복하고자 하였다<sup>[13,14]</sup>. 하지만, 알고리즘에 활용되는 채널 수의 증가 및 픽셀 단위의 Attention 파라미터들의 추정 은 많은 수의 Unknown 파라미터로 인하여 학습에 필요한 충분한 양의 데이터 확보가 필요하고 또한 모델의 복잡도를 증가시킴으로써 학습 및 레이블 데이터의 관리 및 알고리즘 연산의 효율성 측면에서 명확한 한계를 보였다.

본 논문에서는 초음파 이미지의 낮은 CNR과 복잡한 배경 속에서도 BP의 위치를 효과적으로 특정할 수 있는 객체 검출 알고리즘과 영상의 분할 알고리즘을 통합한 Two-stage 기반의 딥러닝 알고리즘을 제안한다. 이를 위해, 1) 전역 정보로부터 지역 정보의 특성을 추출하기 위해 RetinaNet을 활용하여 영상으로부터 관심 영역(ROI)을 제한하고, 2) BP와 관련성이 낮은 영역을 제거함으로써 제한된 영상의 범위에서 분할 기법(U-Net/M-Net)이 적용됨으로써 BP 검출의 정확도를 향상시킨다. 이는, 기존 방법들과 비교 시, 영상의 전역 영역에서 지역 영역으로 범위를 제안하여 학습에 필요한 파라미터의 수를 줄이고, 따라서 적은 Epoch에서 학습을 수렴 가능하게 함으로써 학습에 필요한 연산량을 효과적으로 줄였다. 따라서, 본 논문에 제안된 알고리즘은 정확도와 연산량 사이에서 발생할 수 있는

trade-off를 최소화하여 임상적인 활용도를 높일 수 있고, 이는 UGRA 절차의 정확도와 효율성을 크게 향상시킴으로써 UGRA의 활용 범위를 넓힐 수 있을 것으로 기대된다. 실험적인 검증은 본 알고리즘이 기존의 단일 알고리즘과 비교 시 우수한 성능을 달성할 수 있음을 확인하였다.

## II. MATERIAL AND METHODS

### 1. BP 검출 with RetinaNet

Fig. 1-(a)는 RetinaNet의 구조를 나타내고 Eq. (1)을 통해 클래스 불균형 문제를 해결한다. Focal Loss는 클래스에 대한 확신이 증가함에 따라 스케일링 인자가 0으로 감소하는 동적으로 스케일링된 교차 엔트로피 손실 함수이다. Focal Loss에서 는 클래스의 불균형을 조정하기 위한 가중치이고 는 조정 파라미터로, 잘 분류된 예제에 대한 손실을 얼마나 감소시킬지 조정하는 파라미터이다<sup>[9]</sup>.

$$Focal\ Loss(p_t) = -\alpha(1-p_t)^\gamma \log(p_t) \quad (1)$$

RetinaNet은 단일 통합 네트워크로, 하나의 백본 네트워크와 두 개의 Task별 하위 네트워크로 구성된다. 백본 네트워크로 FPN이 사용된다. 이는 ResNet 구조 위에서 구축된다. 두 개의 하위 네트워크는 객체 분류 하위 네트워크(Classification Sub Network)와 바운딩 박스 하위 네트워크 (Box Regression Sub Network)로 구성되어 있다. 분류 하위 네트워크(Classification Sub Network)는 주어진 피라미드 레벨에서 C 채널을 가진 입력 특징 맵을 받아, 이 서브넷은 C개의 필터를 가진 네 개의 3×3 CNN 층을 적용하고, 각각에 ReLU 활성화 함수를 사용한다. 그 후 K×A개의 필터를 가진 3×3 합성곱 층을 적용하고, 마지막으로 각 공간 위치마다 K×A개의 이진 예측을 출력하기 위해 시그모이드 활성화 함수를 적용한다.

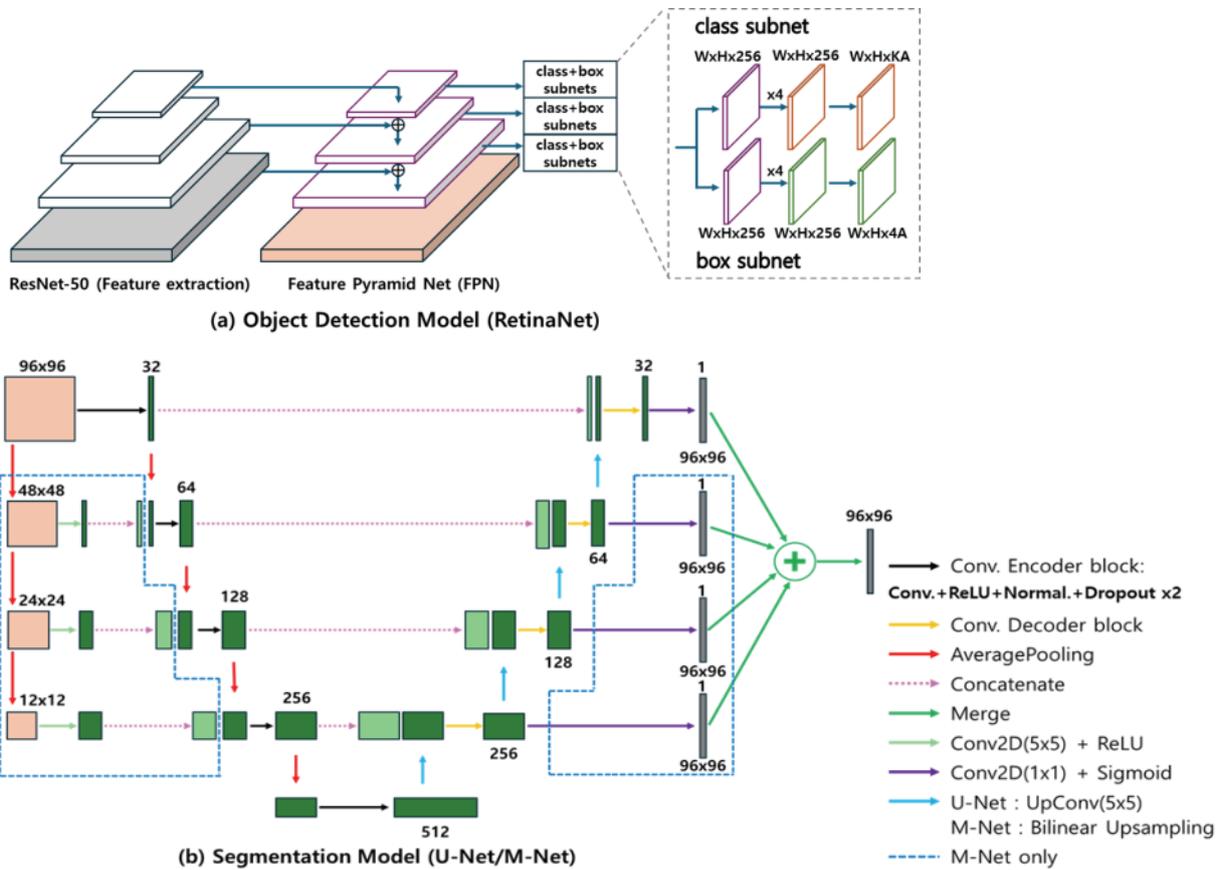


Fig. 1. The architecture of (a) RetinaNet for object detection and (b) U-Net and M-Net for the segmentation.

Table 1. The value of hyper-parameters for RetinaNet, U-Net, M-Net architectures

	RetinaNet	U-Net	M-Net
Optimizer	Adam	RMSprop	RMSprop
Batch Size	8	128	128
Learning Rate	$10^{-4}$	$10^{-4}$	$3 \times 10^{-4}$
Epoch	100	80	80
Kernel Size	3×3	5×5	5×5

바운딩 박스 하위 네트워크는 객체 분류 하위 네트워크와 병렬로 연결되어 있는 Sub Network로서 각 피라미드의 Scale-wise로 연결되어 있는 FCN (Fully Convolutional Network) 구조의 하위 네트워크이다. 이 하위 네트워크의 목적은 해당 객체가 존재하는 경우, 각 앵커 박스에 가까운 실제 객체로의 오프셋을 탐지하는 것이다. 바운딩 박스 하위 네트워크의 설계는 분류 하위 네트워크와 동일하지만, 아웃풋으로 A개의 바운딩 박스 위치 정보 (x,y,w,h)를 출력하고, 이를 바탕으로 실제 바운딩 박스와의 상대적 오프셋을 예측하는 데 사용된다<sup>[15]</sup>.

## 2. BP 분할 with U-Net

첫 번째 분할 모델은 U-Net 구조를 사용한다 (Fig. 1-(b)). 모델의 인코더 부분은 5층의 CNN으로 이루어져 있으며 각각의 층은 ReLU(Rectified Linear Unit) 활성화 함수를 가진 5×5 크기의 CNN 2회, 정규화, 40% 드롭아웃으로 구성되어 있다. 첫 번째 계층에서 필터의 수는 32이며 이후 각 계층에서 필터의 수는 두 배씩 증가하고 평균 풀링을 사용한다. 디코더는 같은 CNN 블록을 사용하지만 블록 내부에서 드롭아웃은 적용하지 않으며 업샘플링 과정에는 전치 컨볼루션(transposed convolution)을 사용한다. 최종 출력은 시그모이드 활성화 함수를 가진 1×1 크기의 CNN을 통해 출력된다.

두 번째 분할 모델은 M-Net 구조를 사용한다 (Fig. 1-(b)). M-Net은 U-Net의 확장으로 입력을 하나만 받는 U-Net과 달리 여러 입력을 받는다. 원본 입력의 2배, 4배, 8배로 다운샘플링된 입력을 U-Net 구조의 인코더 부분에서 이전 층의 특징 맵과 연결하여 후속 층으로 전송된다. 디코더의 각 층에서

시그모이드 활성화 함수를 가진 1×1 CNN을 통해 출력 이미지로 만들고 그 출력을 원본 입력의 형태와 일치하도록 업샘플링 한다. 4개의 출력 이미지는 1개의 출력을 위해 평균화한다. 내부 구조는 U-Net과 유사하나 업샘플링 과정에서 양선형 업샘플링(bilinear upsampling)을 사용한다. 사용된 하이퍼파라미터는 Table 1 에서 제공한다. 구현을 위해 Tensorflow 라이브러리를 사용하였으며 Intel Core i9-10900X CPU, NVIDIA GeForce RTX 3090, 251 GB RAM 환경에서 실험하였다.

## 3. Datasets

본 논문은 [16]의 데이터를 사용한다. 이 데이터셋은 목 부위의 초음파 사진으로 47명의 대상자에게서 얻어졌다 (Fig. 2).

대상자당 119-120개의 사진이 얻어졌으며 모든 사진은 420 × 580 픽셀의 해상도를 가지고 있다. 5,635개의 초음파 사진 중 BP를 담고 있지 않은 사진을 제거하여 총 2,323개의 사진으로 실험을 진행하였으며 각 사진에 대해서 객체 탐지 모델의 정답 데이터로 BP 위치를 탐지한 바운딩 박스 좌표와 분할 모델의 정답 데이터를 마스크 이미지로 구성하였다.

분할 모델의 입력으로 들어가기 전 데이터는 96 × 96으로 다운샘플링 되고 평균이 0, 표준 편차는 1로 설정된다. 훈련 데이터로 2,100개의 이미지, 테스트 데이터로 223개의 이미지를 사용하였으며 훈련 데이터를 8:2로 훈련과 검증 데이터로 분할하였다.

## 4. Experimental setup

BP 분할의 성능에 하이브리드 모델의 영향을 조사하기 위해 3개의 실험을 설정하였다.

실험 1) Without detection Model : 객체 탐지 모델이 없는 분할 모델(U-Net/M-Net)

실험 2) Hybrid Model(with Detection) : RetinaNet으로 BP 탐지 후 분할(U-Net/M-Net)

실험 3) Perfect Detection Model : 완벽한 객체 탐지 후 분할(U-Net/M-Net)

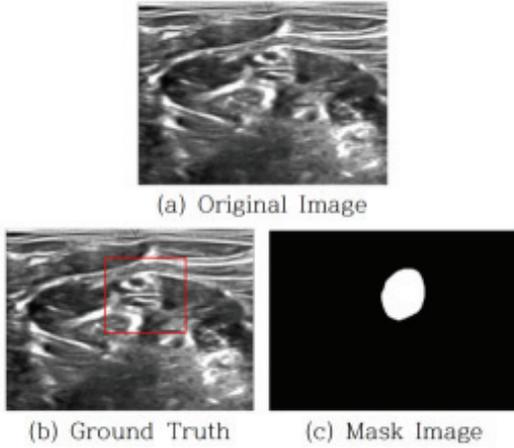


Fig. 2. Dataset consisting of (a) Ultrasound image containing BP, (b) GT bounding box for Object detection, and (c) Mask image for Segmentation.

### 5. Evaluation Metrics

객체 탐지 후 평가를 위해서 Eq. (2)을 사용하였다. IoU(Intersection Over Union)의 값을 통해 실제 객체의 위치와 모델이 예측한 바운딩 박스가 얼마나 정확하게 겹치는지를 수치화한 지표이다. 이 값은 0에서 1 사이의 값을 가지며, 1에 가까울수록 실제 객체와 예측된 바운딩 박스의 겹치는 정도가 높다는 것을 의미한다.

분할 결과의 평가를 위해 Eq. (3)을 사용하였다. DSC (Dice Similarity Coefficient)는 정답 이미지와 대응하는 예측 이미지를 픽셀별로 비교한다. 0에서 1 사이의 값을 가지며 0은 겹치는 부분이 없다는 것이고 1은 완벽히 겹치는 것을 의미한다.

$$IoU = \frac{|M \cap GT|}{|M \cup GT|} \quad (2)$$

$$DSC = \frac{2 \cdot |M \cap GT|}{|M| + |GT|} \quad (3)$$

## III. RESULT

Fig. 3은 Focal Loss의 포커싱 파라미터  $\gamma$  값에 따라 손실이 어떻게 변하는지 보여준다. 이 그래프는 다양한  $\gamma$  값( $\gamma = 1, 2, 3, 5$ )에 대해 검증 손실의 변

Table 2. Comparison of mean IoU (mIoU) for various gamma values ( $\gamma=1, 2, 3, 5$ )

Gamma	mIoU
$\gamma = 1$	0.6794
$\gamma = 2$	0.6691
$\gamma = 3$	0.6845
$\gamma = 5$	0.6858

화를 나타낸다.  $\gamma = 0$ 인 경우, Focal Loss는 교차 엔트로피 손실 함수와 같다.  $\gamma$  값이 증가할수록 모델이 쉽게 분류할 수 있는 예제에 대해서는 손실을 줄이고, 더 어려운 예제에 대해서는 더 큰 손실을 부여하여 어려운 예제에 더 집중할 수 있는 효과를 보여준다. 그래프를 보면 초기에 모든  $\gamma$  값에서 Loss가 급격하게 감소한다. 초기 손실을 보면  $\gamma$ 의 값이 클수록 Loss가 작아지는 경향을 볼 수 있다. 10에서 50 Epoch에서는  $\gamma$  값이 클수록 손실이 더 낮아지고 안정화되는 모습을 볼 수 있다.  $\gamma$ 가 1일 때에는 가장 높은 Loss 값을 가지고,  $\gamma$ 가 5일 때는 가장 낮은 Loss를 유지하고 있다. 이후에는 모든  $\gamma$  값에서 Loss가 거의 일정하게 유지된다.  $\gamma$  값이 클수록 더 낮은 Loss 값으로 수렴하는 것을 확인할 수 있다. 이 그래프를 통해  $\gamma$  값이 클수록 잘 분류된 예제에 대한 손실 기여도가 줄어들어 모델이 어려운 예제에 집중하여 낮은 Loss 값을 갖는 것을 볼 수 있다.

Focal Loss에서  $\gamma$ 의 값은  $[0, 5]$ 에 대한 값을 사용하며, 일반적으로  $\gamma$ 는 2를 많이 사용하지만, 이 실험에서는  $\gamma$ 가 5일 때 더 좋은 성능을 보인다.

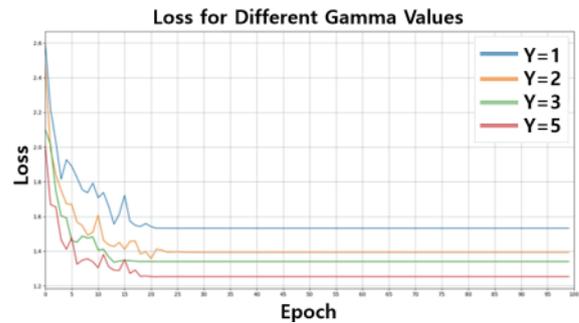


Fig. 3. Loss curves over 100 epochs by varying gamma values ( $\gamma=1, 2, 3, 5$ ) during object detection.

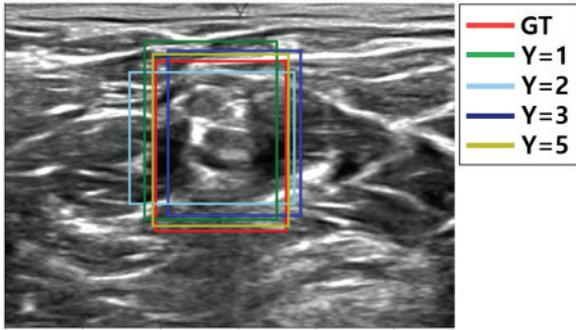


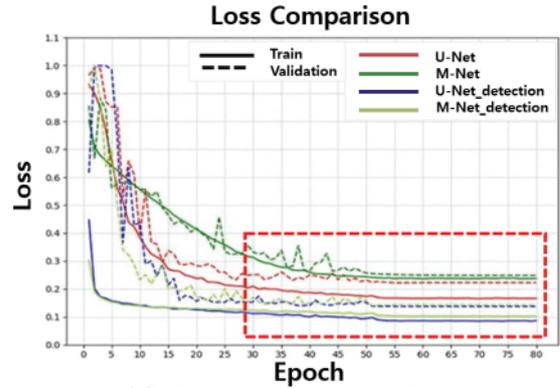
Fig. 4. Comparison of object detection results for different gamma values ( $\gamma=1, 2, 3, 5$ ) compared to GT: Visualization of predicted bounding boxes and GT bounding box on the ultrasound image.

Table 2는 다양한 감마 값에 따른 mIoU 값을 도출한 결과이다. 실험적으로  $\gamma$ 가 5일 때 가장 높은 mIoU 값을 가지고  $\gamma$ 가 2일 때 가장 낮은 mIoU 값을 가진다.

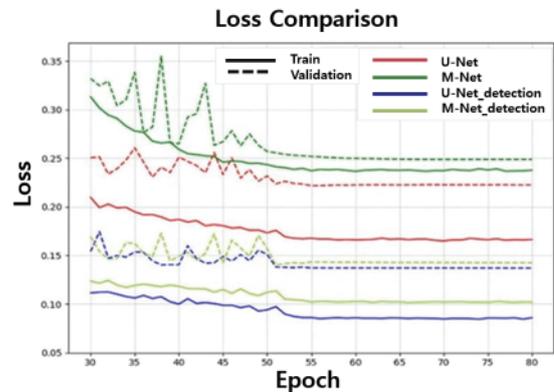
Fig. 4는 각  $\gamma$ 에 따른 바운딩 박스가 GT와 얼마나 일치하는지 비교하는 그림이다.  $\gamma$ 가 2일 때는 다른  $\gamma$  값들에 비교하여 GT와의 오차가 최대임을 확인할 수 있다. 또한  $\gamma$ 가 5일 때가 GT와 가장 비슷한 위치에 있음을 확인할 수 있다. 이를 통해 Table 2의 결과처럼  $\gamma$ 가 5일 때 높은 정확도로 BP를 감지한 것을 볼 수 있다.

Fig. 5는 학습 과정에서 Epoch에 따른 Train/Validation Loss의 변화를 나타내며 (a)는 4가지 경우의 전체 Loss를 Epoch에 따라 비교한 것이다. 모든 경우에서 Loss가 감소하는 추세를 보인다. (b)에서 각 모델의 부분 Loss 추이를 살펴보면, w/ Detection Model이 w/o Detection Model에 비해 w/o Detection Model에 비해 더 빠른 속도로 Loss가 감소하는 경향을 확인할 수 있다. 이는 Detection 기법을 적용함으로써 모델이 중요한 특징을 더 효과적으로 학습할 수 있기 때문으로 해석된다.

Fig. 6은 Test Data에 대한 U-Net 계열 모델의 정성적 지표 GT(Ground Truth)와 U-Net, U-Net-detection, U-Net-Perfect detection 모델의 분할 결과를 보여준다. 녹색 영역은 모델이 예측한 관심 영역을 나타낸다. 전반적으로 U-Net-Perfect detection 모델이 GT에 가장 근접한 결과를 보이고, 그 다음으로 U-Net-detection 모델이 그 다음으로 근접한 결과를 보인다.



(a) Overall Loss Comparison



(b) Partial Loss Comparison

Fig. 5. Train/Validation Loss curves during segmentation training (a) 0~80 Epoch (b) 30~80 Epoch.

U-Net 모델은 상대적으로 가장 낮은 성능을 보이는 것으로 관찰된다.

Table 3은 Test Data에 대한 U-Net과 M-Net 모델의 DSC 실험 결과이다. U-Net 모델의 경우, Detection 정보를 사용하지 않은 모델(w/o Detection Model)의 DSC는 0.7192인 반면, Detection 정보를 사용한 모델(w/ Detection Model)의 DSC는 0.7957로 눈에 띄는 향상을 보인다. 더 나아가 완벽한 Detection 정보를 사용한 경우(Perfect Detection)에는 DSC가 0.8454까지 상승한다.

Table 3. Experiment results (DSC) on Test data

	U-Net	M-Net
w/o Detection Model	0.7192	0.6892
w/ Detection Model	0.7957	0.7720
Perfect Detection	0.8454	0.8384

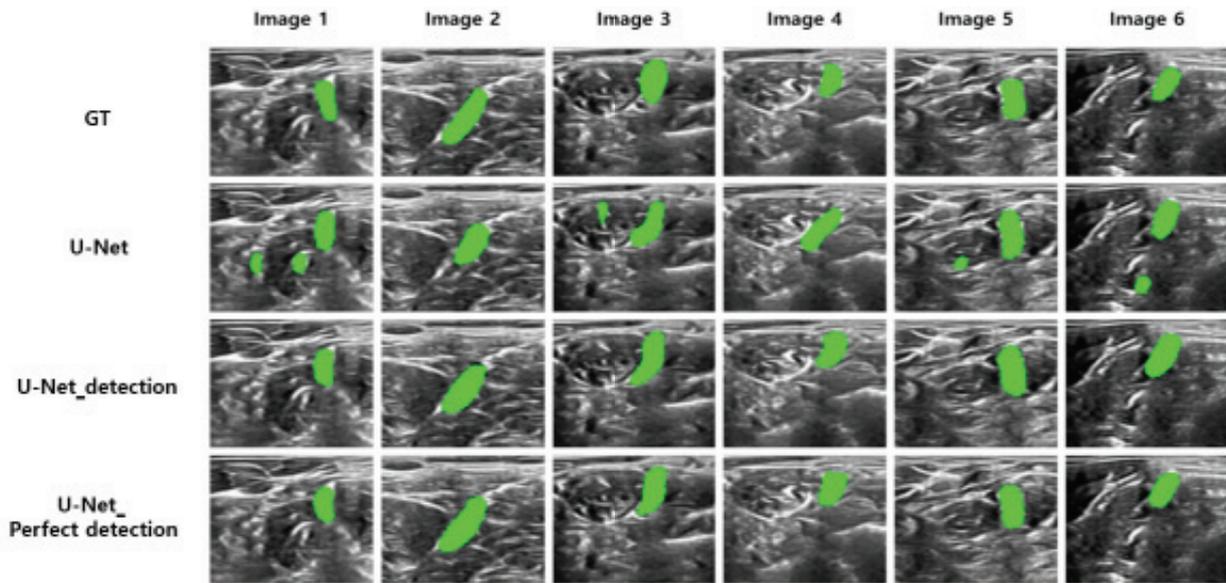


Fig. 6. Comparison of segmentation results of U-Net, U-Net\_Detection, and U-Net\_Perfect detection models on test data.

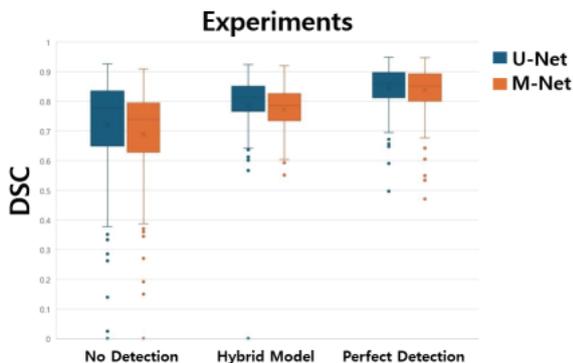


Fig. 7. Boxplots for quantitative analysis w/ and w/o object detection model.

M-Net 모델 역시 비슷한 경향성을 보인다. Detection 정보를 사용하지 않은 모델의 DSC는 0.6892에 그치지만, Detection 정보를 활용한 모델에서는 0.7720으로 크게 향상된다. Perfect Detection의 경우 DSC는 0.8384까지 상승한다. 전반적으로 U-Net 모델이 M-Net 모델에 비해 약간 높은 DSC를 보이고 있지만, 두 모델 모두 Detection 정보를 활용함으로써 분할 성능을 효과적으로 개선할 수 있음을 확인할 수 있다.

Fig. 7은 Detection 모델 사용 여부에 따른 U-Net과 M-Net 모델의 정량적 분할 성능을 Boxplot으로

비교한다. No Detection, Hybrid Model, Perfect Detection의 세 가지 실험조건에서 각 모델의 DSC 분포를 확인할 수 있다. 전반적인 경향성은 Table 3에서 확인한 바와 같이 No Detection, Hybrid Model, Perfect Detection 순으로 성능이 향상되는 것을 볼 수 있다. U-Net과 M-Net 모두 Detection 정보를 활용할수록 DSC의 중앙값과 사분위수 범위가 점차 증가하며, 이는 Detection 정보의 활용이 분할 성능 향상에 기여함을 나타낸다. 특히 Perfect Detection의 경우 두 모델 모두 DSC의 중앙값이 상승하고 분포의 범위가 좁아지는 것을 확인할 수 있다. 이는 Detection 정보의 정확도가 상승할수록 안정적이고 높은 수준의 분할 성능을 얻을 수 있음을 의미한다.

#### IV. DISCUSSION & CONCLUSIONS

이 연구의 목표는 초음파 사진에서 BP 신경 구조의 이미지 분할을 개선하는 것이었으며 이를 위해 Detection Model을 추가하여 Hybrid Model을 사용하였다. RetinaNet이 객체 탐지 모델로 사용되었으며, U-Net과 M-Net이 분할 모델로 사용되었다. 첫 번째 실험은 사전 객체 탐지를 수행하지 않았고 두 번째 실험에서는 RetinaNet을 객체 탐지에 사용

했으며 세 번째 실험에서는 BP 지역을 잘라내어 완벽한 객체 탐지 모델을 시뮬레이션하였다.

실험 결과에서 사전 Detection Model이 추가되었을 때 모든 영상의 분할 모델들의 성능이 향상됨을 보여주었으며, 이는 Detection Model이 Segmentation 성능 향상에 기여하며 Detection 성능과 Segmentation 성능이 비례함을 시사하였다.

이미지 분할은 각 픽셀을 주어진 클래스로 분류하는 작업이다. BP 신경 구조의 이미지 분할은 각 픽셀을 BP에 해당하는 부분인지 아닌지로 분류하는 작업으로 생각할 수 있다. 원본 이미지에서는 BP 부위가 배경에 비해 작아 클래스들 간의 불균형이 발생하였고 이는 Detection Model을 이용한 데이터의 전처리를 통하여 클래스 불균형을 해소할 수 있었다. 따라서, No Detection Model과 비교하여 with Detection의 Segmentation 모델의 성능이 향상된 것으로 해석할 수 있다.

본 논문에서는 두 가지 분할 모델의 경우에 대해서 고려하였고 두 모델의 전반적인 성능의 향상을 확인하였지만, 다음과 같은 한계가 존재한다. 1) 실험에서는 객체 탐지 후 IoU 값이 0.7 이상인 이미지들을 선별하여 Segmentation을 진행하였다. IoU 값을 특정 임계값으로 설정하지 않을 경우 U-Net (M-Net)을 단독으로 사용한 경우와 비슷한 성능이 가진다는 한계가 있었고, 이를 통해 Segmentation의 성능이 Detection 성능에 영향을 많이 받는다는 것을 알 수 있다. 따라서, 향후 알고리즘의 광범위한 활용을 위해서는 BP 영역을 넘어서 다양한 부위에서의 Detection의 IoU 임계값과 Segmentation의 의존도에 대한 정량적인 정확도 분석을 통하여 활용 임계값을 설정할 필요가 있다. 또한, 성능이 높은 Detection Model과 비례하여 Segmentation 성능 또한 높아지는 특성을 활용하여 RetinaNet 모델 대신 Advanced된 Detection 모델을 활용한다면 IoU 임계값에 예민하지 않는 강건한 모델을 구축할 수 있을 것으로 사료된다; 그리고 2) 또한 하이브리드 모델을 활용함으로써 목표 조직을 예측하는 네트워크의 forward problem과정에서 여전히 많은 연산량을 포함하고 있다. 이는, 실제로 GPU 장착되어 있지 않은 많은 의료영상기기에서는 연산의 복잡도로

인하여 활용되기 힘든 어려움이 존재한다. 이를 해결하기 위해, Segmentation에서 사용되는 Convolution 과정을 주파수 도메인으로 영상을 변환하여 point-wise Convolution 연산을 적용함으로써 연산의 복잡도를 효과적으로 줄여 임상으로의 활용 가능성을 높일 수 있을 것으로 예상된다.

또한, 본 연구 결과를 바탕으로 향후에는 Two-Stage 방식으로 학습된 모델이 아닌 One-Stage 방식의 Hybrid Model을 이용하여 Detection Model과 Segmentation Model이 서로 Interactive Learning을 통하여 Complementary features를 학습을 하게 된다면 IoU의 임계값에 대해 강건한 성능을 보여줄 것으로 기대된다.

## Acknowledgement

이 논문은 전남대학교 학술연구비 (과제번호: 2022-2579) 지원에 의하여 연구되었음.

## References

- [1] J. Yang, A. Amini, R. Williamson, L. Zhang, Y. Zhang, R. Komaki, Z. Liao, J. Cox, J. Welsh, L. Court, L. Dong, "Automatic Contouring of Brachial Plexus Using a Multi-Atlas Approach for Lung Cancer Radiation Therapy", *Practical Radiation Oncology*, Vol. 3, No. 4, pp. 139-147, 2013. <http://dx.doi.org/10.1016/j.prro.2013.01.002>
- [2] S. S. Liu, J. E. Ngeow, J. T. Yadeau, "Ultrasound-Guided Regional Anesthesia and Analgesia: A Qualitative Systematic Review", *Regional Anesthesia & Pain Medicine*, Vol. 34, No. 1, pp. 47-59, 2009. <http://dx.doi.org/10.1097/AAP.0b013e3181933ec3>
- [3] J. P. A. van Boxtel, V. R. J. Vousten, J. Pluim, N. M. Rad, "Hybrid Deep Neural Network for Brachial Plexus Nerve Segmentation in Ultrasound Images", 2021 29th European Signal Processing Conference, Dublin, Ireland, pp. 1686-1690, 2021. <http://dx.doi.org/10.23919/EUSIPCO54536.2021.9616329>
- [4] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", 2015 Medical image computing and computer- assisted intervention MICCAI: 18th

- international conference, pp. 234-241, 2015.  
<https://doi.org/10.48550/arXiv.1505.04597>
- [5] N. Abraham, K. Illanko, N. Khan, D. Androutsos, "Deep Learning for Semantic Segmentation of Brachial Plexus Nerves in Ultrasound Images Using U-Net and M-Net", 2019 3rd International Conference on Imaging, Signal Processing and Communication, pp. 85-89, 2019.  
<https://doi.org/10.1109/ICISPC.2019.8935668>
- [6] D. Erhan, C. Szegedy, A. Toshev, D. Anguelov, "Scalable Object Detection Using Deep Neural Networks", 2014 The IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2147-2154, 2013.  
<https://doi.org/10.48550/arXiv.1312.2249>
- [7] S. Bell, C. L. Zitnick, K. Bala, R. Girshick, "Inside-Outside Net: Detecting Objects in Context with Skip Pooling and Recurrent Neural Networks", 2016 The IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2874-2883, 2015.  
<https://doi.org/10.48550/arXiv.1512.04143>
- [8] B. Kong, X. Wang, Z. Li, Q. Song, S. Zhang, "Cancer Metastasis Detection via Spatially Structured Deep Network", in Lecture Notes in Computer Science, Springer International Publishing, pp. 236-248, 2017.  
[https://doi.org/10.1007/978-3-319-59050-9\\_19](https://doi.org/10.1007/978-3-319-59050-9_19)
- [9] T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, "Focal Loss for Dense Object Detection", 2018 Proceedings of the IEEE International Conference on Computer Vision, pp. 2980-2988, 2017.  
<https://doi.org/10.48550/arXiv.1708.02002>
- [10] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, "Feature Pyramid Networks for Object Detection", 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, pp. 2117-2125, 2016.  
<https://doi.org/10.48550/arXiv.1612.03144>
- [11] K. He, X. Zhang, S. Ren, J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 37, No. 9, pp. 1904-1916, 2015.  
<http://dx.doi.org/10.1109/TPAMI.2015.2389824>
- [12] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770-778, 2015.  
<https://doi.org/10.48550/arXiv.1512.03385>
- [13] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, "Squeeze-and-Excitation Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42, No. 8, pp. 2011-2023, 2020.  
<https://doi.org/10.1109/TPAMI.2019.2913372>
- [14] S. Woo, J. Park, J. Y. Lee, I. S. Kweon, "CBAM: Convolutional Block Attention Module", European Conference on Computer Vision, pp. 3-19, 2018.  
[https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
- [15] fizyr, keras-retinanet, "Keras-Retinanet", 2019 Model Retrieved from Github, From URL;  
<https://github.com/fizyr/keras-retinanet/>
- [16] Kaggle, "Ultrasound Nerve Segmentation Dataset", 2016 Data Retrieved From URL;  
<https://www.kaggle.com/c/ultrasound-nerve-segmentation/>

# 객체 검출 알고리즘을 활용한 딥러닝 기반 상완 신경총 초음파 영상의 분할에 관한 연구

조국현<sup>1</sup>, 류현승<sup>1</sup>, 이명진<sup>1</sup>, 박수형<sup>2,\*</sup>

<sup>1</sup>전남대학교 컴퓨터정보통신공학과

<sup>2</sup>전남대학교 전자컴퓨터공학부

## 요 약

초음파 유도 국소마취는 통증 관리와 회복 시간을 개선하여 말초신경 차단에 널리 사용되는 기법이다. 하지만 능숙한 임상 의들에게도 초음파 영상에서 나타나는 speckle 및 Doppler와 같은 영상에 내재되어 있는 artifacts로 인하여 상완 신경총(BP; Brachial Plexus)의 정확한 검출 및 식별이 여전히 난제로 남아있다. 이 문제를 해결하기 위해, 우리는 다중 스케일의 접근법을 기반으로 하는 BP의 객체 검출과 그 결과로부터 U-Net 기반의 의미론적 영상 분할을 수행하는 small target 기반의 BP segmentation 알고리즘을 제안한다. 이를 위해 현재 BP 검출 및 식별은 다음과 같이 진행되었다: 1) 다중 스케일 기반의 RetinaNet 모델을 활용하여 BP 신경 영역을 대략적으로 특정하는 단계와 2) 객체 검출로부터 제한된 영상의 범위를 입력으로 U-Net을 활용함으로써 BP 신경의 영역을 검출하는 단계. 실험 결과는 제안된 모델이 분할 전용 모델 등의 경쟁 방법에 비해 BP 신경 영역을 대략적으로 특정하여 식별 범위를 제한함으로써 BP 신경 범위 분할의 정확도를 높이고 고품질 BP 분할을 생성할 수 있음을 보여준다.

중심단어: 상완신경총, 초음파영상, 객체탐지, 이미지분할

## 연구자 정보 이력

	성명	소속	직위
(제1저자)	조국현	전남대학교 컴퓨터정보통신공학과	학사과정
(공동저자)	류현승	전남대학교 컴퓨터정보통신공학과	학사과정
	이명진	전남대학교 컴퓨터정보통신공학과	학사과정
(교신저자)	박수형	전남대학교 전자컴퓨터공학부	조교수