

U-Net 기반의 좌심방 영역 분할 기술

이진혁¹, 이도희², 왕우진³, 박승⁴, *신용구²

¹ 한남대학교 컴퓨터공학과, ² 한남대학교 AI 융합학과, ³ 한남대학교 호텔항공경영학과, ⁴ 메디브

jinlee382@gmail.com, dlehgm154@gmail.com, wwj604@naver.com, spark.chnuh@gmail.com, *ygshin@hnu.kr

Left Atrial Segmentation Technique based on U-Net

JinHyeok Lee, Do-Hee Lee, WooJin Wang, Seung Park, and Yong-Goo Shin

^{1,2,3}Hannam University and ⁴Mediv, Corp.

요 약

의사들은 심방세동을 예측하기 위해 환자의 컴퓨터 단층 촬영(Computer Tomography, CT)영상에서 좌심방 영역을 관찰한다. 본 논문에서는 의사의 진료를 돕기 위해 CT 영상에서 좌심방 영역을 검출하는 기술을 개발하는 것을 목표로 한다. Biomedical Image Segmentation 분야에서 우수한 성능을 보이는 U-net 모델을 이용하여 좌심방 영역을 분할하는 연구를 진행하였다. 본 연구에서는 적은 수의 CT 영상에서 과적합 (overfitting) 문제를 해결하기 위해 기존 U-Net 모델 대비 1/4 채널 수를 갖는 모델을 사용하였다. 실험을 위하여 충북대병원에서 제공한 심방세동 환자 34 명, 정상인 81 명의 CT 영상을 활용하여 6751 개(훈련 데이터: 5568 개, 평가 데이터: 1183 개)를 이용하였다. 실험 결과 제안하는 기법은 IoU 와 Pixel Accuracy 가 각각 92%, 99% 라는 높은 성능으로 좌심방 영역을 검출한다.

1. 서 론

심방세동은 치료를 요하는 가장 흔한 부정맥 중의 하나로, 심방이 매우 빠르게 뛰면서 심실의 전기전도체계에 이상이 생기는 것이 특징이다[1].

Convolution network 의 일반적인 사용은 분류 작업에 있으며, 수천 개의 훈련 이미지는 일반적으로 생의학 분야에서 처리하기 힘들다. 생의학 이미지는 원하는 출력에 localization 이 포함되어야 하고, 극 소수의 이미지로부터 효과적으로 학습할 수 있어야 한다. Segmentation 에서의 대표적인 모델은 FCN, DeepLab v1/v2, ReSeg 등이 있는데, 앞에서 언급한 조건들을 충족하는 대표적인 모델이 Ronneberger et al 가 제안한 U-Net 이다[2-4]. Ronneberger et al. 가 제안한 U-Net 의 네트워크 및 훈련 전략은 극 소수의 이미지로부터 효과적으로 학습하기 위해 데이터 증강의 사용에 의존한다[5].

따라서, 본 논문은 U-Net 을 기반으로 불규칙하게 박동하는 부분(좌심방)을 검출하여 심방세동을 예측하기 위한 토대를 만들고자 하였다.

2. 본론

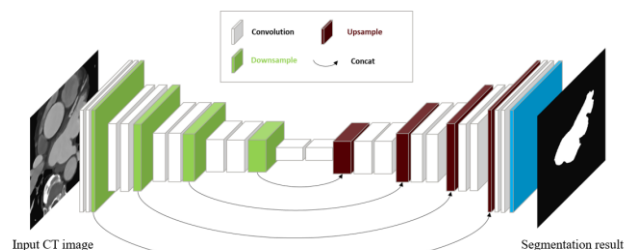
2.1 데이터 셋

충북대병원에서 제공받은 심방세동 환자 34 명, 정상인 81 명의 CT 영상으로부터 데이터 11,004 개를 훈련 데이터 8,896 개, 평가 데이터 2,108 개로 나누어 사용하였다. 원본 CT 영상의 해상도는 1024x1024 이고, 실험 및 평가에서 사용한 CT 영상의 크기는 원본의 절반 크기인 512x512 로 재구성하여 학습시간을 단축하였다.

2.2 네트워크 구조

본 논문에서 사용한 U-Net 은 [그림 1]과 같이 왼쪽의 encoder 부분과 오른쪽의 decoder 부분이 대칭적인 구조를 띄고 있고, 이를 통해 이미지를 학습하고 모든 픽셀의 레이블을 예측한다. U-Net 에서의 encoder 부분은 차원

을 축소하면서 부분적인 feature map 부터 전반적인 feature map 까지 정보를 추출하는 과정으로, Convolution layer(Conv)을 사용하여 이미지를 학습한다. Convolution layer 이후에 배치 정규화(Batch Normalization) 과 활성화 함수(ReLU)를 사용하여 기울기 소실과 폭주를 예방하였다. Encoder 부분은 4 단계의 구조를 가지고, 각 단계마다 Conv 두 번과 Max-pooling 총 세 개의 layer 가 한 묶음으로 설계되어 있다. Decoder 부분은 encoder 과정에서 줄여진 차원을 Input 이미지의 크기와 동일한 크기의 이미지로 만드는 과정으로, up-sampling 을 사용하여 차원을 확대한다. Decoder 부분 또한 encoder 부분과 동일한 구조를 띄고 있다. 다만 이미지를 축소하는 Max-pool layer 대신 up-sample layer 를 사용한다는 점이 상이하다. skip-connection 을 통해 encoder 과정에서 추출되는 feature map 들을 decoder 과정에 결합함으로써 공간정보의 손실을 최소화할 수 있다[7].

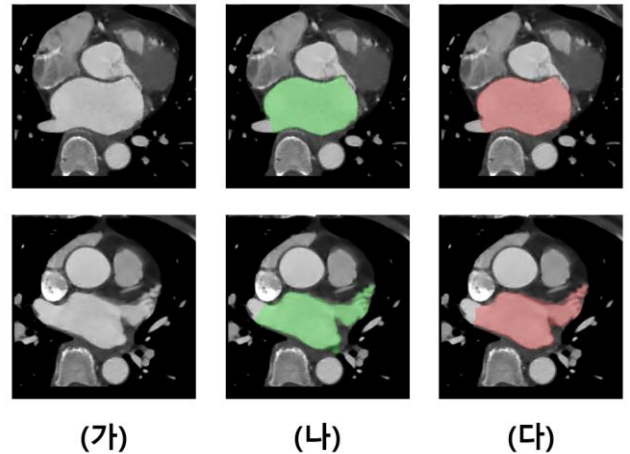


[그림 1] U-Net 의 모델 구조

Ronneberger et al. 가 제안한 U-Net 의 네트워크는 [표 1] 왼쪽의 U-Net 은 많은 채널의 수를 사용하고 있다. 본 논문에서는 [표 1] 오른쪽의 U-Net-Light 와 같이 기존 채널의 수를 1/4 로 축소하여 메모리의 사용량을 줄였다.

U-Net		U-Net-Light	
Encoder	Decoder	Encoder	Decoder
Conv3-64	Upsampling	Conv3-16	Upsampling
Conv3-64	Conv3-1024	Conv3-16	Conv3-128
Maxpool	Conv3-512	Maxpool	Conv3-128
Conv3-128	Upsampling	Conv3-32	Upsampling
Conv3-128	Conv3-512	Conv3-32	Conv3-128
Maxpool	Conv3-256	Maxpool	Conv3-64
Conv3-256	Upsampling	Conv3-64	Upsampling
Conv3-256	Conv3-256	Conv3-64	Conv3-64
Maxpool	Conv3-128	Maxpool	Conv3-32
Conv3-512	Upsampling	Conv3-128	Upsampling
Conv3-512	Conv3-128	Conv3-128	Conv3-32
Maxpool	Conv3-64	Maxpool	Conv3-16
Conv3-1024	Conv3 - 1	Conv3-128	Conv3 - 1
Conv3-1024		Conv3-128	

[표 1] 채널 수 비교



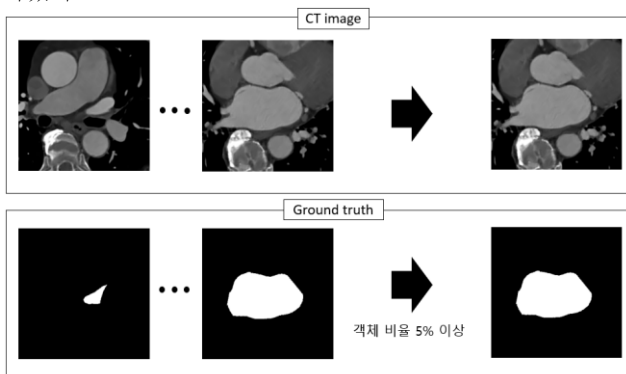
[그림 3] (가) 입력 (나) GT (다) 예측

2.3 실험 방법

모델의 구조 설계, 실험 및 평가를 위해 TensorFlow 와 Keras 를 사용하였고, 이미지를 처리하는 과정에서 Opencv 를 사용하였다.

CT 영상 11,004 개의 데이터에서 훈련 데이터와 평가 데이터를 8 대 2 비율로 적절히 나누었고, 훈련 데이터와 평가 데이터에서 [그림 2]와 같이 Ground truth 에서의 객체 비율이 5% 이상인 이미지들을 추출하여 최종적으로 훈련 데이터 5,568 개 평가 데이터 1,183 개를 학습과 평가에 사용하였다.

훈련 배치의 크기는 8, 평가 배치의 크기는 100 으로 설정하였고 예측 확률 값을 시그모이드 함수로 다시 조정하여 출력 값이 0.5 이상일 경우만 참으로 판단하게 하였다.



[그림 2] 객체 비율 5% 이상인 데이터 추출

2.4 실험 결과

학습한 모델로 평가 데이터를 예측하였을 때 출력된 값을 기반으로 Ground truth와의 비교를 통해 Intersection over Union(IoU)와 Pixel Accuracy를 측정하였고, [표 2]에서 보이는 것과 같이 IoU : 92%, Pixel Accuracy : 99%가 나왔다 [그림 3]에서 보이는 것과 같이 모델이 심방영역을 잘 예측하는 것을 볼 수 있다..

IoU	Pixel Accuracy
92%	99%

[표 2]

3. 결론

본 논문에서는 심방세동 예측에 필요한 좌심방 영역을 높은 정확도로 Segmentation 하는 기법을 제안하였다. 제안하는 U-Net-Light 네트워크를 사용하였을 때, 기존 U-Net 대비 비슷한 성능이 나오고, 네트워크 변수(parameter) 숫자가 적기에 더 빠른 속도를 보였다. 하지만 연구에서 사용한 U-Net 모델 자체의 성능 개선에 대한 부분에서 미흡함을 느꼈다. 본 연구를 통해, 적은 데이터에서도 사용할 수 있는 가벼운 네트워크 구조를 구축하여 CT 영상을 Segmentation 할 수 있다는 성과를 거두었다. Segmentation 결과를 기반으로 의사들의 심방세동의 여부 예측을 도와줄 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

This research was supported by the the National IT Industry Promotion Agency(NIPA) grant funded by the Korea government (Ministry of Science and ICT, MSIT) (NIPA-2022-A1504-22-1021)

참 고 문 헌

- [1] 한아름, and 권수혜, "심방세동 환자의 질병 적응과정: 소리없는 심장의 반란을 다스리며 살아가기." *Korean Journal of Adult Nursing* 32.3 (2020): 326-340.
- [2] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- [3] LC. Chen, et al. "Semantic image segmentation with deep convolutional nets and fully connected crfs." *arXiv preprint arXiv:1412.7062* (2014).
- [4] F. Visin, M. Francesco, et al. "Reseg: A recurrent neural network-based model for semantic segmentation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2016.
- [5] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.
- [6] S. Ioffe, C. Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *International conference on machine learning*. PMLR, 2015.