석사학위논문

2차원 복부 CT 영상에서 딥러닝 기법을 이용한 정확한 근육 분할 기법 연구

2022년

한 성 대 학 교 대 학 원

컴 퓨 터 공 학 과

컴 퓨 터 공 학 전 공

조 다 솜

석 사 학 위 논 문
지도교수 계희원

2차원 복부 CT 영상에서 딥러닝 기법을 이용한 정확한 근육 분할 기법 연구

Accurate Muscle Segmentation Using Deep Learning Method in 2D Abdominal CT Image

2022년 6월 일

한 성 대 학 교 대 학 원

컴 퓨 터 공 학 과

컴 퓨 터 공 학 전 공

조 다 솜

석 사 학 위 논 문
지도교수 계희원

2차원 복부 CT 영상에서 딥러닝 기법을 이용한 정확한 근육 분할 기법 연구

Accurate Muscle Segmentation Using Deep Learning Method in 2D Abdominal CT Image

위 논문을 공학 석사학위 논문으로 제출함

2022년 6월 일

한 성 대 학 교 대 학 원

컴 퓨 터 공 학 과

컴 퓨 터 공 학 전 공

조 다 솜

조다솜의 공학 석사학위 논문을 인준함

2022년 6월 일

심사위원장 이정진 (인)

심 사 위 원 계희원 (인)

심 사 위 원 김진모 (인)

국 문 초 록

2차원 복부 CT 영상에서 딥러닝 기법을 이용한 정확한 근육 분할 기법 연구

한 성 대 학 교 대 학 원
컴 퓨 터 공 학 과
컴 퓨 터 공 학 전 공
조 다 솜

체내 근육의 정량화는 사람의 건강을 평가하기 위한 필수적인 요소이다. 오랜 시간 동안 사람의 건강을 효과적으로 평가하기 위해 다양한 연구가 진행되어 왔고 그중 이미지 양식을 사용한 근육의 질량적 품질적 평가가 발전해 왔다. CT 영상은 근육과 지방 조직 영역에 대한 정확한 시각화가 가능하기 때문에 체내 근육과 지방을 평가하는 데 유용하며 특히 복부 CT 영상을 기반으로 한 근육의 측정은 실제 근육량과 밀접한 연관이 있다. 하지만 단면 영상에서 근육을 분할하는 작업은 전문적인 지식과 복잡한 프로세스를 요구로 하며 시간이 많이 걸리는 번거로운 수작업이고 제한된 인적 자원과 시간으로 인해 대규모 데이터 세트에 적용시키는 데에 어려움을 겪었다. 본 논문에서는 복부 CT 영상에서 딥러닝을 이용한 완전 자동화된 근육 분할 방법에 대해 제안한다. 분할 모델로는 속도가 빠르고 생물의학 이미지 분할에 좋은

성능을 보여주는 U-net을 사용했다. 강체 변환을 이용해 데이터 증강을 진행한 훈련 데이터를 이용해 훈련을 진행했으며 훈련 결과에 이미지 후처리를 통해 근육 영역 마스크에서 근육 외에 해당하는 영역을 제거했다. 학습된 모델은 임상적으로 사용될 수 있는 정확도를 넘는 98%의 정확도를 보였으며 학습된 모델을 이용해 한 장의 복부 CT 영상에서 근육을 분할하는 데에 약 1초가 소요되었다.

【주요어】 딥러닝, 영상 처리, 분할, 근감소증

목 차

표 목 차

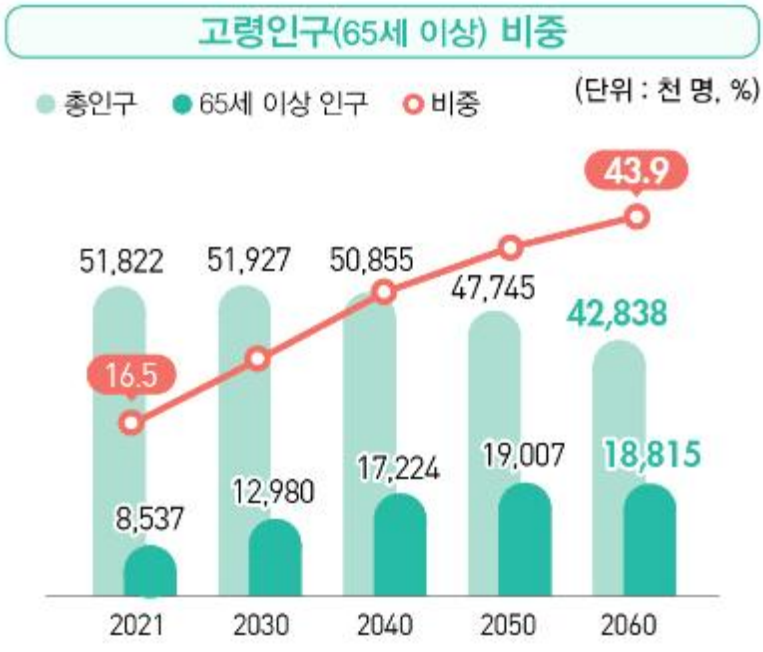
그 림 목 차

수 식 목 차

제 1 장 서 론

제 1 절 연구 배경

근감소증(sarcopenia)이란 노화와 함께 발생하는 근육량과 힘의 상실을 정의하는 데 사용되는 용어이다(Morley, 2001). 최근 한국의 노년층이 예상보다 빠르게 증가하고 있어 근감소증에 대한 관심도 급증하고 있다(홍상모, 2012). 그에 따른 연구로 근감소증이 암 환자의 수술 후 예후에 미치는 영향, 근감소증이 암 환자의 생존율에 미치는 영향 등이 있다(Blauwhoff-Buskernolen, 2016)(Reisinger, 2015)(Fukuda, 2016)(Bokshan, 2011). [그림 1-1]은 한국의 65세 이상 고령인구 비중을 예측한 그래프이다.



[그림 1-1] 65세 이상 고령인구 비중(통계청, 2021)

오랜 시간 동안 사람의 건강을 효과적으로 평가하기 위한 다양한 연구가

진행되어왔고 그중 이미지 양식을 사용하여 근육의 질량과 품질에 대해 평가하는 분야가 발전해왔다. CT(computed tomography) 영상은 X-선을 투과시켜 그 흡수 차이를 컴퓨터로 재구성하여 인체의 단면 영상 또는 3차원 입체 영상을 얻는 영상진단 기법(Buzug, 2011)이다. 체내 근육과 지방의 정량화는 건강을 평가하기 위해서 필수적인 요소이다. CT는 근육과 지방 조직 영역에 대한 정확한 시각화가 가능하기 때문에 체내 근육과 지방을 평가하는 데 유용한 도구이다. 특히 복부 CT 영상을 기반으로 한 근육과 지방의 측정은 실제 근육량과 밀접한 연관(Hong, 2014)이 있다.

단면 영상에서 근육을 분할하기 위해서는 전문적인 지식과 복잡한 수작업 프로세스를 요구로 하며 시간이 많이 걸리는 번거로운 작업(Jones, 2015)(Shen, 2004)(Zhao, 2019)이었다. 제한된 인적 자원과 시간으로 인해 대규모 데이터 세트에 이러한 분할 방법을 적용하는 데에는 어려움이 있었다.

본 논문에서는 복부 CT 영상에서 딥러닝을 이용한 완전 자동화된 근육 분할 방법에 대해 제안한다.

제 2 절 연구 내용

본 논문에서는 여러 의미론적 분할(semantic segmentation) 모델 중 U-net(Ronneberger, 2015)을 이용해 복부 CT 영상에서 근육 분할을 수행한다. U-net은 생물의학 이미지에서 뛰어난 성능을 모이는 분할 모델로 빠른 훈련 시간과 높은 예측 정확도, 적은 이미지 양으로도 정확한 분할이 가능하도록 고안된 모델이다.

훈련 데이터에는 강체 변환(rigid transformation)을 이용한 데이터 증강(data augmentation)을 수행했다. 의료 영상 데이터는 시간, 비용, 사생활 등의 문제로 다량의 데이터를 수집하기 어려운 경우가 많은데 이러한 문제점을 데이터 증강을 통해서 극복할 수 있다. 또한 데이터 증강은 훈련 모델이 과적합(overfitting)되는 것을 방지할 수 있다.

근육 영역 마스크에 대해 훈련된 모델은 후처리를 통해 근육 외에 해당하는 영역을 제거하는 과정을 거친다.

제 2 절 논문 구성

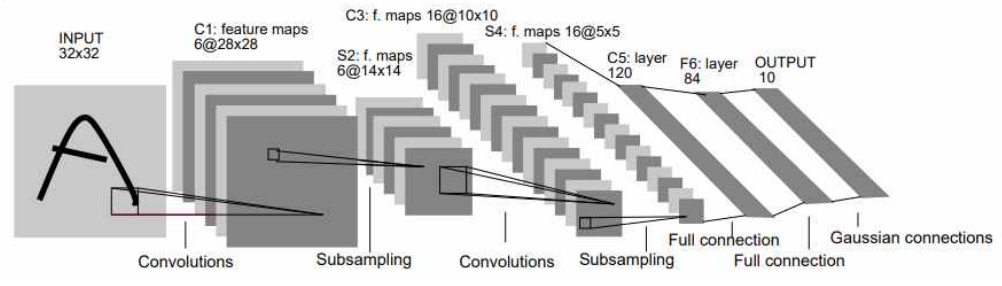
본 논문은 2장에서 관련 연구인 여러 분할 네트워크에 대해 알아본다. 3장에서는 분할 네트워크 U-net에 관한 설명과 본 논문에서 적용된 이미지 영상 처리에 대해 설명한다. 4장에서는 본 논문에서 제안된 분할 네트워크에 대한 실험 결과와 성능 평가에 대해 설명하고, 5장에서는 결론과 향후 연구에 대한 제시로 논문을 마무리한다.

제 2 장 관련 연구

제 1 절 CNN

컴퓨터 과학 분야에서 딥러닝 알고리즘의 발전은 많은 성과들을 불러왔다. 특히 CNN(convolutional neural network)은 컴퓨터 비전, 이미지 인식 등의 분야에서 뛰어난 성능을 보였다. CNN은 DNN(deep neural network) 중에 하나로 인간의 시신경을 모방한 인공 신경망 구조이다.

LeNet-5(LeCun, 1998)는 수표에 쓰인 손글씨를 인식하는 모델로 현대 CNN의 초석이 되는 모델이다. CNN에서는 주로 컨볼루션 레이어(convolution layer)와 풀링 레이어(pooling layer)가 사용된다. 컨볼루션 레이어에서는 특징 맵(feature map)을 수신한 후 컨볼루션 레이어의 수용 영역에 대한 픽셀을 계산해 특징 값을 추출한다. 풀링 레이어에서는 특징 맵의 크기를 줄이고 특정된 값을 강조한다. [그림 2-1]은 LeNet-5의 네트워크 구조를 나타낸 그림이다.



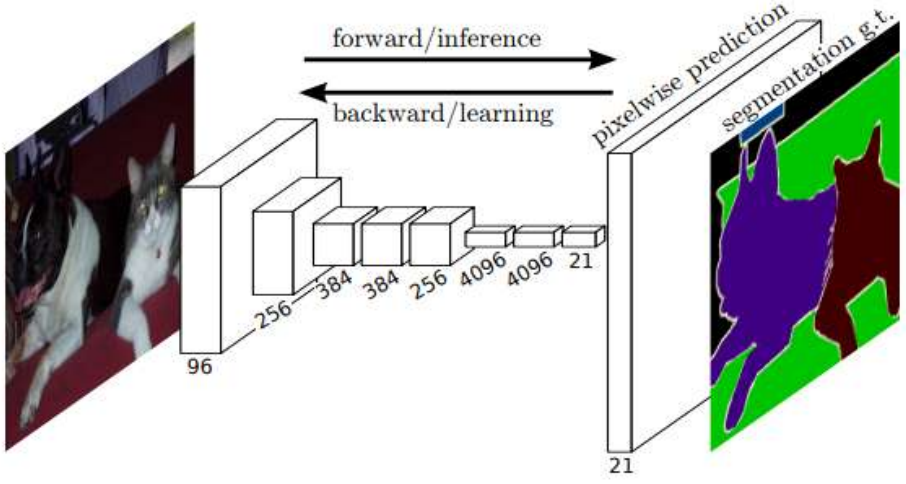
[그림 2-1] LeNet-5 네트워크 구조(LeCun, 1998)

제 2 절 분할 네트워크

이미지 분할(image segmentation)이란 이미지 전체의 클래스(class)를 예측하는 것에서 더 나아가 이미지를 이루는 모든 픽셀들의 클래스를 예측하는 것이다. 이미지 분할은 의료 영상, 자율주행 자동차, 위성 영상 등의 다양한 분야에서 활용

되고 있다. 이미지 분할을 위해서는 영상의 픽셀 단위의 마스크를 출력하도록 신경망을 훈련시켜야 한다. 이미지 분할 모델 종류 중 대표적인 것들을 설명해 보고자 한다.

1) FCN



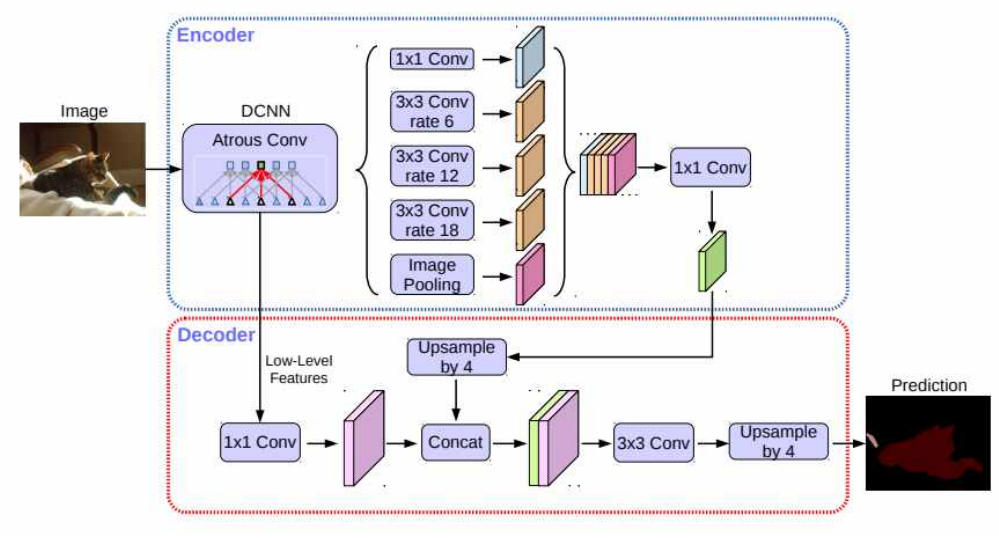
[그림 2-2] FCN 네트워크 구조(Long, 2015)

FCN(fully convolutional networks)(Long, 2015)은 이미지 분류에서 성능을 검증받은 기존의 네트워크를 이용해 의미론적 분할에 맞게 파인 튜닝(fine-tuning)한 모델이다. 기존 분류 모델에서는 클래스 분류를 위해 네트워크 마지막에 완전 연결 계층(fully connected layer)을 삽입했다. 이러한 방식은 고정된 크기의 입력(input)만을 받아들일 수 있고, 완전 연결 계층을 거치고 나면 위치 정보가 사라지기 때문에 이미지 분할에 있어서 적합하지 않다. 따라서 FCN에서는 완전 연결 계층을 컨볼루션 레이어로 대체하여 입력 크기에 제약을 받지 않고, 위치 정보가 손실되지 않게 변경했다.

컨볼루션 레이어를 거치고 나서 얻게 되는 특징 맵은 대략적인 정보만 가지고 있기 때문에 특징 맵의 크기를 원래 이미지의 크기로 다시 복원해 줄 필요가 있다. FCN에서는 컨볼루션 연산을 반대로 수행하는 디컨볼루션(deconvolution)을 수행해 자연스러운 업 샘플링(up-sampling) 효과를 얻어낸다. 하지만 확대하기 전

특징 맵이 매우 작기 때문에 확대했을 시 상당히 많은 정보를 상실했을 가능성이 높아 더 정확하고 상세한 분할 정보를 위해 스킵 커넥션 아키텍처(skip connection architecture)를 도입했다. [그림 2-2]는 FCN 네트워크 구조를 나타낸 그림이다.

2) DeepLab v3+



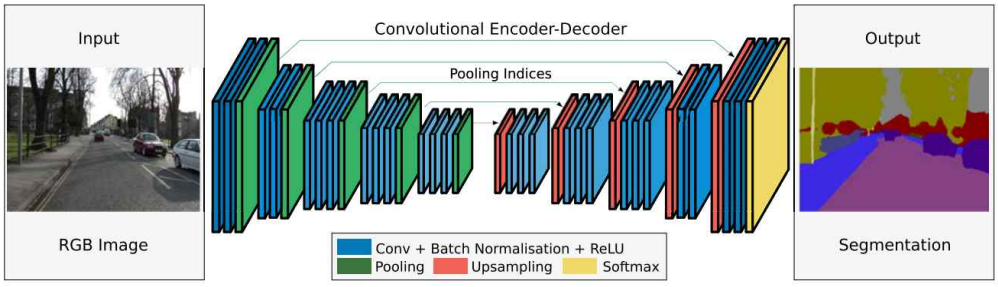
[그림 2-3] DeepLab v3+ 네트워크 구조(Chen, 2018)

DeepLab(Chen, 2014)은 아트러스 컨볼루션(Atrous convolution)을 처음으로 적용한 의미론적 분할 모델이다. 아트러스 컨볼루션은 이미지 분류 모델 또는 객체 탐지 모델에서 사용되는 신경망들을 의미론적 분할에 사용할 경우, 컨볼루션과 풀링(pooling)을 거치면서 특징 맵이 추상화되는 문제점을 해결하고자 도입된 방법이다. 아트러스 컨볼루션을 이용하면 기존 컨볼루션과 동일한 양의 파라미터와 계산량을 유지할 수 있다. 또한 큰 수용 영역(receptive field)을 통해 정보의 손실을 최소화할 수 있어 해상도가 좋은 결과를 얻어낼 수 있다.

DeepLab은 여러 번의 버전 업을 진행했고 가장 최신 버전의 DeepLab은 DeepLab v3+이다. DeepLab v3+(Chen, 2018)는 아트러스 컨볼루션에 깊이 단위 분리 컨볼루션(Depthwise separable convolution)을 결합한 아트러스 분리 컨

볼루션(Atrous separable convolution)을 제안한다. 단위 분리 컨볼루션이란 컨볼루션 필터가 동시에 공간 특징(spatial feature)과 채널 특징(channel feature)을 처리하던 것에서 분리하여 각각 처리하는 방법이다. 분리를 했을 뿐 두 가지 모두 처리했기 때문에 기존 컨볼루션이 수행하던 역할을 대체할 수 있고 파라미터 수와 연산량을 상당히 줄였기 때문에 이러한 구조를 채택했다. [그림 2-3]은 DeepLab v3+의 네트워크 구조를 나타낸 그림이다.

3) SegNet



[그림 2-4] SegNet 네트워크 구조(Badrinarayanan, 2017)

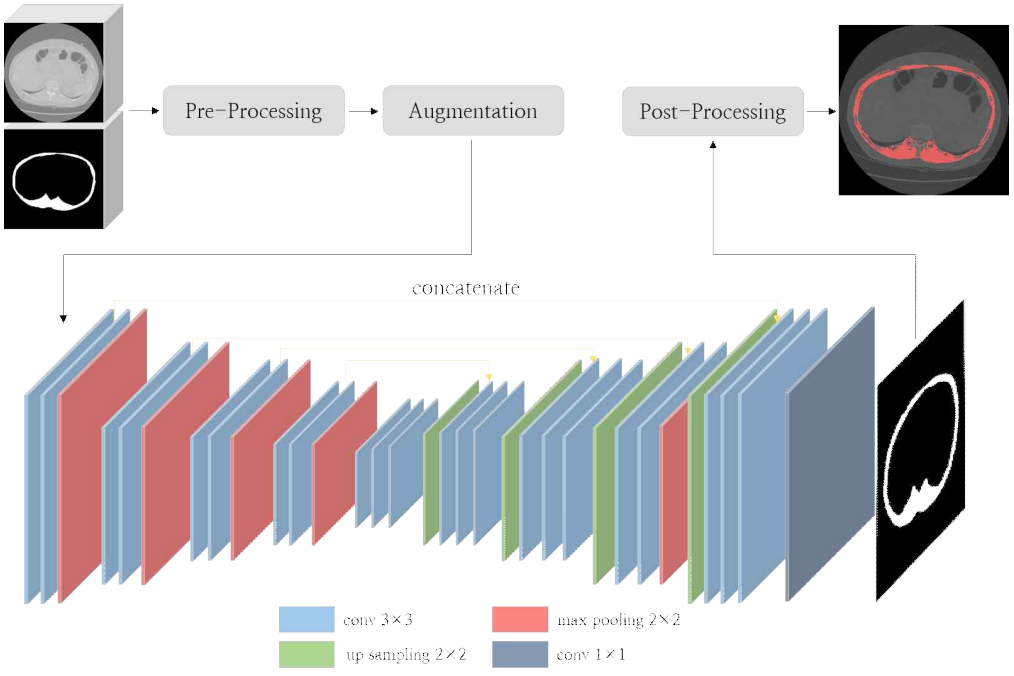
기존의 의미론적 분할 모델들은 맥스 풀링(max pooling), 서브 샘플링(subsampling) 등의 연산을 수행해 특징 맵을 만들었는데, 이러한 방식을 사용하면 해상도가 매우 떨어져 픽셀 단위로 정교하게 분할할 수 없게 된다. 또한 계산량이 많거나 파라미터 수가 많다면 정확도가 높을지라도 빠른 분할을 할 수 없게 된다. SegNet은 이러한 문제들을 해결하고자 했으며 도로 및 실내 장면에 대해 메모리와 계산 시간 측면에서 효율적으로 이해하기 위한 아키텍처 설계가 목적이다.

SegNet(Badrinarayanan, 2017)은 의미론적 픽셀 단위 분할(semantic pixel-wise segmentation)을 위한 FCNN(fully convolutional neural network) 아키텍처이다. 인코더(encoder)에는 VGG16(Simonyan, 2014) 네트워크 구조에서 완전 연결 계층를 제외한 13개의 레이어를 사용했다. 완전 연결 계층은 상당히 많은 계산량을 요구로 하고, 이미지 분류(image classification) 작업이 필요한 것

이 아니라면 반드시 필요하지 않기 때문에 이 부분을 제거함으로써 연산량을 줄일 수 있다. 디코더(decoder)에서는 인코더에서 저장한 맥스 풀링 인덱스들을 받아 이후 맥스 풀링을 진행한다. 이러한 방식을 이용하면 정확도가 아주 조금 감소하더라도 좋은 수치의 정확도를 유지하면서 메모리를 크게 아낄 수 있다는 장점이 있다. [그림 2-4]는 SegNet 네트워크 구조를 나타낸 그림이다.

제 3 장 본 론

제 1 절 네트워크 아키텍처

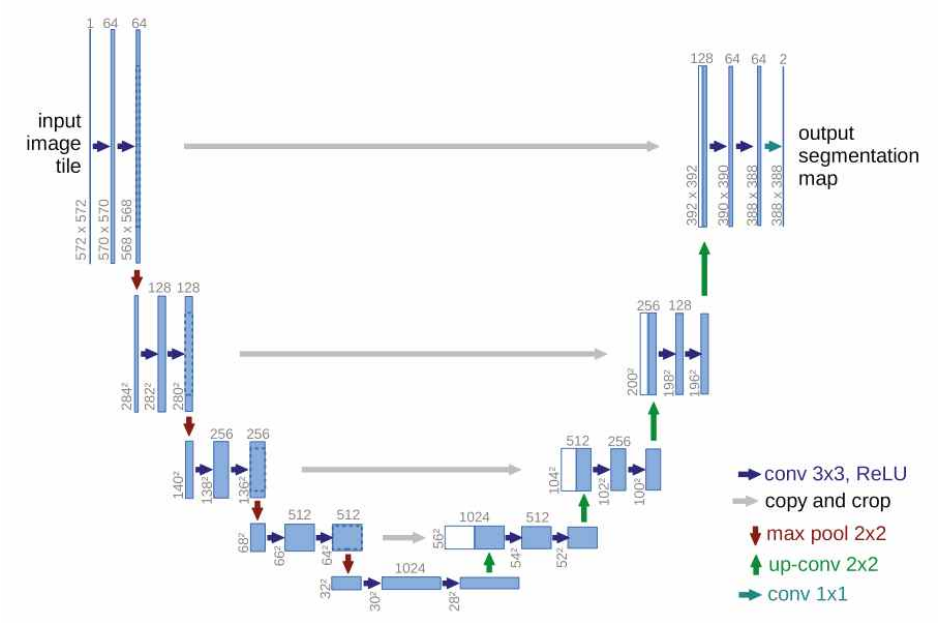


[그림 3-1] 제안된 분할 모델 구조

다양한 이미지 분할 모델 중에 본 논문에서는 U-net 분할 모델을 사용했다. [그림 3-1]은 U-net 네트워크를 이용한 본 논문에서 제안하는 분할 모델의 전반적인 구조를 나타내는 그림이다.

U-net은 생물의학 이미지 분할을 위해 고안된 FCN 기반 종단 간(end-to-end) 방식 모델이다. CNN으로 구성되어 있어 입력 영상 크기에 제약을 받지 않는다. 종단 간 방식 학습은 입력에서 출력까지 파이프라인 네트워크 없이 신경망으로 한 번에 처리할 수 있어 직접 파이프라인을 설계할 필요가 없다는 장점이 있다. 또한 이미 검증이 끝난 부분을 다시 검증하는 슬라이딩 윈도(sliding window) 방식 대신 검증이 끝난 부분은 건너뛰고 다음 패

치(patch) 구역부터 검증을 진행하는 방식을 도입해 속도가 빠르다. [그림 3-2]는 U-net 네트워크 구조를 나타낸 그림이다.



[그림 3-2] U-net 네트워크 구조(Ronneberger, 2015)

1) 인코더

수축 경로(contracting path)라고도 부르는 인코더에서는 입력 이미지의 컨텍스트(context) 정보를 얻는다. 각각의 단계마다 두 개의 3×3 컨볼루션을 통과한다. 이때 컨볼루션은 패딩(padding)이 적용되지 않아 특징 맵이 단계를 거침에 따라 조금씩 줄어들며 활성화 함수는 ReLU(Agarap, 2018)를 사용한다. 각각의 다운 샘플링(down sampling)에는 스트라이드(stride) 크기가 2인 2×2 맥스 풀링 연산을 수행한다. 이 단계에서 특징 맵의 크기는 절반으로 줄어들고 특징 채널(feature channel) 수는 두 배로 늘어난다.

2) 디코더

확장 경로(expansive path)라고도 부르는 디코더에서는 정확한 위치 식별(localization) 정보를 얻는다. 디코더는 인코더와 대칭 구조를 이루고 있고 각각의 단계마다 업 샘플링(up sampling)을 진행한다. 단계마다 2×2 업 컨볼루션(up-convolution)을 수행하는 데 이때 특징 채널의 수가 절반으로 줄어들고 특징 맵의 크기가 두 배로 늘어난다. 또한 단계마다 인코더 단계에서 잘렸던 특징 맵과 업 컨볼루션된 특징 맵을 연결해 ReLU가 포함된 3×3 컨볼루션 연산을 수행한다.

3) 스킵 커넥션 아키텍처

수축 경로를 통해 얻어진 특징 맵은 의미론적 분할이 픽셀 단위로 예측한다는 점으로 미루어 봤을 때 해상도가 현저하게 낮다. 이러한 상태의 특징 맵을 거친 특징 맵(coarse feature map)이라고 부른다. 의미론적 분할을 하기 위해서는 거친 특징 맵을 원본 이미지 크기에 가깝게 키워야 하는데 이 과정에서 발생할 수 있는 정확도 손실을 인코더에서 대응되는 단계의 특징 맵을 연결함으로써 정교한 예측이 가능해진다.

제 2 절 이미지 처리

1) 데이터 증강

CNN 네트워크의 깊이가 깊어짐에 따라 과적합을 피하기 위해서는 빅데이터에 의존하게 된다(Shorten, 2019). 과적합이란 과하게 학습되어 훈련 데이터 세트 안에서는 일정 수준 이상의 예측 정확도를 보이지만 실제 데이터에 대한 오차가 증가하는 현상을 말한다. 하지만 의료 영상 데이터와 같이 시간, 비용 등의 문제로 다량의 데이터를 수집하기 어려운 경우가 있다. 이러한 경우 데이터 증강 통해서 극복할 수 있다. 데이터 증강은 데이터의 양을 늘리

기 위해 원본 데이터에 다양한 변환을 적용시켜 데이터의 개수를 늘리는 방법이다. 변환에도 여러 종류가 있는데 본 논문에서는 강체 변환과 크기 변환, 대칭 변환을 적용시켰다. 유클리드 변환(Euclidean transformation)이라고도 불리는 강체 변환은 형태와 크기는 유지한 상태에서 위치와 회전을 조정하는 변환 방법이다. [수식 3-1], [수식 3-2], [수식 3-3], [수식 3-4]는 각각 크기, 회전, 대칭, 평행이동에 관한 행렬식이다.

$$\begin{bmatrix} x_{scale} \\ y_{scale} \end{bmatrix} = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

[수식 3-1] 크기 변환 수식

$$\begin{bmatrix} x_{rotation} \\ y_{rotation} \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

[수식 3-2] 회전 변환 수식

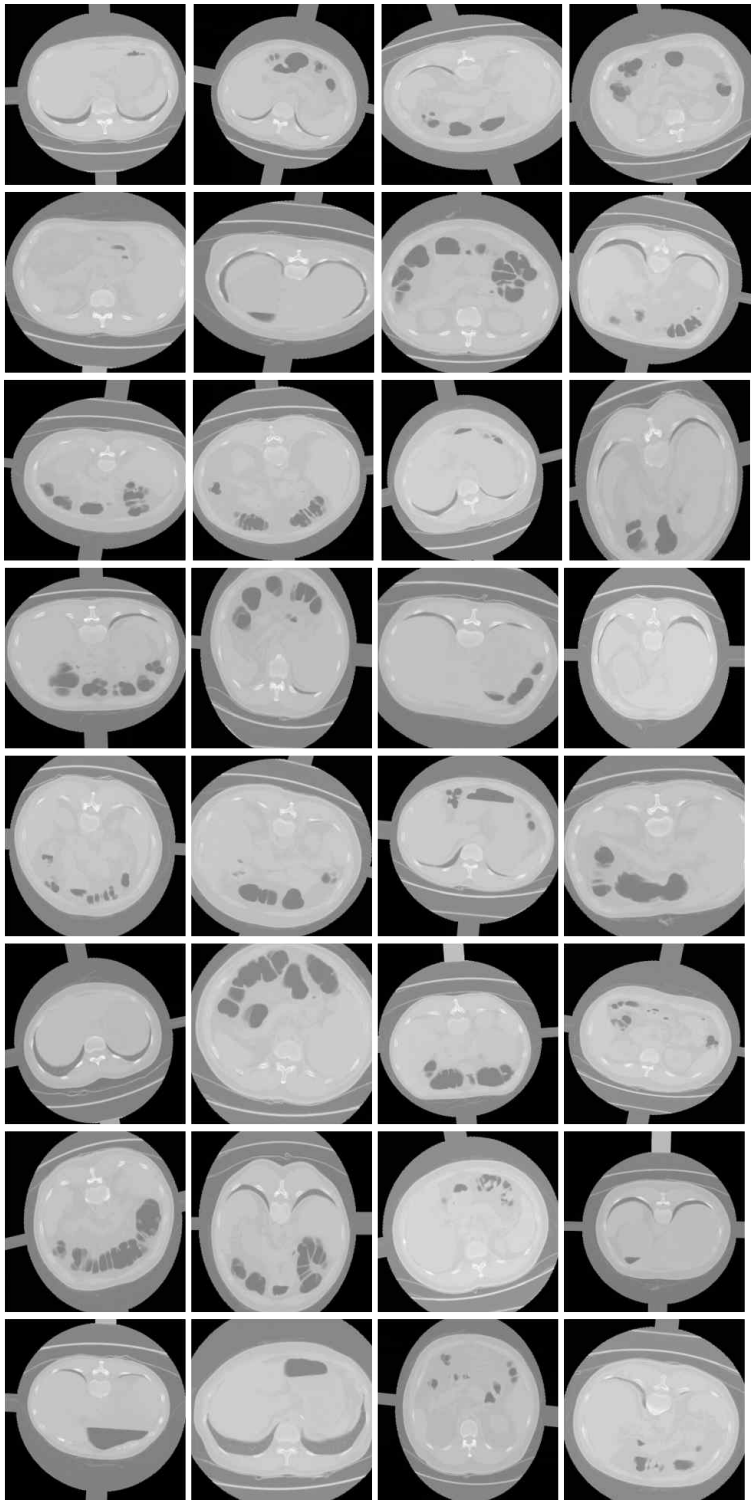
$$\begin{bmatrix} x_{reflect} \\ y_{reflect} \end{bmatrix} = \begin{bmatrix} r_x & 0 \\ 0 & r_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, r_x = \begin{cases} 1 \\ -1 \end{cases}, r_y = \begin{cases} 1 \\ -1 \end{cases}$$

[수식 3-3] 대칭 변환 수식

$$\begin{bmatrix} x_{translation} \\ y_{translation} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

[수식 3-4] 평행이동 변환 수식

훈련 데이터 세트를 증강한 결과의 일부는 [그림 3-3]과 같다.



[그림 3-3] 강체 변환을 이용한 데이터 증강

2) 이미지 전처리

HU(hounsfield unit)란 CT 영상 촬영할 때 기준이 되는 단위이다. 정확한 CT 영상 처리를 위해 다이콤(dicom) 파일로부터 추출한 픽셀 값들을 HU로 정규화하는 과정을 거쳐야 한다. 이때 다이콤 메타 태그(meta tag)에 있는 리스케일 슬로프(rescale slope)와 리스케일 인터셉트(rescale intercept) 정보를 이용해 픽셀 값을 정규화한다. 이러한 과정을 거쳐 훈련에 들어가기 전 영상을 HU로 변환했다. [수식 3-5]는 픽셀 값을 HU로 변환하는 식이다.

$$hu = (slope) \times (pixel) + (intercept)$$

[수식 3-5] CT 데이터를 HU로 변환

3) 이미지 후처리

근육 영역 마스크에는 근육 외에도 뼈나 지방 같은 다른 조직이 포함되어 있을 가능성이 있다. 따라서 이미지 후처리를 통해 근육을 제외한 나머지를 근육 영역 마스크에서 제거해 주는 과정을 수행한다.

CT 영상을 HU로 변경하면 HU의 범위에 따라 영상 내의 근육을 알아낼 수 있다. CT 이미지의 그레이 스케일을 구성하는데 범위는 −1024HU ~ 3071HU이다. 기준값 0HU인 −1024HU 부근은 공기, 3071HU 부근은 치아나 뼈 같은 밀도 높은 조직을 나타낸다. 이 중 근육은 100HU 부근이며 본 논문에서는 −29HU ~ 150HU를 근육의 HU 범위로 잡았다. [그림 3-4]에서 (a), (c)는 근육 영역 마스크이며 (b), (d)는 후처리를 진행한 마스크이다.

[그림 3-4] 예측 마스크와 후처리를 거친 마스크

제 4 장 실험 및 결과

제 1 절 실험 환경

본 논문에서 제안된 방법에 대한 실험 환경은 다음과 같다.

[표 4-1] 실험 환경

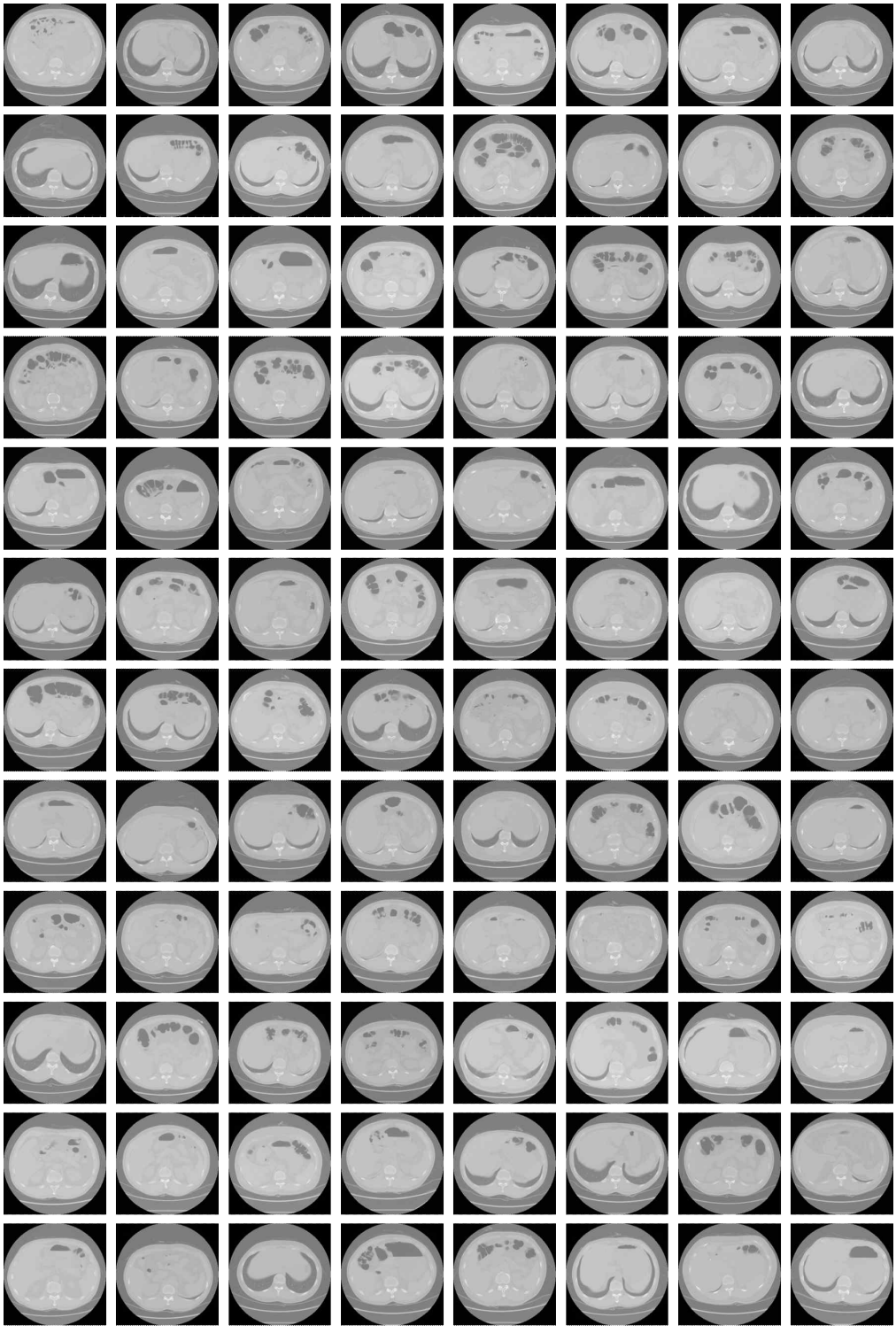
구분	개발환경
CPU	AMD Ryzen 7 5800H Processor 3.0GHz
GPU	NVIDIA GeForce RTX 3060 6GB
RAM	16GB
OS	Windows 10 Home
Language	3.8.11
Deep Learning Tool	Pytorch 1.9.1 / Tensorflow 2.6.0
Library	CUDA 11.1 / CuDNN 8.1.1 / OpenCV 4.0.1

제 2 절 실험 데이터

데이터 세트는 GE 사의 BrightSpeed, Discovery CT750 HD, HiSpeed, LightSpeed, Optimal CT660, Hitachi 사의 Presto, Philips 사의 iCT 256, Siemens 사의 Aquilion, Definition, Emotion, Sension, Somatom 스캐너를 사용해 얻은 CT 이미지로 구성되었다. CT 영상의 크기는 512×512이다. 그라운드 트루쓰(ground truth)는 8년 경력의 의료 영상 분석가가 수동으로 식별하고 생성했고, 임상 경력 5년의 방사선 전문의가 검토하고 수정했다.

총 873명의 복부(abdominal) CT 영상과 그라운드 트루쓰 데이터를 사용해서 실험을 진행했다. 약 30%에 해당하는 260건의 데이터를 테스트 데이터로, 약 30%에 해당하는 260건의 데이터를 검증 데이터로 사용했고, 나머지

353건의 데이터를 훈련 데이터로 사용했다. 입력 CT 영상 데이터는 복셀(voxel)로 구성되어 있으며, 정답 데이터의 배경 클래스는 0, 근육 클래스는 1인 이진 클래스로 구성되어 있다. [그림 4-1]은 데이터 세트의 일부를 나타낸 그림이다.



[그림 4-1] 데이터 세트 이미지

제 3 절 손실 함수와 평가 지표

1) 손실 함수

손실 함수(loss function)는 측정한 데이터를 토대로 산출한 모델의 예측값과 실제 값의 차이를 나타내는 지표이며 이를 통해 예측값과 실제 값에 대한 오차를 줄이는 데 유용하게 사용할 수 있다. 본 논문에서는 이진 교차 엔트로피(binary cross entropy)를 손실 함수로 사용했다. 최적화 함수(optimizer function)로는 Adam을 사용했다.

교차 엔트로피(cross entropy)는 신경망을 훈련하는 데 사용되는 일반적인 손실 함수 중 하나로, 머신 러닝(machine learning)의 분류 모델이 얼마나 잘 수행되는지 측정하기 위해 사용되는 지표이다. 그중 이진 교차 엔트로피는 참과 거짓처럼 2개의 클래스로 분류할 수 있는 이진 분류기에서 사용하기 적절한 손실 함수이며 식은 다음과 같다.

$$BinaryCrossEntropy = -(ylog(p) + (1-y)\log(1-p))$$

[수식 4-1] 이진 교차 엔트로피 수식

2) 평가 지표

모델의 성능을 평가하는 것은 머신 러닝 모델을 만들 때 핵심적인 작업 중 하나다. 정확도(accuracy)는 전체 데이터 중 데이터가 예측에 성공할 확률을 의미하며 그 식은 다음과 같다.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

[수식 4-2] 정확도 수식

이때 TP(true positive)는 실제 참인 정답을 참으로 예측한 경우, TN(true

negative)은 실제 거짓인 정답을 거짓이라고 예측한 경우, FP(false positive)는 실제 거짓인 정답을 참이라고 예측한 경우, FN(false negative)은 실제 참인 정담을 거짓이라고 예측한 경우이다. 값이 1에 가까울수록 예측에 성공했다는 것을 의미한다. [표 4-2]는 혼동 행렬(confusion matrix)의 관해 나타낸 것이다.

[표 4-2] 혼동 행렬

실제 값	예측 결과	
	positive	negative
positive	TP	FN
negative	FP	TN

IoU(intersection over union)은 실제 값과 예측값이 얼마나 일치하는지를 의미하며 그 식은 다음과 같다.

$$IoU = \frac{A \cap B}{A \cup B}$$

[수식 4-3] IoU 수식

이때 A는 예측 영역, B는 실제 영역이다. IoU를 측정하기 위해서는 기준이 되는 임곗값(threshold)이 있어야 하는데 본 논문에서는 0.5를 임곗값으로 사용했다. 값이 1에 가까울수록 물체가 잘 검출되고 있다는 것으로 해석된다.

DSC(dice similarity coefficient)는 영상 분할 결과를 비교하여 얼마나 유사한지 평가하는 것으로 식은 다음과 같다.

$$DSC = \frac{2TP}{2TP + FP + FN}$$

[수식 4-4] DSC 수식

값이 1에 가까울수록 실제 값과 예측한 영역이 같다는 것으로 해석할 수 있다. [표 4-3]은 학습 결과에 대한 정확도, IoU, DSC의 표준편차와 평균을

나타낸 것이다.

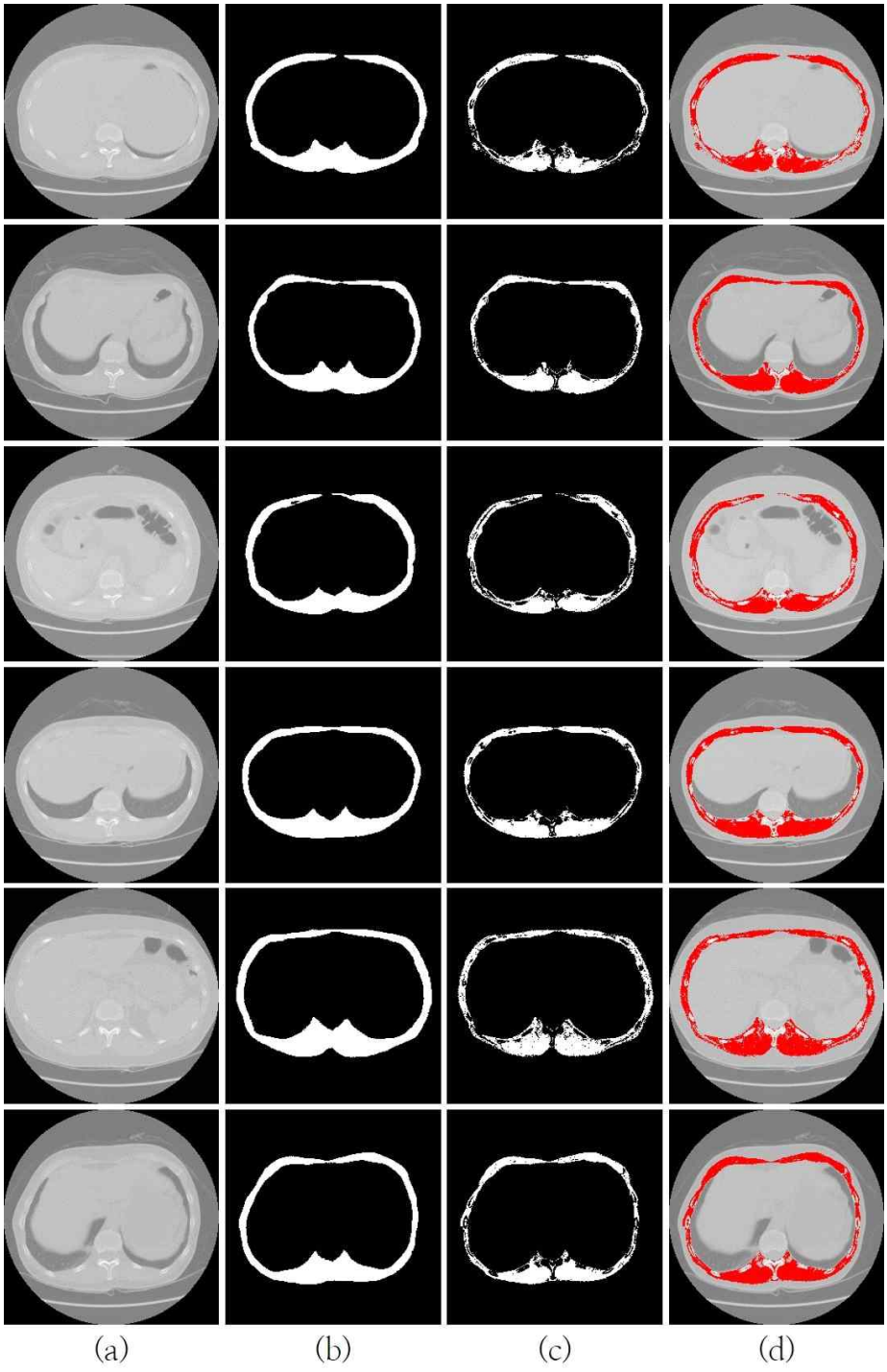
[표 4-3] 평가 지표와 결과

	평균	표준 편차
정확도	0.98	0.03
IoU	0.78	0.15
DSC	0.87	0.14

제안된 방법은 98%의 정확도를 나타내며 이는 임상적으로 자동 분할된 결과를 활용하기 위해서 필요한 정확도인 95%을 넘는 수치이다.

제 4 절 실험 결과

데이터 증강이 적용된 훈련 데이터를 이용해 제안된 모델의 훈련을 진행했다. 실험에 사용된 배치의 크기는 32이고, 총 10 에폭(epoch)를 수행했다. 훈련된 모델을 통해 한 장의 복부 CT 영상에서 근육 분할을 진행하는 데에는 약 1초가 소요되었다. [그림 4-2]에서 (a)는 원본 복부 CT 영상, (b)는 훈련 결과를 바탕으로 생성된 예측 마스크, (c)는 예측 마스크에 이미지 후처리를 진행한 마스크, (d)는 (a)에 (c)를 오버 레이한 이미지이다.



[그림 4-2] 학습된 결과

제 5 장 결론

본 논문은 신체 근육의 정량적 평가를 위한 복부 CT 영상에서 완전 자동화된 근육 분할 기법을 연구하였다. 근육을 정량적으로 평가하기 위한 다양한 방법이 연구되어 왔는데 그중 CT 이미지를 이용한 근육의 평가는 실제 근육량과 밀접한 연관이 있다. 사람이 직접 CT 영상에서 근육을 분할하는 것은 시간이 오래 걸릴 뿐 아니라 전문적인 지식을 요구로 하는 복잡한 작업이기 때문에 대규모 데이터 세트에 적용하는데 어려움이 있었다. 이러한 문제점에 대한 해결책으로 본 논문은 U-net 네트워크를 이용한 완전 자동화된 복부 CT 영상에서 근육 분할 기법을 제안했다.

U-net은 생물의학 이미지 분할을 위해 고안된 분할 네트워크로 빠른 훈련 속도와 높은 정확도가 특징이다. 훈련 데이터에는 강체 변환을 이용한 데이터 증강을 진행했고 CT 영상을 HU로 변환하는 작업을 전처리로 수행했다. 근육 분할은 예측된 근육 영역 마스크에 이미지 후처리를 통해 근육 외의 부분을 제거시켜 진행했다. 하나의 실제 복부 CT 이미지를 분할하는 데에 약 1초가량의 시간이 소요되었으며 98%의 정확도를 보였다. 이는 임상적으로 사용 가능한 95%의 정확도를 넘는 수치이다.

본 논문에서는 CT 영상에서 근육에 관한 분할만을 진행했지만 근육과 더불어 중요한 정보인 내장지방(visceral fat)과 피하지방(subcutaneous fat)을 함께 분할하는 연구를 진행할 수 있을 것으로 생각한다. 또한 네트워크의 인코더 부분에 다른 네트워크를 접목시켜 의미있는 분할 결과를 도출할 수 있을 것으로 생각한다.

참 고 문 헌

1. 국내문헌

홍상모, & 최웅환. (2012). 종설: Sarcopenia 의 최신지견; 근감소증. Korean Journal of Medicine (구 대한내과학회지), 83(4), 444-454.

2. 국외문헌

Morley, J. E., Baumgartner, R. N., Roubenoff, R., Mayer, J., & Nair, K. S. (2001). Sarcopenia. Journal of Laboratory and Clinical Medicine, 137(4), 231-243.

Blauwhoff-Buskermolen, S., Versteeg, K. S., de van der Schueren, M. A., den Braver, N. R., Berkhof, J., Langius, J. A., & Verheul, H. M. (2016). Loss of muscle mass during chemotherapy is predictive for poor survival of patients with metastatic colorectal cancer. Journal of Clinical Oncology, 34(12), 1339-1344.

Reisinger, K. W., van Vugt, J. L., Tegels, J. J., Snijders, C., Hulsewé, K. W., Hoofwijk, A. G., ... & Poeze, M. (2015). Functional compromise reflected by sarcopenia, frailty, and nutritional depletion predicts adverse postoperative outcome after colorectal cancer surgery. Annals of surgery, 261(2), 345-352.

Fukuda, Y., Yamamoto, K., Hirao, M., Nishikawa, K., Nagatsuma, Y., Nakayama, T., ... & Tsujinaka, T. (2016). Sarcopenia is associated with severe postoperative complications in elderly gastric cancer patients undergoing gastrectomy. Gastric cancer, 19(3), 986-993.

Bokshan, S. L., Han, A. L., DePasse, J. M., Eltorai, A. E., Marcaccio, S. E., Palumbo, M. A., & Daniels, A. H. (2016). Effect of

sarcopenia on postoperative morbidity and mortality after thoracolumbar spine surgery. Orthopedics, 39(6), e1159-e1164.

Buzug, T. M. (2011). Computed tomography. In Springer handbook of medical technology (pp. 311-342). Springer, Berlin, Heidelberg.

Hong, H. C., Hwang, S. Y., Choi, H. Y., Yoo, H. J., Seo, J. A., Kim, S. G., ... & Choi, K. M. (2014). Relationship between sarcopenia and nonalcoholic fatty liver disease: the Korean Sarcopenic Obesity Study. Hepatology, 59(5), 1772-1778.

Jones, K. I., Doleman, B., Scott, S., Lund, J. N., & Williams, J. P. (2015). Simple psoas cross-sectional area measurement is a quick and easy method to assess sarcopenia and predicts major surgical complications. Colorectal Disease, 17(1), O20-O26.

Shen, W., Punyanitya, M., Wang, Z., Gallagher, D., St.-Onge, M. P., Albu, J., ... & Heshka, S. (2004). Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. Journal of applied physiology, 97(6), 2333-2338.

Zhao, A., Balakrishnan, G., Durand, F., Guttag, J. V., & Dalca, A. V. (2019). Data augmentation using learned transformations for one-shot medical image segmentation. In Proceedings of the ieee/cvf conference on computer vision and pattern recognition (pp. 8543-8553).

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2014). Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Agarap, A. F. (2018). Deep learning using rectified linear units (relu). arXiv preprint arXiv:1803.08375.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of big data, 6(1), 1-48.

ABSTRACT

Accurate Muscle Segmentation Using Deep Learning Method in 2D Abdominal CT Image

Cho, Da-Som

Major in Computer Engineering

Dept. of Computer Engineering

The Graduate School

Hansung University

Quantification of muscles in the body is an essential factor in evaluating human health. For a long time, various studies have been conducted to effectively evaluate human health, and among them, qualitative evaluation of muscles using image forms has developed. CT images are useful for evaluating muscle and fat because they allow accurate visualization of muscle and adipose tissue areas, especially muscle measurements based on abdominal CT images are closely related to actual muscle mass. However, the task of segmenting muscles in cross-sectional images requires professional knowledge and complex processes, and is time-consuming, cumbersome, and limited human resources and time have made it difficult to apply to large datasets. In

this paper, we propose a fully automated muscle segmentation method using deep learning in abdominal CT images. As a segmentation model, we used U-net, which is fast and shows good performance for biomedical image segmentation. Training was conducted using learning data that conducted data augmentation using rigid transformation, and areas other than muscles were removed from the muscle area mask through image post-processing in the training results. The learned model showed 98% accuracy beyond the clinically usable accuracy, and using the learned model, it took approximately 1 second to segment the muscles in a single abdominal CT image.

KEYWORD: Deep learning, image processing, Segmentation