일반논문 (Regular Paper)
방송공학회논문지 제29권 제5호, 2024년 9월 (JBE Vol.29, No.5, September 2024)
https://doi.org/10.5909/JBE.2024.29.5.676
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

U-Net 기반 의미분할 알고리즘의 개선 기술

김 민 균[a)], 한 종 기[a)‡]

An efficient Semantic Segmentation Algorithm based on U-Net

Min-Gyun Kim[a)] and Jong-Ki Han[a)‡]

요 약

본 논문에서는 U-Net 기반의 의미분할 알고리즘을 개선하여 성능을 향상시키기 위한 새로운 접근 방안을 제시하였다. 제안된 방법은 초해상도(SR) 기술을 도입하여 입력 이미지의 해상도를 개선하고, 컨볼루션 블록 어텐션 모듈(CBAM)을 활용하여 중요한 특징을 강조하며, 푸리에 변환을 통해 전역적 정보를 보존하였다. 이러한 개선 요소들은 의미분할의 정확도를 높이기 위해 효과적으로 작용하였다. 특히, 초해상도 기술은 입력 이미지의 디테일을 살려 의미분할의 정밀도를 높이는 데 기여하였다. 또한, CBAM을 통해 각 특징 맵의 중요한 부분을 강조하여 네트워크의 성능을 극대화하였다. 마지막으로, 푸리에 변환을 사용하여 전역적인 주파수 정보를 보존함으로써 보다 일관된 분할 결과를 얻을 수 있었다. 실험 결과, 제안된 방법은 BCEWithLogitsLoss, IoU, Dice 계수 등 다양한 평가 지표에서 기존의 U-Net 및 다른 변형 모델보다 우수한 성능을 보였다. 특히, 제안된 방법은 IoU와 Dice 계수에서 현저한 성능 향상을 보여주었으며, 이는 실제 의미분할 작업에서의 실용성을 높이는 데 중요한 역할을 하였다. 이러한 결과는 제안된 방법이 의료 영상 분석, 자율 주행, 위성 이미지 분석 등 다양한 응용 분야에서의 의미분할 작업에 기여할 수 있음을 시사한다. 본 논문은 향후 연구에 있어 의미분할 알고리즘의 성능을 더욱 향상시키기 위한 기초 자료로 활용될 수 있을 것이다.

Abstract

This paper presents a novel approach to improving the performance of U-Net-based semantic segmentation algorithms. The proposed method enhances input image resolution using super-resolution (SR) technology, emphasizes key features through the Convolutional Block Attention Module (CBAM), and preserves global information using Fourier Transform. These improvements effectively enhance the accuracy of semantic segmentation. Specifically, the super-resolution technology contributes to increasing the precision of segmentation by preserving details in the input images. Additionally, CBAM maximizes network performance by highlighting important regions in each feature map. Lastly, the use of Fourier Transform allows for more consistent segmentation results by maintaining global frequency information. Experimental results demonstrate that the proposed method outperforms traditional U-Net and other modified models across various evaluation metrics, including BCEWithLogitsLoss, IoU, and Dice coefficient. Notably, the proposed method shows significant improvement in IoU and Dice coefficients, playing a crucial role in enhancing the practical applicability of semantic segmentation tasks. These results suggest that the proposed method can contribute to various applications of semantic segmentation, such as medical image analysis, autonomous driving, and satellite image analysis. This paper can serve as a foundational resource for future research aimed at further enhancing the performance of semantic segmentation algorithms.

Keyword : U-Net, Semantic Segmentation, Super resolution, Fourier Transform, Convolutional Block Attention Module

Ⅰ. 서 론

의미분할(Semantic Segmentation)은 이미지나 비디오에서 각 픽셀이 특정 클래스에 속하는지를 예측하는 컴퓨터 비전 기술이다[1][2]. 이 기술은 단순히 객체의 위치를 파악하는 것을 넘어, 각 픽셀 단위에서 객체를 분류하여 더 정밀하고 세밀한 분석을 가능하게 한다. 이 기술이 중요한 이유는 다음과 같다. 첫째, 의미분할은 고정밀 객체 인식이 요구되는 응용 분야에서 필수적인 역할을 한다. 자율 주행, 의료 영상 분석, 정밀 농업 등 여러 분야에서 높은 수준의 정확도를 요구하는 작업들이 많다. 둘째, 의미분할 기술은 다른 컴퓨터 비전 기술들과 결합하여 더욱 강력한 시스템을 만들 수 있다. 예를 들어, 객체 탐지(Object Detection)와 결합하면 객체의 위치뿐만 아니라 형태와 경계를 정확히 파악할 수 있다. 셋째, 최근 딥러닝 기술의 발전으로 의미분할의 정확도와 효율성이 크게 향상되었으며, 다양한 실제 응용 분야에서의 활용 가능성이 증대되고 있다. 이와 같은 이유로, 의미분할은 현재와 미래의 다양한 응용 분야에서 중요한 역할을 수행하는 핵심 기술로 자리 잡고 있다. 최근 연구에 따르면, 의미분할 기술에 대한 수요는 매년 20% 이상 증가하고 있으며, 헬스케어, 자동차, 보안 분야에서의 적용이 늘어나고 있다[3]. 이러한 추세는 의미분할이 향후 인공지능 기술 발전에서 중요한 역할을 할 것임을 시사한다.

지금까지 의미분할을 연구해온 많은 연구자들이 있었는데, [4]에서는 Fully Convolutional Networks (FCN)을 제안하여, 이미지 분할 문제를 처음으로 엔드-투-엔드(End-to-End) 방식의 딥러닝 모델로 해결하였다. 이는 기존의 기계학습 기법보다 훨씬 향상된 성능을 보였으며, 이후의 연구에 큰 영향을 끼쳤다. [5]에서는 U-Net을 제안하여, 특히 의료 영상 분석 분야에서 뛰어난 성능을 보이는 모델을 개발하였다. U-Net은 대칭적인 인코더-디코더 구조를 가지며, 세밀한 정보 손실을 최소화하기 위해 스킵 연결을 사용하였다. 이로 인해 적은 양의 학습 데이터셋에서도 높은 성능을 유지할 수 있었다. [6]에서는 U-Net의 성능을 더욱 향상시키기 위해 U-Net++를 제안하였다. U-Net++는 중첩된 스킵 경로와 깊이 방향으로 조밀한 연결을 추가하여 분할 성능을 개선하였다. 이 모델은 특히 복잡한 구조를 가진 이미지에서 더 높은 정확도를 보였으며, 다양한 의료 영상 분할 작업에서 성공적으로 적용되었다.

지금까지 연구된 연구들은 다음과 같은 한계들이 있었다. [4]의 FCN은 전역적인 맥락 정보를 충분히 반영하지 못하여, 세밀한 분할에서는 한계를 보였다. [5]의 U-Net은 적은 양의 훈련 데이터셋에서도 성능이 우수하였지만, 복잡한 장면에서는 성능이 저하되는 문제가 있었다. [6]의 U-Net++는 이러한 문제를 어느 정도 해결하였으나, 여전히 해상도 제한과 중요한 특징 강조의 부족이라는 한계가 존재하였다. 이러한 점을 해결하고자 본 논문에서는 입력 영상에 초해상도(Super Resolution) 기술을 도입함과 동시에 U-Net 기반 모델에 CBAM 및 Fourier 변환 기법을 결합하여, 전역적이고 세밀한 정보 모두를 효과적으로 활용할 수 있는 알고리즘을 제안한다. 이를 통해 기존 연구들이 해결하지 못한 문제들을 보완하고, 다양한 응용 분야에서 높은 성능을 발휘할 수 있는 의미분할 모델을 개발하고자 한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 기존의 U-net을 이용한 의미분할 연구를 소개하고 기존 연구의 문제점을 파악한다. Ⅲ장에서는 기존 연구에서 발생한 문제점을 보안하는 방안과 더불어, 새로운 구조의 U-net 구조를 제안한다. Ⅳ장에서는 데이터 생성 과정과 실험 환경을 설명하고 기존의 구조를 사용하여 학습한 결과와 제안한 구조를 사용하여 학습한 결과를 비교 및 분석한다. Ⅴ장에서는 결론을 내린다.

a) 세종대학교 전자정보통신공학과(Electrical Engineering, Sejong University)
‡ Corresponding Author : 한종기(Jong-Ki Han)
E-mail: hjk@sejong.edu
Tel: +82-2-3408-3739
ORCID: https://orcid.org/0000-0002-5036-7199
※ This work was supported by the National Research Foundation of Korea (NRF) under Grant 2022R1F1A1071513 funded by the Korea government through the Ministry of Science and ICT (MSIT).
· Manuscript July 17, 2024; Revised August 16, 2024; Accepted August 19, 2024.

Ⅱ. 기존 연구

1. Fully Convolutional Networks

기존 연구 [4]에서는 Fully Convolutional Networks

(FCN)의 구조를 사용하여 의미분할을 진행하고 있다. FCN은 전통적인 컨볼루션 신경망(CNN)을 의미분할 작업에 맞게 확장한 모델로, 이미지의 각 픽셀을 특정 클래스에 할당하는 문제를 처음으로 엔드-투-엔드 방식으로 해결하였다. FCN은 전통적인 CNN의 구조에서 완전 연결 계층(Fully Connected Layer)을 제거하고, 대신 모든 계층을 컨볼루션 레이어로 구성하여 입력 이미지에서 출력 세그멘테이션 맵까지 모든 과정을 하나의 네트워크로 처리한다. 이는 기존의 이미지 분할 기법들과 달리, 특징 추출, 분류 등의 단계를 별도로 나누지 않고 통합적으로 학습하는 방식을 도입한 것이다.

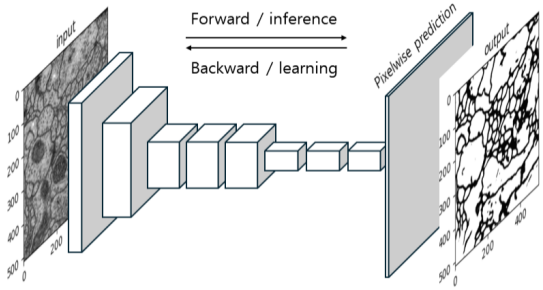

그림 1. FCN의 구조
Fig. 1. FCN Structure

그림 1의 FCN의 구조는 인코더-디코더 형태로 구성되어 있다. 첫 번째 pass인 다운샘플링 패스(down-sampling pass)에서는 이미지의 전역적인 의미를 분석하여 이미지의 주요 특징을 추출하고, 두 번째 pass인 업샘플링 패스(up-sampling pass)에서는 이러한 전역적인 의미를 바탕으로 각 픽셀의 로컬 의미를 분석하여 원래 해상도로 복원하면서 세밀한 분할을 수행한다. 이 과정에서 FCN은 업샘플링 레이어를 사용하여 풀링 계층에서 손실된 공간 정보를 복원하고 원래 해상도로 되돌리는 역할을 한다.

FCN은 엔드-투-엔드 학습을 통해 이미지의 크기에 관계없이 다양한 해상도의 입력을 처리할 수 있으며, 이는 전통적인 방법론보다 훨씬 높은 성능을 보인다. 그러나 FCN은 전역적인 맥락 정보를 충분히 반영하지 못하여, 복잡한 장면이나 세밀한 객체 경계의 분할에서는 한계를 보인다. 이는 주로 업샘플링 과정에서의 정보 손실과 관련이 있다.

2. U-net

기존 연구 [5]에서 제안하는 U-Net의 주요 특징은 대칭적인 인코더-디코더 구조와 스킵 연결(Skip Connections)이다. 인코더 부분에서는 입력 이미지에서 중요한 특징을 추출하기 위해 컨볼루션 연산과 풀링 연산이 반복적으로 적용된다. 디코더 부분에서는 업샘플링을 통해 원래의 해상도로 복원하면서 세밀한 특징을 강조한다. 이 과정에서 인코더 단계에서 추출된 특징 맵을 디코더 단계로 직접 연결하는 스킵 연결이 사용되어, 고해상도의 정보를 유지하며 세밀한 분할을 가능하게 한다.

그림 2의 U-Net의 구조는 전형적인 인코더-디코더 네트워크와 유사하지만, 각 인코더 단계의 출력이 대응하는 디코더 단계로 전달되면서 고해상도 정보를 보존하는 스킵 연결이 추가된다. 이러한 스킵 연결은 정보의 손실을 줄이고, 디테일한 경계와 작은 객체들을 정확하게 분할하는 데 중요한 역할을 한다. 특히, U-Net은 작은 데이터셋에서도 높은 성능을 발휘할 수 있어, 의료 영상 분석과 같은 분야에

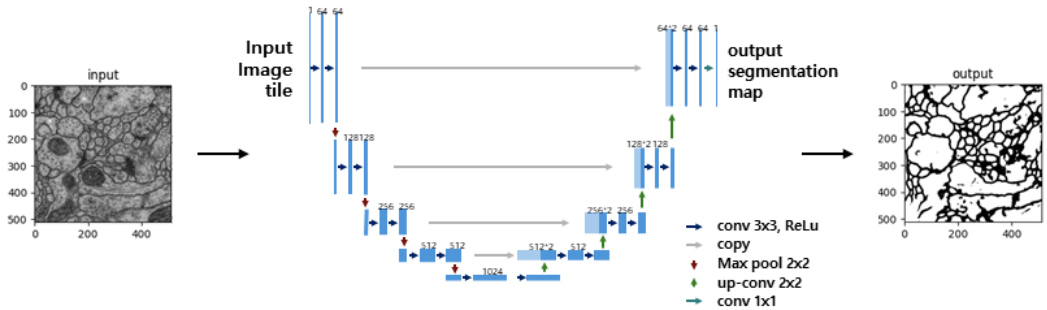

그림 2. U-net 기반 의미분할 알고리즘 구조[5]
Fig. 2. U-net based semantic segmentation algorithm structure[5]

서 널리 사용되고 있다.

U-Net의 구조를 보면, 먼저 인코더 부분에서 입력 이미지를 반복적인 컨볼루션과 풀링 과정을 통해 점진적으로 축소하여 점진적으로 축소하여 중요한 특징을 추출한다. 컨볼루션 연산을 수식적으로 나타내면 다음과 같다.

$$Y_{conv}(i,j) = (X^{*}K)(i,j) = \sum_{m}\sum_{n} X(i+m,j+n)\cdot K(m,n) \quad (1)$$

여기서 X는 입력 이미지, K는 필터(커널), Y_{conv}는 출력 이미지이다. 이 연산은 입력 이미지 X에 필터 K를 적용하여 출력 이미지 Y_{conv}를 생성하는 과정을 나타낸다. 여기서 i와 j는 출력 이미지의 픽셀 위치를 나타내며, m과 n은 필터의 크기를 나타낸다.

이후 디코더 부분에서는 업컨볼루션을 통해 이미지를 점진적으로 확대하며, 인코더에서 전달된 스킵 연결을 통해 고해상도 정보를 결합하여 원래 해상도의 세그멘테이션 맵을 복원한다. 다음은 업컨볼루션 연산과정이다.

$$Y_{upconv}(i,j) = (X^{*}K^{T})(i,j) = \sum_{m}\sum_{n} X(i-d_i+m,j-d_j+n)\cdot K(m,n) \quad (2)$$

여기서 X는 입력 이미지, K^{T}는 트랜스포즈(Transpose) 필터(커널), Y_{upconv}는 출력 이미지이다. 이 연산은 입력 이미지 X를 업컨볼루션 필터 K^{T}로 확장하여 출력 이미지 Y_{upconv}를 생성하는 과정을 나타낸다. 여기서 d_i와 d_j는 업컨볼루션 과정에서의 스트라이드(stride)를 나타낸다.

U-Net은 전역적인 문맥 정보와 세밀한 로컬 정보를 모두 효과적으로 통합하여 높은 정확도의 분할 결과를 제공한다. 그러나 복잡한 장면에서는 여전히 성능이 저하되는 문제가 존재하며, 이러한 문제를 해결하기 위해 다양한 개선 기법들이 제안되고 있다.

3. U-Net++

기존 연구 [6]에서는 U-Net의 성능을 더욱 향상시키기 위해 U-Net++를 제안하였다. U-Net++의 주요 특징으로는 중첩된 스킵 경로(nested skip pathways)를 사용하여 서로 다른 해상도 수준에서 특징 맵을 통합한다는 점이다. 그림 3에 포함된 녹색의 점선이 중첩된 스킵 경로를 나타내며, 주황색의 원들은 이를 통해 생성된 특징 맵을 나타낸다. 이러한 구조는 의미분할에 있어 세밀한 구조와 복잡한 패턴을 더욱 효과적으로 분할할 수 있게 한다.

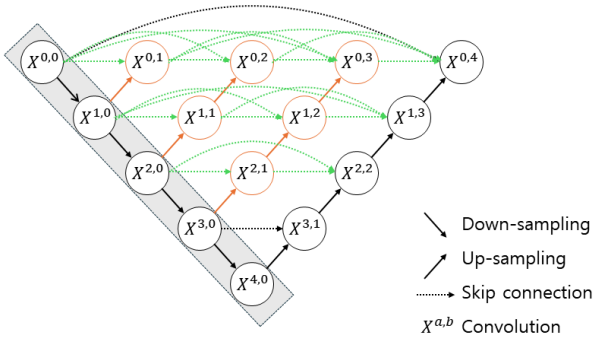

그림 3. 중첩된 스킵 경로를 포함한 U-net++ 구조
Fig. 3. U-net++ structure with nested skip pathways

중첩된 스킵 경로는 다음과 같은 수식으로 나타낼 수 있다.

$$x^{a,b} = \begin{cases} Y_{conv}(x^{a-1,b}), & \text{if } b=0 \\ Y_{conv}\left(\left[x^{a,k}\right]_{k=0}^{b-1}, Y_{upconv}(x^{a+1,b-1})\right), & else \text{ if } b>0 \end{cases} \quad (3)$$

여기서 $x^{a,b}$는 깊이 a와 레벨 b에서의 특징 맵을 나타내며, 이 수식은 서로 다른 해상도에서의 특징 맵을 결합하여 더 풍부한 정보를 포함하는 새로운 특징 맵을 생성하는 과정을 설명한다. Y_{conv}과 Y_{upconv}는 각각 식 (1), (2)에 해당하며 컨볼루션과 업컨볼루션의 연산을 나타낸다.

U-Net++는 더 깊은 네트워크 구조를 가지며, 이는 더 많은 컨볼루션 레이어와 조밀한 연결을 포함하여, 더 높은 수준의 특징을 추출하고 복잡한 패턴을 학습할 수 있게 한다. U-Net++는 이러한 구조적 개선을 통해 기존 U-Net의 한계를 극복하고, 더 높은 정확도의 의미분할 결과를 제공한다. 특히, 복잡한 의료 영상 분석 작업에서 뛰어난 성능을 발휘하며, 다양한 실제 응용 분야에서 성공적으로 적용되고 있다. 기존의 FCN, U-Net과 비교할 때, U-Net++는 중첩된 스킵 경로와 조밀한 연결구조로 인해 복잡한 구조를 가

진 이미지에서도 더 높은 정확도를 제공하며, 복잡한 장면이나 세밀한 객체 경계의 분할에서도 높은 성능을 발휘 한다.

그러나 U-Net++는 중첩된 스킵 경로와 조밀한 연결을 도입함으로써 네트워크의 복잡성이 증가한다는 한계가 있다. 이는 더 많은 연산 자원과 메모리를 요구하며, 학습 시간과 추론 시간이 길어질 수 있다. 또한 이와 같은 복잡한 구조로 인해 연산 비용이 높아져, 실시간 적용이 어려울 수 있다. 특히, 의료 영상 분석과 같이 대용량 데이터셋을 처리해야 하는 경우, 계산 비용이 큰 부담으로 작용할 수 있다.

Ⅲ. 제안하는 방법

1. 제안 기술의 요약

이 시스템은 SR을 이용한 해상도 개선, 푸리에 변환을 이용한 정보 보존, 그리고 컨볼루션 블록 어텐션 모듈(CBAM)을 결합하여 기존의 U-Net 구조를 개선한 모델로 구성된다. 그림 4 (a)의 경우 초해상도(SR) 기술을 이용하여 입력 이미지의 해상도를 개선하며, (b)의 경우 CBAM을 통해 중요한 채널과 위치 정보를 강조하여 분할 성능을 향상시킨다. (c)는 푸리에 변환을 사용하여 전역적이고 세밀한 정보를 스킵 연결을 통해 효과적으로 활용한다.

2. 초해상도 (Super Resolution)

초해상도(SR)는 저해상도 이미지를 고해상도로 변환하는 기술로, 의미분할의 정확도를 높이기 위해 중요한 역할을 한다. 기존 기술인 FCN과 U-Net에서는 해상도 문제가 발생하여 정확한 분할이 어려웠다[4][5]. 저해상도 이미지에서는 세밀한 특징이 손실될 수 있으며, 이는 객체 경계와 작은 구조를 정확히 분할하는 데 어려움을 초래한다. SR 기술을 통해 입력 이미지의 해상도를 개선하면, 분할 네트워크가 더 많은 세밀한 정보를 학습할 수 있게 되어, 최종 분할 결과의 정확도가 향상된다.

SR 기술의 대표적인 기법으로는 bicubic interpolation, Lanczos resampling 등이 있다[7][8]. 다른 방식으로는 기하학적 변환 기반 기법(Geometry Transformation Method)으로 저해상도 이미지를 여러 방향으로 변환한 후 고해상도 이미지를 재구성하는 기법이다[9]. 다음으로는 신경망 기반 기법(Deep Learning Method)으로 최근에는 딥러닝을 이용

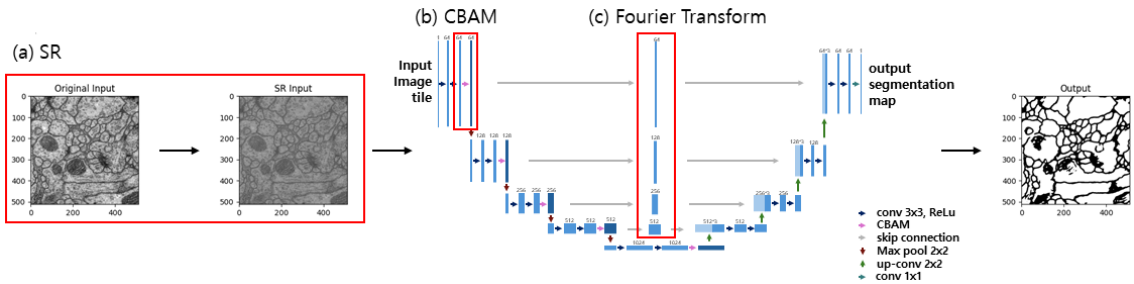

그림 4. 개선된 U-net 기반 의미분할 알고리즘 구조
Fig. 4. Improved U-net based semantic segmentation algorithm structure

표 1. SRCNN (super-Resolution Convolution Neural Network) 구성 단계
Table 1. SRCNN (super-Resolution Convolution Neural Network) construction steps

Order	Step	Function
1	Patch Extraction	Extract small patches from the input low-resolution image and map them to a high-dimensional space.
2	Non-Linear Mapping	Perform a nonlinear mapping that converts the input patches in the high-dimensional space into high-resolution patches.
3	Reconstruction	Combine the transformed high-resolution patches to reconstruct the final high-resolution image.

한 SR 기술이 많이 사용되고 있으며, 특히 컨볼루션 신경망(CNN)이 그 중 가장 두드러진 성과를 보이고 있다[10]. 본 논문에서는 SRCNN (Super-Resolution Convolutional Neural Network)[10], EDSR (Enhanced Deep Residual Networks for Single Image Super-Resolution)[11], ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks)[12]과 같은 다양한 SR 기법을 제안하는 기술에 적용하였다.

SRCNN[10] 기술의 경우 표 1과 같은 3단계의 네트워크로 구성되며, 식 (4)와 같이 표현된다.

$$F(Y) = W_3 * \max(0, W_2 * \max(0, W_1 * Y + B_1) + B_2) + B_3 \quad (4)$$

여기서 Y는 입력 저해상도 이미지, $F(Y)$는 출력 고해상도 이미지, W_i와 B_i는 각 레이어의 가중치와 바이어스를 나타낸다. 이와 같은 네트워크 구조는 저해상도 이미지의 고빈도 성분을 효과적으로 복원하여, 의미분할의 정확도를 높이는 데 기여한다.

EDSR[11]의 경우 고해상도 이미지를 생성하기 위해 잔차 네트워크를 강화한 모델이다. 이 모델은 기존의 초해상도 네트워크에 비해 더 깊은 구조를 가지고 있으며, 배치 정규화를 제거하여 성능을 최적화하고 있다. EDSR은 이를 통해 입력 이미지의 세부 사항을 더 정밀하게 복원하며, 결과적으로 고해상도 이미지에서 더 선명하고 세밀한 표현을 가능하게 하고 있다. EDSR의 이러한 특성은 특히 고주파수 정보를 유지해야 하는 의미분할 작업에서 유용하게 적용될 수 있다.

ESRGAN[12]은 perceptual loss를 개선하고, Residual-in-Residual Dense Block (RRDB)를 도입하여 더 깊은 네트워크를 효과적으로 훈련시킬 수 있도록 한다. Perceptual loss는 VGG 네트워크의 중간 특징 맵을 사용하여 주관적인 화질을 높이고, 수식으로는 다음과 같이 표시된다.

$$L_{percep} = \sum_{i=1}^{N} \lambda_i \| \phi_i(I_{HR}) - \phi_i(I_{SR}) \|_2^2 \quad (5)$$

여기서 ϕ_i는 VGG 네트워크의 i번째 레이어에서 추출된 피처 맵을 나타내며, I_{HR}은 고해상도(ground truth) 이미지, I_{SR}은 생성된 초해상도 이미지를 나타낸다. λ_i는 각 레이어의 중요성을 나타내는 가중치이다.

RRDB는 여러 dense block을 통합하여 정보의 손실을 줄이며, 더욱 풍부하고 세밀한 특징을 추출할 수 있게 한다. RRDB는 다음과 같은 수식으로 표현된다.

$$F_{RRDB}(X) = X + B_n(B_{n-1}(\cdots B_1(X))) \quad (6)$$

여기서 B_i는 각 Residual Block을 나타내며, F_{RRDB}는 전체 Residual-in-Residual Dense Block의 출력을 의미한다. 이 블록은 입력 x와 블록 내에서 반복적으로 계산된 출력을 더하여 최종 출력을 생성한다.

ESRGAN은 최적화 과정에서 relativistic discriminator를 사용하여 생성된 이미지와 실제 이미지 간의 차이를 더 현실감 있게 학습할 수 있도록 한다. Relativistic Discriminator의 Loss는 다음과 같이 정의된다.

$$L_{relativistic} = -E[\log(D_{rel}(I_{HR}, I_{SR}))] \quad (7)$$

여기서 $D_{rel}(I_{HR}, I_{SR})$은 상대적인 진실도를 나타내는 함수로, 이 함수는 실제 이미지와 생성된 이미지의 차이를 평가한다. 이러한 기술적 개선을 통해 ESRGAN은 이미지의 고주파수 정보를 더욱 잘 보존하고, 세부 묘사가 뛰어난 고해상도 이미지를 생성할 수 있다.

3. 컨볼루션 블록 어텐션 모듈 (Convolutional Block Attention Module, CBAM)

컨볼루션 블록 어텐션 모듈(CBAM)은 피드포워드 컨볼루션 신경망의 성능을 향상시키기 위한 간단하면서도 효과적인 어텐션 모듈이다. CBAM은 중간 피처 맵을 입력으로 받아 채널 어텐션과 공간 어텐션을 순차적으로 계산하여 입력 피처 맵을 적응적으로 보정(refinement)한다. CBAM은 경량 모듈로, 거의 모든 CNN 아키텍처에 매끄럽게 통합할 수 있으며 기본 CNN과 함께 End-to-End 방식으로 학습할 수 있다. CBAM은 다양한 모델에서 일관된 성능 향상을 보여주었으며, 분류 및 검출 성능을 개선하는 데 광범위하게 적용될 수 있음이 입증되었다[13].

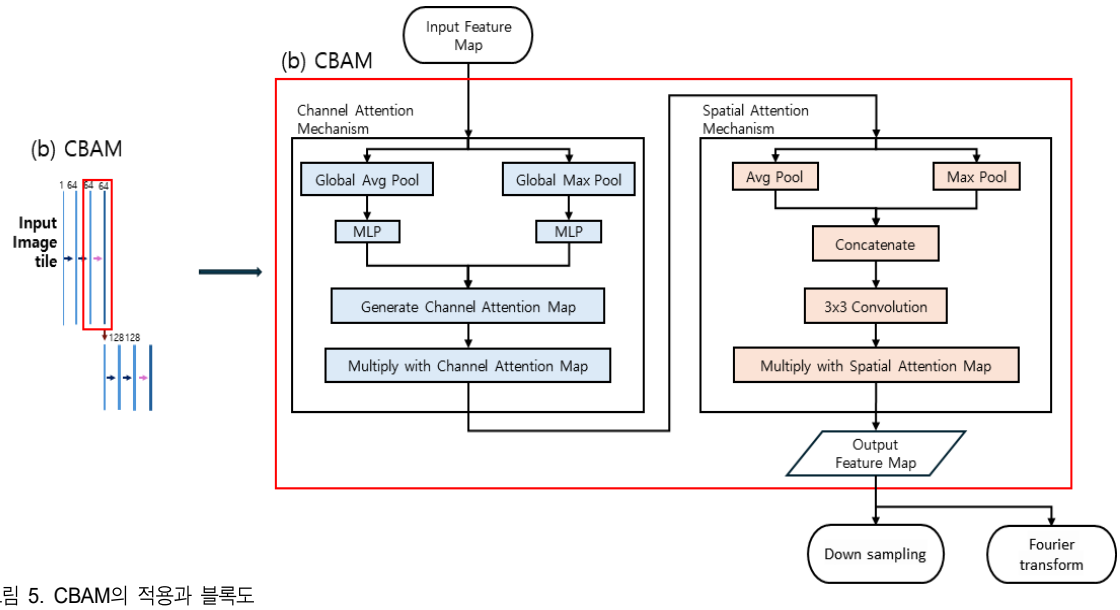

그림 5. CBAM의 적용과 블록도
Fig. 5. Application and block diagram of CBAM

그림 5의 경우 본 논문의 아이디어 중 CBAM이 사용되는 방식과 블록도를 나타낸다. CBAM의 과정은 피처 맵을 입력으로 받아 채널 어텐션 과정과 공간 어텐션 과정을 거치게 된다. 채널 어텐션 과정의 경우 전역 평균 풀링 및 전역 최대 풀링 후, 다층 퍼셉트론(MLP)을 통해 얻은 가중치를 통해 채널 주의맵을 생성 하게 된다. 이후 이를 입력 피처 맵과 곱하여 채널 주의가 적용된 피처 맵을 생성하는 과정을 거친다. 공간 어텐션 과정의 경우 평균 풀링 및 최대 풀링 후, 두 풀링 결과를 합쳐서 하나의 피처 맵을 만든다. 이후 합쳐진 피처 맵에 대해 3x3 합성곱 연산을 적용하여 공간 주의 맵을 생성하게 되고, 채널 주의가 적용된 피처 맵과 공간 주의 맵을 곱하여 최종 출력 피처 맵을 얻게 된다. 이는 각각 다운 샘플링과 스킵 연결을 위한 푸리에 변환에 적용된다.

CBAM의 작동 원리는 다음과 같다. 차원의 크기가 각각 C(채널 수), H(높이), W(너비)의 경우, 중간 피처 맵 $F \in \mathbb{R}^{(C \times H \times W)}$이 주어지면, CBAM은 채널 어텐션 맵 $Mc \in \mathbb{R}^{(C \times 1 \times 1)}$과 공간 어텐션 맵 $Ms \in \mathbb{R}^{(1 \times H \times W)}$을 순차적으로 추론한다. 전체 어텐션 과정은 다음과 같이 요약할 수 있다.

$$F' = Mc(F) \otimes F \tag{8}$$

$$F'' = Ms(F') \otimes F' \tag{9}$$

여기서 $\otimes$는 요소별 곱셈을 나타낸다. 곱셈 동안 어텐션 값은 각 채널과 공간 차원에 맞게 브로드캐스트된다. F''는 최종 리파인된 출력이다.

CBAM의 구조는 채널 어텐션 모듈, 공간 어텐션 모듈 두 가지 주요 구성 요소로 이루어진다. 채널 어텐션 맵은 피처의 채널 간 관계를 이용하여 생성된다. 각 채널이 특정 피처 디텍터로 간주될 때, 채널 어텐션은 주어진 입력 이미지에서 중요한 채널을 강조한다. 채널 어텐션을 효율적으로 계산하기 위해 입력 피처 맵의 공간 차원을 압축한다. 평균 풀링과 최대 풀링을 사용하여 공간 정보를 집계한 후, 이를 통해 생성된 두 개의 공간 컨텍스트 디스크립터를 공유 네트워크에 전달하여 채널 어텐션 맵 Mc를 생성한다. 이 과정은 다음과 같이 수식으로 나타낼 수 있다.

$$Mc(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \tag{10}$$

여기서 σ는 시그모이드 함수, MLP는 다층 퍼셉트론, AvgPool과 MaxPool은 각각 평균 풀링과 최대 풀링을 나타낸다. 공간 어텐션 맵은 피처의 공간 간 관계를 이용하여 생성된다. 공간 어텐션은 입력 이미지에서 중요한 위치를

강조한다. 공간 어텐션을 계산하기 위해, 채널 축을 따라 평균 풀링과 최대 풀링을 적용하여 두 개의 2D 맵을 생성하고 이를 결합한 후, 7×7 필터 크기의 컨볼루션 연산을 적용하여 최종 공간 어텐션 맵 Ms를 생성한다. 이 과정은 다음과 같이 수식으로 나타낼 수 있다.

$$Ms(F) = \sigma(f_{7\times7}([AvgPool(F);MaxPool(F)])) \quad (11)$$

여기서 σ는 시그모이드 함수, $f_{7\times7}$는 7×7 필터 크기의 컨볼루션 연산을 나타낸다. CBAM의 채널 어텐션과 공간 어텐션 모듈은 순차적으로 배치되어 더 나은 성능을 발휘한다. 이 모듈을 통해 네트워크는 중요한 피처를 강조하고 불필요한 피처를 억제하여 의미분할의 정확도를 향상시킬 수 있다. 본 논문에서는 CBAM을 인코더 단계에서 각 블록의 출력에 적용하여, 중요한 채널 및 공간 정보를 강조함으로써 의미분할의 성능을 향상시킨다. CBAM은 인코더 블록의 출력을 리파인하고, 이를 푸리에 변환과 결합하여 디코더로 전달하는 과정에서 중요한 역할을 한다.

4. 푸리에 변환 (Fourier Transform)

푸리에 변환(Fourier Transform)은 이미지 처리에서 주파수 도메인으로 변환하여 전역적이고 세밀한 정보를 분석하는 강력한 도구이다[3]. 본 논문에서는 의미분할의 정확도를 높이기 위해 푸리에 변환을 U-Net의 스킵 연결(Skip Connections)에 활용한다. 위에서 언급한 CBAM 과정을 거쳐 중요한 피처가 강조된 이미지에 푸리에 변환을 통해 이미지의 고주파 성분을 강조하고, 불필요한 저주파 성분을 제거하여 피처맵을 생성하고 이를 스킵 연결을 통해 전달함으로서 의미분할의 성능을 향상시킨다.

그림 6의 경우 본 논문의 아이디어 중 푸리에 변환이 사용되는 방식과 블록도를 나타낸다. 본 논문에서는 푸리에 변환을 적용한 방식의 자세한 설명은 다음과 같다. 우선 CBAM의 결과인 입력 이미지의 각 채널에 대해 이차원 푸리에 변환(2D Fourier Transform)을 수행하여 주파수 도메인으로 변환한다. 이후 주파수 중심을 시프트하여 저주파 성분을 제거하고 고주파 성분을 강조한다. 변환된 주파수 도메인을 역 푸리에 변환(Inverse Fourier Transform)하여

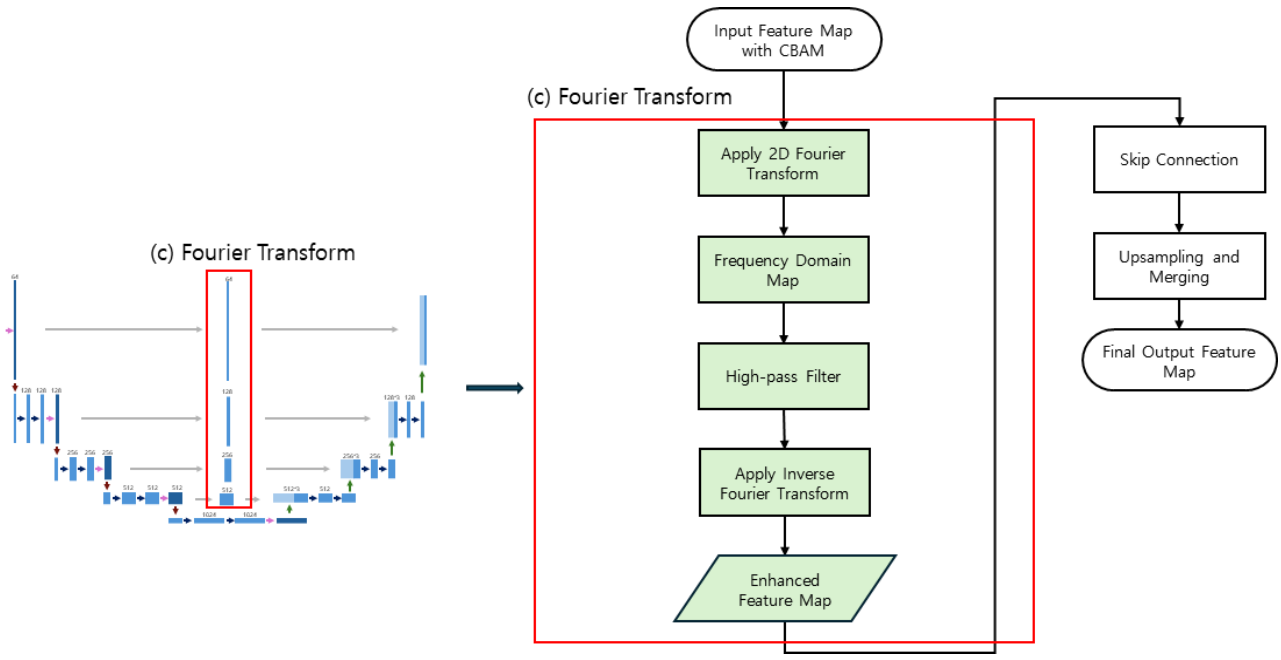

그림 6. 푸리에 변환의 적용과 블록도
Fig. 6. Application and block diagram of Fourier transform

다시 공간 도메인으로 복원한다. 이 과정에서 주파수 도메인에서의 저주파 성분을 제거함으로써, 전역적인 문맥 정보를 보존하면서도 세밀한 로컬 정보를 효과적으로 활용할 수 있다. 특히, 인코더와 디코더 사이의 스킵 연결에 푸리에 변환을 적용함으로써, 고해상도의 세밀한 정보를 보다 효과적으로 전달할 수 있게 된다.

Ⅳ. 실험 결과 및 분석

1. 알고리즘 구현 및 실험 조건

제안한 아이디어를 실험하기 위해 공개된 오픈 소스 코드를 사용하였다. 본 연구에서는 [5]에서 제시된 U-Net을 기반으로 한 딥러닝 모델 구현체를 활용하였다. 해당 구현체는 GitHub에서 제공되는 youtube-cnn-002-pytorch-unet 레포지토리에서 가져왔다. 이를 바탕으로, 초해상도 기술(SR)과 컨볼루션 블록 어텐션 모듈(CBAM)을 추가하여 U-Net 구조를 개선하였다. 구체적으로, SR 모델을 사용하여 입력 이미지를 초해상도로 변환한 후, 변환된 이미지를 U-Net의 입력으로 사용하였다. 또한, CBAM을 각 인코더 블록에 추가하여 중요한 특징을 강조하고, 푸리에 변환을 이용하여 주파수 도메인에서 전역적 정보를 보존하였다.

딥러닝 실험을 위해 “ISBI 2012 EM segmentation Challenge”에서 사용된 membrane 데이터셋을 활용하였다. 데이터셋은 위 GitHub 레포지토리 내에 포함된 .tif 파일로부터 가져왔다. 이 파일에는 512x512 크기의 이미지가 30개 포함되어 있으며, 이를 각각 .npy 파일로 변환하여 학습, 검증, 테스트 데이터로 분류하였다. 각 이미지는 학습 데이터는 24개, 검증 데이터는 3개, 테스트 데이터는 3개로 나누어진다. 또한, 추가적인 검증을 위해 Broad Bioimage Benchmark Collection (BBBC)에서 제공하는 BBBC007 데이터를 사용하여 실험을 진행하였다. 이 데이터셋은 세포핵의 플루오레선스 이미지를 포함한 데이터셋으로, 형광현미경으로 촬영된 이미지들로 구성되어 있다.

훈련 과정에서는 전체 모델을 처음부터 학습시키는 대신, 사전 학습된 SR 모델을 사용하여 입력 이미지를 초해상도로 변환하고, 변환된 이미지를 U-Net의 입력으로 사용하였다. 이 과정을 통해 기존의 U-Net 구조에 CBAM과 푸리에 변환을 결합한 새로운 구조를 fine-tuning하였다. 네트워크의 파라미터는 초기화했으며, 학습률은 1e-3으로 설정하고 Adam 옵티마이저를 사용하여 최적화하였다. 또한, 학습 과정에서는 BCEWithLogitsLoss를 손실 함수로 사용했으며, 각 에포크마다 모델의 성능을 평가하기 위해 검증 데이터를 사용하여 IoU와 Dice 계수를 계산하였다.

2. 실험 결과 분석 지표

의미 분할 모델의 성능을 평가하기 위해 다양한 평가 지표가 사용된다. 이 연구에서는 Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss), Intersection over Union (IoU), 그리고 Dice 계수를 주요 지표로 사용하였다. 각 지표의 의미와 특징은 다음과 같다.

BCEWithLogitsLoss는 이진 분류 문제에서 사용되는 손실 함수이다. 이 함수는 이진 교차 엔트로피 손실(Binary Cross-Entropy Loss)과 로짓스(Logits)를 결합한 형태로, 로짓스는 예측된 확률 값에 시그모이드 함수를 적용하기 전의 값이다. BCEWithLogitsLoss는 다음과 같이 계산된다.

$$BCEWithLogitsLoss = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\sigma(x_i)) + (1-y_i)\log(1-\sigma(x_i))] \quad (12)$$

여기서 y_i는 실제 레이블, x_i는 예측된 로직스, σ는 시그모이드 함수, N은 데이터의 개수이다. 이 손실 함수는 값이 작을수록 모델의 예측이 실제 레이블과 가깝다는 것을 의미한다.

IoU는 예측된 분할 영역과 실제 정답 분할 영역 간의 겹치는 부분을 측정하는 평가 지표이다. IoU는 다음과 같이 계산된다.

$$IoU = \frac{TruePositive(TP)}{TruePositive(TP) + FalsePositive(FP) + FalseNegative(FN)} \quad (13)$$

$$Dice = \frac{2 \cdot TruePositive(TP)}{2 \cdot TruePositive(TP) + FalsePositive(FP) + FalseNegative(FN)} \quad (14)$$

여기서 TP는 정확하게 예측된 영역, FP는 잘못 예측된 영역, FN은 놓친 영역을 의미한다. IoU 값은 0에서 1 사이의 값을 가지며, 값이 클수록 예측이 정확함을 의미한다. 일반적으로 IoU 값이 0.5 이상이면 예측이 유의미하다고 간주된다. IoU는 모델이 실제 객체를 얼마나 잘 탐지하고 분할하는지를 평가하는 데 유용하다.

Dice 계수는 두 샘플 간의 유사성을 측정하는 지표로, 주로 이진 분할 문제에서 사용된다. Dice 계수는 0에서 1 사이의 값을 가지며, 값이 클수록 예측이 정확함을 의미한다. Dice 계수는 특히 데이터셋이 불균형할 때 더 유리한 평가 지표로 작용한다. 이 지표는 분할 영역의 중첩도를 평가하여 모델이 얼마나 잘 분할했는지를 나타낸다. Dice 계수는 식 (14)와 같이 계산된다.

3. 실험 결과 및 분석

표 2과 그림 7은 다양한 SR 기술을 적용했을 때의 의미분할 정확도를 비교한 결과를 나타낸다. 여기서는 기본 U-Net 모델과 SR 기술로 SRCNN, EDSR, ESRGAN을 각각 결합하여 성능 변화를 분석하였다. Loss는 분할된 결과와 실제 레이블 간의 손실을 측정하는 지표로, 값이 작을수록 모델의 예측이 실제 레이블과 가까움을 의미한다. IoU는 예측된 분할 영역과 실제 분할 영역 간의 겹침 정도를 나타내며, 값이 클수록 좋은 성능을 의미한다. Dice 계수는 두 샘플 간의 유사성을 측정하는 지표로써 값이 클수록 좋은 성능을 의미한다.

표 2. 다양한 SR 기술 적용에 따른 의미분할 정확도 분석
Table 2. Analysis of semantic segmentation accuracy using various SR technologies

	Loss	IoU	Dice
U-Net	0.2186	0.8985	0.9464
U-Net with SRCNN	0.2084	0.8884	0.9409
U-Net with EDSR	0.2082	0.8921	0.9399
U-Net with ESRGAN	0.2079	0.9021	0.9485

표 2의 결과에서 SRCNN을 결합한 U-Net 모델은 약간의 성능 저하를 보였다. 손실 함수 값이 0.2084로 감소했지만, IoU와 Dice 계수는 각각 0.8884와 0.9409로, 기본 U-Net에 비해 약간 낮아졌다. 이는 SRCNN이 일부 세부 정보를 보존하는 데 기여했지만, 전체적인 성능 향상에는 크게 기여하지 못했음을 시사한다.

EDSR을 적용한 U-Net 모델에서는 손실 함수 값이 0.2082로 더욱 낮아졌으며, IoU와 Dice 계수는 각각 0.8921과 0.9399로 측정되었다. EDSR은 기본 U-Net에 비해 더 나은 고해상도 정보를 제공했지만, IoU와 Dice 계수의 결과를 보면 여전히 의미분할 성능 향상에는 제한적인 영향을 미친 것으로 나타난다.

마지막으로, ESRGAN을 결합한 U-Net 모델은 가장 우수한 성능을 보였다. 손실 함수 값은 0.2079로 가장 낮았으며, IoU와 Dice 계수는 각각 0.9021과 0.9485로, 모든 평가 지표에서 가장 높은 값을 기록하였다. 이는 ESRGAN이 입력 이미지의 세밀한 디테일을 효과적으로 복원하고, 고해상도 정보를 보존하여 의미분할 성능을 크게 향상시켰음을 나타낸다.

본 논문에서는 위에서 언급한 바와 같이 여러 SR 기술 중에서 성능이 가장 뛰어난 ESRGAN을 사용하여 이후 실험을 진행하였다. ESRGAN을 통해 얻어진 향상된 의미분할 성능은 SR 기술의 중요성을 보여주지만, 이 연구의 주된 기여는 기술의 결합에 대한 접근 방식에 있다. 이는 더 나은

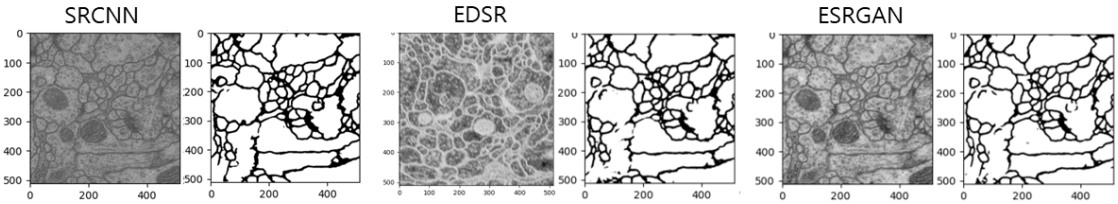

그림 7. 다양한 SR 기술 적용에 따른 의미분할 정성적 결과 비교
Fig. 7. Comparison of the qualitative results of semantic segmentation according to various SR technologies

SR 기술이 개발된다면, 이를 결합하여 더욱 향상된 의미분할 성능을 달성할 수 있음을 시사한다. 따라서, 본 논문에서 제안한 기술 결합 방법론은 향후 SR 기술의 발전에 따라 더 높은 정확도와 일관된 결과를 제공할 수 있는 잠재력을 가지고 있다.

표 3과 그림 8은 U-Net에 다양한 Attention 기술을 적용하여 의미분할 성능을 비교한 결과를 보여준다. 실험에서는 SE-Net과 CBAM을 각각 U-Net 구조에 결합하여 성능 변화를 분석하였다.

표 3. 다양한 Attention 기술 적용에 따른 의미분할 정확도 분석
Table 3. Analysis of semantic segmentation accuracy according to various Attention technologies

	Loss	IoU	Dice
U-Net	0.2186	0.8985	0.9464
U-Net with SENet	0.2178	0.8988	0.9465
U-Net with CBAM	0.1990	0.8980	0.9462

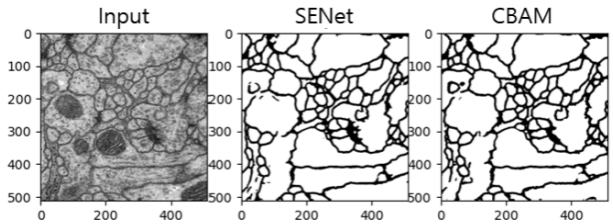

그림 8. 다양한 Attention 기술 적용에 따른 의미분할 정성적 결과 비교
Fig. 8. A comparison of qualitative results of semantic segmentation according to various Attention technologies

SE-Net은 Attention 메커니즘을 사용하여 이미지의 각 채널 간 중요도를 학습하는 방식으로, 주로 ResNet과 같은 구조에 결합되어 사용된다. SE-Net의 Squeeze-and-Excitation 모듈은 각 채널의 특징을 전역적으로 집약(Squeeze)하고, 그 중요도를 조절(Excitation)하여 네트워크가 중요한 정보에 집중할 수 있게 한다. 이러한 방식은 다양한 컴퓨터 비전 작업에서 성능을 크게 향상시키는 것으로 입증되었다. 실험 결과를 보면, U-Net에 SE-Net을 결합했을 때 기본 U-Net에 비해 약간의 성능 향상이 있었지만, IoU와 Dice 계수에서는 큰 차이가 나타나지 않았다.

CBAM은 SE-Net과 유사하게 채널 간 중요도를 학습할 뿐만 아니라, 공간적 Attention 메커니즘도 결합하여 특징 맵의 중요한 영역을 강조한다. 실험 결과, U-Net에 CBAM을 적용한 경우 손실 값이 더 낮고, IoU와 Dice 계수에서도 약간의 향상이 있었다. 이는 CBAM이 중요한 정보의 강조와 공간적 특징을 모두 고려하여 의미분할 성능을 개선할 수 있음을 보여준다.

그러나, 이 연구의 핵심은 단순히 특정 Attention 기술의 적용이 아니라, 다양한 기술의 효과적인 결합에 있다. SE-Net과 CBAM 모두 의미분할 성능을 향상시키는 데 기여할 수 있지만, 더 나은 결과를 얻기 위해서는 새로운 기술과의 결합을 고려해 볼 필요가 있다. 예를 들어, 최신의 Attention 기반 메커니즘인 Vision Transformers (ViT) 혹은 Self-Attention과 같은 기술을 통해 성능을 더욱 발전시킬 수 있을 것이다[14]. 결론적으로, 이 실험 결과는 제안된 방법론이 다양한 Attention 메커니즘과 결합할 수 있음을 보여주며, 더 나은 기술을 적용하면 의미분할 성능이 더욱 향상될 가능성이 있다는 점을 시사한다. 향후 연구에서는 이러한 결합 방법을 더욱 발전시켜, 다양한 응용 분야에서 일관된 성능 향상을 달성할 수 있는 방안을 모색해야 할 것이다.

표 4와 표 5, 그리고 그림 9와 그림 10은 U-Net에 다양한 핵심 기술들을 개별적으로 적용한 경우와, 이들 기술을 결합하여 적용한 경우의 의미분할 성능을 비교한 결과를 보여준다.

표 4. 각 기술 적용에 따른 의미분할 정확도 분석
Table 4. Analysis of performance of the proposed algorithm according to the applied core technologies

	Loss	IoU	Dice
U-Net	0.2186	0.8985	0.9464
U-Net with SR(ESRGAN)	0.2079	0.9021	0.9485
U-Net with CBAM	0.1990	0.8980	0.9462
U-Net with Fourier Transform	0.2131	0.8986	0.9468

표 5. 각 기술 결합에 따른 의미분할 정확도 분석
Table 5. Analysis of semantic segmentation accuracy to the combinations of the core technologies

	Loss	IoU	Dice
U-Net	0.2186	0.8985	0.9464
U-Net with ESRGAN, CBAM	0.1940	0.8998	0.9473
U-Net with CBAM, Fourier Transform	0.1982	0.9042	0.9497
U-Net with ESRGAN, Fourier Transform	0.1992	0.8995	0.9471
Proposed method	0.1895	0.9049	0.9499

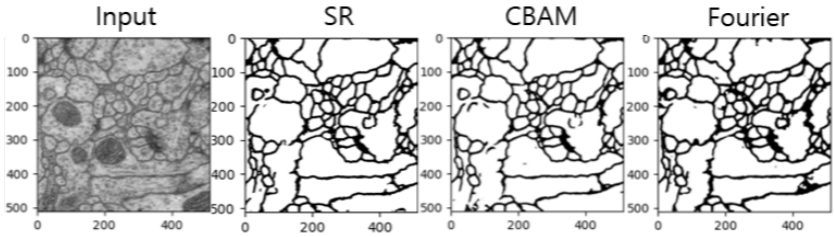

그림 9. 각 기술 적용에 따른 의미분할 정성적 결과 비교
Fig. 9. Comparison of semantic segmentation qualitative results for each technology application

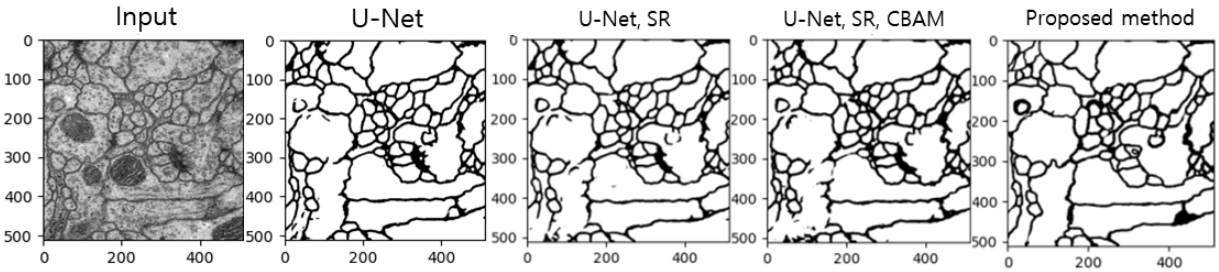

그림 10. 각 기술 결합에 따른 의미분할 정성적 결과 비교
Fig. 10. Comparison of semantic segmentation qualitative results accuracy to the combinations of the core technologies

표 4의 결과를 보면, U-Net을 기준으로 했을 때, 초해상도 기술(SR)을 적용한 경우 IoU와 Dice 계수가 각각 0.9021과 0.9485로 개선되었다. 이는 SR 기술이 세밀한 구조 표현을 개선하여 의미분할 성능을 높이는 데 기여했음을 나타낸다. 또한, CBAM을 적용한 경우에도 손실 값(Loss)이 가장 낮아지는 결과를 보였으나, IoU와 Dice 계수에서 큰 변화는 없었다. 푸리에 변환(Fourier Transform)은 전역적인 정보 보존에 중점을 두었으나, 성능 개선이 제한적이었다.

표 5에서는 이러한 기술들을 결합하여 적용한 경우의 성능 변화를 분석하였다. SR과 CBAM을 결합한 모델은 IoU와 Dice 계수가 각각 0.8998과 0.9473으로 나타났으며, 손실 값은 0.1940으로 줄어들었다. 이는 SR이 세밀한 구조를 보존하고, CBAM이 중요한 정보를 강조하여 전체적인 성능을 향상시킨 것을 의미한다. CBAM과 푸리에 변환을 결합한 경우에는 IoU가 0.9042, Dice 계수가 0.9497로 나타나 가장 우수한 성능을 기록하였다. SR과 푸리에 변환을 결합한 모델 역시 안정적인 성능을 보여주었다. 마지막으로, 제안된 방법은 SR, CBAM, 푸리에 변환을 모두 결합하여 적용한 결과로, IoU와 Dice 계수가 각각 0.9049와 0.9499로 가장 높은 성능을 기록하였다. 이는 세 가지 기술의 강점을 최대한으로 결합함으로써 의미분할의 정확도와 일관성을 극대화할 수 있음을 보여준다. 결론적으로, 본 연구에서는 다양한 기술의 결합이 U-Net 기반 의미분할 성능을 크게 향상시킬 수 있음을 입증하였다. 특히, CBAM과 푸리에 변환의 결합이 IoU와 Dice 계수 모두에서 뛰어난 성능을 보였으며, 제안된 방법은 이러한 기술들을 종합하여 가장 우수한 결과를 도출하였다. 이는 초해상도 기술과 CBAM, 푸리에 변환을 결합한 것이 각 특징 맵의 중요한 부분을 강조하고, 전역적인 주파수 정보를 보존하여 의미분할의 정밀도를 높이는 데 기여했음을 시사한다.

제안하는 기술이 다양한 데이터셋에 대해서도 우수한 성능을 보이는지 확인하기 위하여, 추가적인 데이터셋에 대한 실험을 실시하였다. 표 6과 그림 11은 Broad Bioim- age Benchmark Collection에서 제공하는 세포핵의 이미지(BBBC007)에 대한 의미 분할 결과이다. 이 결과에서 확인할 수 있듯이, 추가 데이터셋에 대해서도, 제안하는 기술들이 기존 방법보다 우수한 성능을 보임을 확인할 수 있었다.

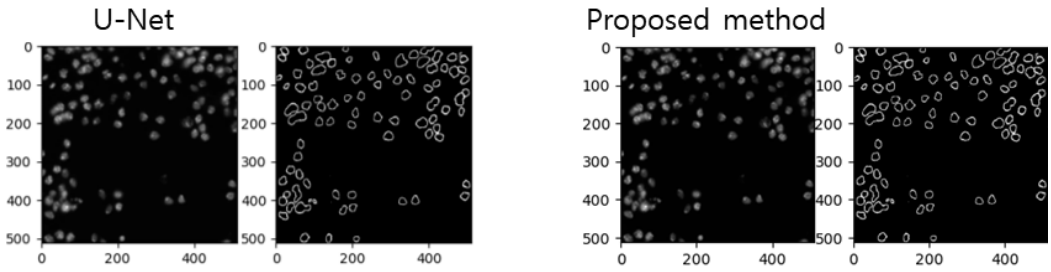

그림 11. BBBC007 데이터셋에 대한 의미분할 정성적 결과 비교
Fig. 11. Comparison of qualitative results of dataset semantic segmentation for BBBC007 dataset

표 6. BBBC007 데이터셋에 대한 의미분할 정확도 분석
Table 6. Analyzing the accuracy of semantic segmentation for BBBC007 dataset

	Loss
U-Net	0.1522
Proposed method	0.1262

표 7은 다양한 네트워크 방식을 사용한 의미분할 정확도를 분석한 결과를 보여주며, 그림 12의 경우 정성적인 결과를 나타내고 있다. FCN, U-Net, U-Net++ 그리고 제안된 방법을 비교하였다.

표 7. 네트워크 방식에 따른 의미분할 정확도 비교
Table 7. Comparison of the accuracy performances of various semantic segmentation algorithms

	Loss	IoU	Dice
FCN	0.2277	0.8933	0.9436
U-Net	0.2186	0.8985	0.9464
U-Net++	0.2105	0.9016	0.9483
Proposed method	0.1895	0.9049	0.9499

FCN은 상대적으로 높은 손실값과 낮은 IoU 및 Dice 계수를 보였다. 이는 FCN이 고해상도 특성을 잘 보존하지 못하여 분할 성능이 다른 네트워크에 비해 떨어질 수 있음을 의미한다. U-Net은 FCN에 비해 낮은 손실값과 높은 IoU 및 Dice 계수를 보였다. 이는 U-Net의 스킵 연결(skip connection)이 분할 성능을 향상시키는데 기여했음을 보여준다. U-Net++는 U-Net보다 더 낮은 손실값과 더 높은 IoU 및 Dice 계수를 기록하였다. 이는 U-Net++의 네스트 된 구조가 더 나은 특징 표현을 가능하게 하여 성능을 향상시켰음을 나타낸다. 제안된 방법은 가장 낮은 손실값과 가장 높은 IoU 및 Dice 계수를 기록하였다. 이는 제안된 방법이 다른 네트워크 방식보다 더 높은 성능을 보여줬음을 의미한다.

그림 12는 다양한 네트워크 방식에 따른 의미분할의 정성적 결과를 보여준다. Input 이미지에 대해 FCN, U-Net, U-Net++ 그리고 제안된 방법을 적용한 결과를 비교하고 있다. 각 네트워크 방식의 출력 이미지를 통해 의미분할의 성능을 시각적으로 평가할 수 있다. 제안된 방법은 다른 네트

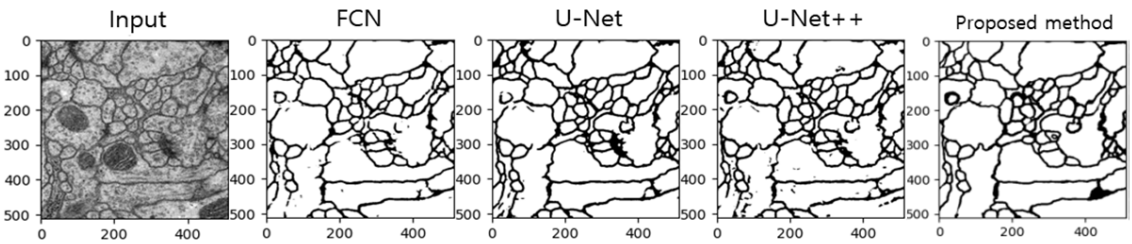

그림 12. 네트워크 방식에 따른 의미분할 정성적 결과 비교
Fig. 12. Comparison of semantic segmentation qualitative results by network method

워크 방식에 비해 더 정확하고 세밀한 의미분할 결과를 제공함을 확인할 수 있다. 특히 제안된 방법은 경계 부분의 세밀한 표현과 객체의 일관된 분할을 잘 수행하는 것으로 나타났다.

Ⅴ. 결 론

본 연구에서는 기존의 U-Net 모델을 개선하여 의미분할 성능을 향상시키기 위한 방법을 제안하였다. 제안된 방법은 다음과 같이 세가지 핵심 요소 기술들이 이었다. 첫째는, 초해상도 기술(Super-Resolution)을 도입했다. 본 논문에서는 ESRGAN 모델을 사용하여 입력 이미지를 초해상도로 변환함으로써 더 고해상도의 입력 데이터를 제공하여 모델의 학습 성능을 향상시켰다. 둘째는, 컨볼루션 블록 어텐션 모듈(CBAM)을 사용했다. 각 인코더 블록에 CBAM을 추가하여 중요한 특징을 강조함으로써, 모델이 더 중요한 영역에 집중할 수 있게 하여 성능을 개선하였다. 셋째는 푸리에 변환을 이용한 전역적 정보 보존하는 과정을 적용했다. 푸리에 변환을 통해 주파수 도메인에서 전역적 정보를 보존함으로써, 모델이 전체적인 이미지 구조를 더 잘 이해할 수 있도록 하였다.

실험 결과, 제안된 방법은 BCEWithLogitsLoss, IoU, Dice 계수 등 다양한 평가 지표에서 기존의 U-Net보다 우수한 성능을 보였다. 특히, 제안된 방법은 더 낮은 손실 값과 더 높은 IoU 및 Dice 계수를 기록하여, 의미분할 작업에서 더 정확하고 일관된 결과를 제공함을 확인할 수 있었다. 이러한 성능 향상은 제안된 기술들이 유효함을 입증하며, 의미분할 분야에서의 실질적인 응용 가능성을 보여준다. 앞으로의 연구에서는 제안된 방법을 다양한 데이터셋과 실제 응용 사례에 적용하여 그 성능을 더욱 검증하고, 추가적인 개선 방안을 모색할 것이다. 또한, 다른 딥러닝 모델과의 비교 연구를 통해 제안된 방법의 장단점을 더욱 명확히 규명할 필요가 있다.

참 고 문 헌 (References)

[1] Y. Liu; C. Liu; K. Han; Q. Tang; Z. Qin, “Boosting Semantic Segmentation from the Perspective of Explicit Class Embeddings,” proceedings of International Conference on Computer Vision, Paris, France, pp. 821-831, 2023.
doi: https://doi.org/10.1109/ICCV51070.2023.00082

[2] G. Lin, A. Milan, C. Shen, I. Reid, “RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation,” Proceedings of Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 1925-1934, 2017.
doi: https://doi.org/10.48550/arXiv.1611.06612

[3] Y. Yang, S. Soatto, “FDA: Fourier Domain Adaptation for Semantic Segmentation,” Proceedings of Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, pp. 4084 – 4094, 2020.
doi: https://doi.org/10.1109/CVPR42600.2020.00414

[4] J. Long, E. Shelhamer, T. Darrell, “Fully convolutional networks for semantic segmentation,” Proceedings of Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp. 3431-3440, 2015.
doi: https://doi.org/10.1109/CVPR.2015.7298965

[5] O. Ronneberger, P. Fischer, T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany, pp. 234-241, 2015.
doi: https://doi.org/10.1007/978-3-319-24574-4_28

[6] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, “UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation,” Proceedings of International Workshop on Deep Learning in Medical Image Analysis, Québec City, Canada, pp. 3-11, 2017.
doi: https://doi.org/10.1007/978-3-030-00889-5_1

[7] D. Khaledyan, A. Amirany, K. Jafari, M.H. Moaiyeri, A. Zargari Khuzani, N. Mashhadi, “Low-Cost Implementation of Bilinear and Bicubic Image Interpolation for Real-Time Image Super-Resolution,” Proceedings of Global Humanitarian Technology Conference (GHTC), Seattle, WA, USA, pp. 1-5, 2020.
doi: https://doi.org/10.1109/GHTC46280.2020.9342625

[8] S. Boukhtache, B. Blaysat, M. Grédiac, F. Berry, “FPGA-based architecture for bi-cubic interpolation: the best trade-off between precision and hardware resource consumption,” Journal of Real-Time Image Processing, pp. 901 – 911, 2021.
doi: https://doi.org/10.1007/s11554-020-01035-1

[9] Y. Shi, K. Wang, C. Chen, L. Xu, L. Lin, “Structure-Preserving Image Super-Resolution via Contextualized Multi-Task Learning,” Proceedings of Transactions on Multimedia, pp. 2804-2815, 2017.
doi: https://doi.org/10.1109/TMM.2017.2711263

[10] C. Dong, C. C.Loy, K. He, X. Tang, “Image Super-Resolution Using Deep Convolutional Networks,” proceedings of Transactions on Pattern Analysis and Machine Intelligence, Shenzhen, China, pp. 295-307, 2015.
doi: https://doi.org/10.1109/TPAMI.2015.2439281

[11] B. Lim, S. Son, H. Kim, S. Nah, K. Lee “Enhanced Deep Residual Networks for Single Image Super-Resolution,” proceedings of Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 1132-1140, 2017.
doi: https://doi.org/10.48550/arXiv.1707.02921

[12] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. Loy, Y. Qiao, X. Tang, “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks,” proceedings of Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, pp. 63-79, 2018.
doi: https://doi.org/10.48550/arXiv.1809.00219

[13] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” proceedings of European Conference on Computer Vision (ECCV), Munich, Germany, pp. 3-19, 2018.
doi: https://doi.org/10.1007/978-3-030-01234-2_1

[14] G. Brauwers, F. Frasincar, “A General Survey on Attention Mechanisms in Deep Learning,” proceedings of IEEE Transactions on Knowledge and Data Engineering, pp. 3279 – 3298, 2021.
doi: https://doi.org/10.48550/arXiv.2203.14263

저 자 소 개

김 민 균

- 2019년 ~ 현재 : 세종대학교 전자정보통신공학과 학사과정
- ORCID : https://orcid.org/0009-0009-8926-3766
- 주관심분야 : 영상처리, 인공지능, 의미분할

한 종 기

- 1992년 : KAIST 전기및전자공학과 공학사
- 1994년 : KAIST 전기및전자공학과 공학석사
- 1999년 : KAIST 전기및전자공학과 공학박사
- 1999년 3월 ~ 2001년 8월 : 삼성전자 DM연구소 책임연구원
- 2001년 9월 ~ 현재 : 세종대학교 전자정보통신공학과 교수
- 2008년 9월 ~ 2009년 8월 : University California San Diego (UCSD) Visiting Scholar
- ORCID : https://orcid.org/0000-0002-5036-7199
- 주관심분야 : 비디오 코덱, 영상 신호처리, 정보 압축, 방송 시스템