

## Barlow Twins 기반 VAE를 활용한 제한된 레이블 환경의 의료영상 분할

박민영<sup>1</sup> · 박채현<sup>1</sup> · 류현승<sup>2</sup> · 박수형<sup>3\*</sup><sup>1</sup>전남대학교 컴퓨터정보통신공학과 학사과정<sup>2</sup>전남대학교 지능전자컴퓨터공학과 석사과정<sup>3</sup>전남대학교 전자컴퓨터공학부 교수

## Medical Image Segmentation in Limited-Label Environments Using a Barlow Twins-Based VAE

Minyeong Park<sup>1</sup> · Chaehoon Park<sup>1</sup> · Hyunseung Ryu<sup>2</sup> · Suhung Park<sup>3\*</sup><sup>1</sup>Bachelor's Course, Department of Computer Engineering, Chonnam National University, Gwangju 61186, Korea<sup>2</sup>Master's Course, Department of Intelligent Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea<sup>3</sup>Associate Professor, Department of Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea

### [요약]

의료영상 분할(segmentation)은 대규모의 정밀한 레이블 데이터를 요구하지만, 전문가 레이블링에 따른 높은 비용과 임상 데이터 내 병변이 존재하지 않는 음성(negative) 샘플의 포함으로 인해 기존 지도학습 기반 모델은 성능 저하의 한계를 가진다. 특히 U-Net과 같은 결정론적 구조는 negative 샘플에 대해서도 특정 패턴의 출력을 생성하는 경향이 있어, 불균형 데이터 환경에서 학습 불안정성 및 과적합이 발생할 수 있다. 이를 해결하기 위해 본 연구에서는 자기지도 학습 기법인 Barlow Twins를 확률적 생성 모델인 VAE(Variational Autoencoder)에 결합한 이단계 학습 프레임워크 BT-VAE를 제안한다. 사전학습 단계에서는 레이블 없이 회전 기반 증강 뷰를 활용하여 VAE 인코더가 방향 불변성과 채널 간 중복성이 감소된 잠재 표현을 학습하도록 한다. 미세조정 단계에서는 사전학습된 인코더를 기반으로 제한된 레이블 데이터만을 사용하여 분할 마스크를 예측한다. 또한 Barlow Twins 기반 사전학습은 채널 간 중복성을 감소시켜 일반화 가능한 특징 학습을 유도하며, VAE의 확률적 잠재 공간은 음성 샘플의 불확실성을 반영하여 오검출(false positive)을 감소시킨다. COVID-19 CT 데이터셋과 상완신경총(Brachial Plexus, BP) 초음파 데이터셋에 대한 실험 결과, 제안한 방법은 비교 모델 대비 우수한 분할 성능을 보였으며, 특히 음성 샘플이 포함된 환경에서 안정적인 성능을 나타냈다.

### [Abstract]

Accurate medical image segmentation requires large-scale annotated data; however, the high cost of expert labeling and presence of negative samples (i.e., images without lesions) in clinical datasets pose challenges for conventional supervised approaches. In particular, deterministic architectures, such as U-Net, tend to generate pattern-specific outputs even for negative inputs, leading to training instability and overfitting in imbalanced settings. To address this, we propose BT-VAE, a two-stage framework that integrates the self-supervised Barlow Twins method into a Variational Autoencoder (VAE). In the pre-training stage, the VAE encoder learns orientation-invariant and redundancy-reduced latent representations using rotation-based augmented views without labels. In the fine-tuning stage, the pretrained encoder predicts segmentation masks using limited labeled data. Barlow Twins-based pretraining reduces channel-wise redundancy, enabling generalizable feature learning, while the probabilistic latent space of VAE captures uncertainty in negative samples, reducing false positives. Experiments on the COVID-19 CT and Brachial Plexus ultrasound datasets showed that the proposed method achieves superior segmentation performance compared to baseline models, particularly in the presence of negative data.

**색인어** : 의료영상, 딥러닝, 영상 분할, 유넷(U-Net), 변분 오토인코더(VAE)**Keyword** : Medical Imaging, Deep Learning, Image Segmentation, U-Net, Variational Autoencoder<http://dx.doi.org/10.9728/dcs.2026.27.5.1391>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 13 March 2026; Revised 17 April 2026

Accepted 11 May 2026

\*Corresponding Author; Suhung Park

Tel : +82-62-530-1797

E-mail: suhyung@jnu.ac.kr

## I. 서론

의료영상은 질병의 진단과 병변의 위치 확인, 치료 계획 수립, 수술 및 시술 가이드 등 다양한 임상 과정에서 핵심적인 역할을 수행한다. 최근에는 의료영상 분석 분야에서 딥러닝 기반 기법이 널리 활용되며, 질병 검출, 분류, 분할 등 다양한 과제를 효과적으로 수행하고 있다. 특히 의료영상 분할은 장기 및 병변(종양, 신경, 혈관 등)을 픽셀 단위로 구분함으로써 정량적 분석을 가능하게 하며, 자동화된 진단 보조 및 임상 의사결정 지원 시스템의 기반 기술로 활용되고 있다[1]. 그러나 의료영상 분할은 영상 획득 장비 및 촬영 조건에 따른 도메인 차이, 환자 간 해부학적 구조의 변동성, 병변 형태의 불규칙성 등으로 인해 높은 난이도를 가지며, 안정적인 성능을 확보하기 위해서는 대규모의 정밀한 레이블 데이터가 요구되는 경우가 많다.

대부분의 딥러닝 기반 분할 모델은 지도학습(supervised learning)에 기반하여 학습되며, 이에 따라 충분한 양의 레이블 데이터가 성능을 좌우하는 핵심 요소로 작용한다. 하지만 의료영상에서의 레이블링은 전문 지식이 요구되는 작업으로, 임상 의의 시간과 인력 등 상당한 자원이 필요하며 데이터 구축 비용 또한 높다[2]. 이러한 한계는 제한된 레이블 환경에서의 학습 성능 저하로 이어질 수 있으며, 의료영상 분할 분야에서 효과적인 학습 전략의 필요성을 더욱 부각시킨다[3].

이러한 문제를 완화하기 위한 대안으로, 레이블 없이도 일반화 가능한 표현을 학습할 수 있는 비지도 사전학습(self-supervised pre-training) 기법들이 주목받고 있다[4]. 그중 Barlow Twins는 동일한 입력으로부터 생성된 서로 다른 변형에서도 일관되게 유지되는 표현을 학습하고, 채널 간 특징 표현의 상관성을 최소화하여 채널별로 중복되지 않는 특징을 학습하도록 함으로써 강건한 표현 학습을 수행하는 방법이다[5]. 이러한 특성은 제한된 레이블 데이터 환경에서도 성능 향상을 기대할 수 있다는 점에서, 의료영상 분할 분야에 적용 가능성이 크다.

이와 함께, 의료영상 분할 모델 구조 측면에서도 다양한 접근이 제안되고 있다. 의료영상 분할에서는 U-Net과 같은 CNN 기반 구조부터 Transformer 기반 모델에 이르기까지 다양한 딥러닝 모델들이 제안되어 왔다[6]-[10]. 특히, 최근에는 Transformer와 CNN의 특성을 동시에 활용하는 MIST와 같은 hybrid 기반의 구조의 최신 분할 모델이 제안되었으며, 다양한 의료영상 데이터셋에서 높은 성능을 보이고 있다[11]. 또한 최근에는 대규모 사전학습 기반 범용 분할 모델과 반지도 및 자기지도 학습을 활용하여 제한된 레이블 환경에서의 성능을 향상시키려는 연구가 활발히 진행되고 있다[12]-[14]. 이러한 모델들은 복잡한 구조와 대규모 학습을 통해 우수한 성능을 달성하고 있으나, 제한된 레이블 환경이나 negative 데이터가 포함된 상황에서의 안정성 측면에서는 여전히 한계가 존재한다.

이러한 한계는 특히 결정론적(deterministic) 구조에서 두드러지게 나타난다. 대표적으로 U-Net은 의료영상 분할에서 가장 널리 사용되는 구조 중 하나로, 다양한 데이터셋과 환경에서 우수한 성능을 보여왔다. 또한 기존 연구에서는 Barlow Twins 기법을 U-Net 기반 구조에 적용한 BT U-Net과 같은 방식이 제안되며, 제한된 레이블 환경에서의 표현 학습 성능을 개선하고자 하였다[15]. 그러나 U-Net은 기본적으로 입력에 대해 결정론적 예측을 수행하는 구조로, 동일 입력에 대해 하나의 고정된 결과로 수렴하는 경향이 있다. 특히 임상 환경에서 수집되는 데이터는 병변이 존재하는 양성(positive) 샘플뿐만 아니라 병변이 존재하지 않는 음성(negative) 샘플도 함께 포함되는데, U-Net과 같은 결정론적 구조는 negative 샘플에 대해서도 특정 패턴의 출력을 생성하려는 경향을 보인다. 이러한 특성은 negative 데이터가 포함된 학습 환경에서 학습 불안정성을 유발하거나 특정 패턴에 과적합되는 현상으로 이어져 성능 저하가 발생할 수 있다는 한계를 가진다.

반면, 생성형 모델의 대표적인 구조 중 하나인, VAE(Variational Autoencoder)는 잠재공간(latent space)에서 데이터 분포를 생성하는 확률 모델로서, 단일한 결정론적 예측에만 의존하지 않고 다양한 형태의 출력을 생성할 수 있는 특성을 가진다. 이러한 생성 모델의 특성은 Negative 데이터가 포함된 상황에서도, 보다 안정적인 표현 학습을 가능하게 하며, 결과적으로 의료영상 분할 성능의 강건성 측면에서 U-Net 대비 장점으로 작용할 수 있다.

본 연구에서는 결정론적 분할 구조의 한계를 극복하기 위해, Barlow Twins 기반의 자기지도 학습 전략을 확률적 생성 모델인 VAE(Variational Autoencoder)에 통합한 BT-VAE를 제안한다. VAE의 확률적 특성을 활용함으로써, 음성(negative) 샘플이 포함된 학습 환경에서도 보다 안정적이고 강건한 표현 학습이 가능하다. 또한 Barlow Twins 목적 함수를 통해 잠재 특징 간의 중복성을 억제함으로써, 제한된 레이블 환경에서도 모델의 일반화 성능을 향상시키고자 한다. 제안 방법의 유효성을 검증하기 위해, 다음의 두 가지 실험을 수행하고, 1) 레이블 데이터 비율 변화에 따른 모델 성능을 기존 방법과 비교 분석, 2) 양성 및 음성 샘플이 혼재된 데이터셋에서의 강건성을 평가, 각 실험의 정량적 결과와 함께 제안된 BT-VAE의 우수한 성능을 제시한다.

## II. 연구 방법

### 2-1 학습 프레임워크 구조

본 연구에서 제안하는 BT-VAE는 자기지도 기반 사전학습(self-supervised pre-training)과 생성적 지도학습(generative supervised learning)을 순차적으로 결합한 이

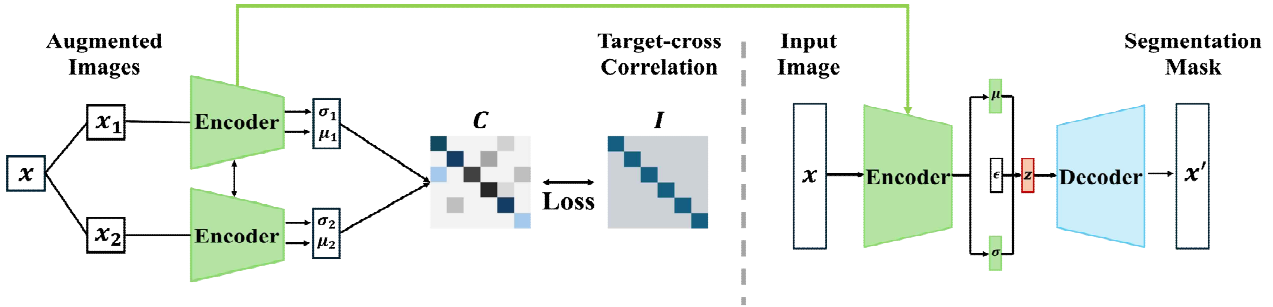


그림 1. 제안된 BT-VAE 전체 학습 프레임워크. Barlow Twins 기반 사전학습과 VAE 학습을 결합한 전체 구조  
**Fig. 1.** Overall training framework of the proposed BT-VAE. The overall architecture combining Barlow Twins-based pretraining and VAE-based learning

단계(two-stage) 학습 프레임워크로 구성된다. 1단계 사전 학습에서는 레이블 정보를 사용하지 않고, 동일 입력 영상으로부터 생성된 두 개의 증강 뷰(augmented view)를 VAE 인코더에 통과시켜 잠재 분포 파라미터를 추정하고, Barlow Twins 손실 함수를 통해 채널 간 중복성이 억제된 강건한 표현을 학습한다. 이를 통해 인코더는 레이블 데이터 없이도 의미적으로 일관되고 독립적인 잠재 표현을 획득한다. 2단계 미세조정에서는 사전학습된 인코더 가중치를 초깃값으로 활용하여, 소량의 레이블 데이터를 기반으로 VAE 전체 구조(인코더-잠재공간 샘플링-디코더)를 학습하고 분할 마스크를 예측한다. 본 프레임워크는 비지도 표현 학습의 일반화 능력과 생성 모델 기반의 확률적 추론을 결합함으로써, 제한된 레이블 환경 및 음성(negative) 샘플이 포함된 불균형 데이터 환경 모두에서 안정적인 분할 성능을 달성하도록 설계되었다(그림 1).

## 2-2 Barlow Twins 기반 사전학습

### 1) Barlow Twins 구조

본 연구의 사전학습 단계는 Barlow Twins의 핵심 원리를 VAE의 확률적 잠재 표현 구조에 맞게 재구성한 분포 기반 자기지도 표현 학습 방식을 따른다. 기존 Barlow Twins가 결정론적 특징 벡터를 대상으로 교차 상관 행렬을 단위행렬에 수렴시키는 것과 달리, 본 구조에서는 잠재 분포의 파라미터(평균 벡터, 표준편차 벡터)를 표현 학습의 대상으로 활용한다. 구체적으로, 1) 원본 입력 영상에 대해 좌측 90도 회전과 우측 90도 회전을 각각 적용하여 두 개의 증강 뷰를 생성, 2) 회전 기반 증강은 동일한 해부학적 구조를 서로 다른 방향에서 관찰하게 함으로써, 모델이 방향 불변(orientation-invariant)하고 일반화 가능한 표현을 학습하도록 유도, 3) 생성된 두 뷰는 가중치를 공유하는(weight-sharing) 동일한 VAE 인코더에 입력되어 각각의 잠재 분포 파라미터를 추정한다. 이 사전학습 단계에서는 재매개변수화(reparameterization)를 통한 잠재 벡터 샘플링을 수행하지 않고, 분포 파라미터 자체를 표현 벡터로 직접 활용함으로써 데이터의 내재적 불확실성과 변동성이 표현에 반영되도록 한다. 이후 두 인코더의 잠재 표

현을 바탕으로 채널 간 상호 상관 행렬(cross-correlation matrix)를 산출하고, 상관 행렬을 단위 행렬에 수렴시키는 방향으로 인코더를 최적화한다.

### 2) Barlow Twins Loss

BT-VAE의 사전학습 손실 함수  $L_{bt}$  는 잠재 분포 정렬을 위한 KL 기반 항  $L_{kl}$  과 평균 벡터 기반 정규화  $L_{\mu}$  를 결합한 형태로 정의된다:

$$L_{bt} = L_{kl} + L_{\mu} \tag{1}$$

KL 기반 항은 두 잠재 분포 간의 KL 발산을 기반으로 채널 간 관계를 학습하는 항으로, 동일 채널 쌍에 대한 KL 값은 0에 가깝게 최소화하고, 서로 다른 채널 쌍에 대한 KL 값은 임계값에 수렴하도록 유도한다. 이는 동일 채널 내에서는 두 뷰의 잠재 분포가 일관된 의미 표현을 공유하도록 하고, 서로 다른 채널 간에는 중복성과 상관성을 억제하는 Barlow Twins의 핵심 목적을 확률적 잠재 공간에서 구현한 것이다.

$$L_{kl} = \sum_i C_{kl_i}^2 + \sum_i \sum_{j \neq i} (C_{kl_{ij}} - \tau)^2 \tag{2}$$

평균 벡터 기반 항은 평균 벡터 기반의 보조 정규화 항으로, 표준적인 Barlow Twins 손실 형태를 따르며 사전학습 과정에서 잠재 표현이 특정 방향으로 과도하게 편향되거나 불안정해지는 현상을 완화한다.

$$L_{\mu} = \sum_i (1 - C_{\mu_{ii}})^2 + \lambda \sum_i \sum_{j \neq i} C_{\mu_{ij}}^2 \tag{3}$$

$C_{kl_{ij}}$  는 두 증강 뷰로부터 얻어진 잠재 표현 간의 KL divergence 값을 나타내며,  $i$ 와  $j$ 는 각각 채널 인덱스를 나타낸다.  $\tau$ 는 서로 다른 채널 간 KL divergence의 목표 값으로, 채널 간 중복 표현을 억제하기 위해 사용된다. 또한  $C_{\mu_{ij}}$  는 평균 벡터 기반 잠재 표현으로부터 계산된 채널 간 상관 행렬을 의미한다. 여기서  $\lambda$ 는 대각 항과 비대각 항 사이의 상대적 기

여도를 조절하는 하이퍼파라미터이고, 두 손실 항의 결합을 통해 잠재 표현은 부 간 의미적 일관성과 채널 간 독립성을 동시에 확보하며, 이는 제한된 레이블 환경에서의 강력한 분할 성능의 기반이 된다.

### 2-3 Variational Autoencoder(VAE)

#### 1) VAE 구조

본 연구에서는 의료영상 분할을 위한 생성 모델로 VAE를 채택한다. VAE는 인코더-잠재공간-디코더로 구성된 확률적 생성 프레임워크로, 결정론적 표현에 의존하는 일반 오토인코더 및 U-Net과 달리, 입력 데이터의 변동성과 불확실성을 잠재 공간의 확률 분포로 명시적으로 모델링한다. 이를 위해, 제한된 구조에서 입력 영상은 인코더를 통과하며, 인코더는 잠재 변수의 분포를 나타내는 평균 벡터와 표준편차 벡터를 추정하고, 이를 바탕으로 재매개변수화 기법을 통해 잠재 벡터를 샘플링하고, 이를 디코더에 입력하여 최종적으로 분할 마스크를 생성한다. 이러한 특성은 병변이 존재하지 않는 음성(negative) 샘플이 포함된 학습 환경에서도 특정 패턴으로의 과적합을 억제하고, 다양한 입력 변형에 대해 유연한 표현 학습이 가능하게 하는 핵심 요인이다.

#### 2) VAE Loss

VAE 학습은 분할 마스크 예측의 정확도를 높이기 위한 손실과, 잠재 공간에서의 분포 학습을 안정화하기 위한 정규화 손실을 함께 고려하여 수행된다. 본 연구에서는 분할 성능을 효과적으로 반영하기 위해 Dice Loss와 BCE를 결합한 손실 함수를 사용하였다.

$$L_{vae} = \alpha * L_{Dice} + (1 - \alpha) * L_{BCE} \tag{4}$$

가중치  $\alpha$ 는 두 손실 항의 상대적 기여도를 조절하는 하이퍼파라미터로, 두 손실의 상보적 특성을 결합함으로써 전반적인 분할 정확도와 경계 예측 성능을 동시에 향상시키도록 구성하였다. Dice Loss는 예측 마스크와 정답 마스크 간의 중첩 정도를 직접적으로 반영하여 관심 영역이 희소한 불균형 데이터에서 안정적인 학습을 유도한다[16].

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N \hat{y}_i y_i}{\sum_{i=1}^N \hat{y}_i + \sum_{i=1}^N y_i} \tag{5}$$

BCE(Binary Cross Entropy)는 픽셀 단위의 예측 확률과 정답 레이블 간의 차이를 측정하여 경계 영역을 포함한 세밀한 분할 예측을 가능하게 한다[17].

$$L_{BCE} = -\frac{1}{N} \sum_{i=0}^N y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i) \tag{6}$$

또한 VAE의 잠재 분포 학습을 안정화하기 위한 KL 정규화는 근사 사후 분포와 표준 정규 사전 분포  $p(z)=N(0,1)$  간의 KL 발산을 최소화하는 방식으로 수행되며, 이는 잠재 공간의 구조적 연속성을 보장하고 과적합을 방지하는 역할을 한다.

### 2-3 실험 환경 및 평가 지표

#### 1) 데이터셋

##### • COVID-19 CT Segmentation Dataset(COVID Dataset)

COVID Dataset은 폐 CT 이미지와 코로나 병변 영역에 대한 레이블로 구성되어 있으며, 병변이 존재하는 픽셀은 1, 병변이 없는 픽셀은 0으로 표시된 이진 마스크를 사용한다 [18]-[20]. 전체 데이터는 총 2,729쌍의 이미지-마스크로 이루어져 있다. 학습을 위해 데이터셋을 8:1:1 비율로 train, validation, test로 분할하였으며, 각각 2,267장, 231장, 231장을 사용하였다.

##### • Ultrasound Nerve Segmentation Dataset(BP Dataset)

BP Dataset은 상완 신경총 초음파 이미지와 마취가 필요한 영역에 대한 레이블로 구성되어 있으며, 마취가 필요한 픽셀은 1, 불필요한 픽셀은 0으로 표시된 이진 마스크를 사용한다. 전체 데이터는 총 5,635쌍의 이미지-마스크로 이루어져 있다[21]. 또한 Covid Dataset과 달리 병변(또는 관심 영역)이 존재하지 않는 Negative 데이터(마스크가 전부 0인 샘플)도 포함되어 있어, Negative 환경에서의 분할 성능을 평가할 수 있다. 학습을 위해 데이터셋을 8:1:1 비율로 train, validation, test로 분할하였으며, 각각 4,508장, 563장, 564장을 사용하였다.

#### 2) 학습 설정 및 구현 세부사항

본 연구에서는 사전학습(pre-training) 단계와 VAE 기반 지도학습 단계를 구분하여 학습을 수행하였다. optimizer는 사전학습과 지도학습 단계에서 동일하게 AdamW optimizer를 사용하였다. 사전학습 단계에서의 학습률은  $3 \times 10^{-4}$ , weight decay는  $1 \times 10^{-4}$ 로 설정하였고 배치 크기는 16, 총 학습 에폭은 200으로 설정하였다. Barlow Twins 기반 손실 함수의 가중치  $\lambda$ 는 0.2로 설정하였다. 이후 VAE 기반 지도학습 단계에서의 학습률은  $1 \times 10^{-4}$ 로 설정하였고, 배치 크기는 16, 총 학습 에폭은 400으로 설정하였으며, 잠재 공간은  $1024 \times 8 \times 8$  크기의 feature map 형태로 구성되며, 손실 함수의 가중치  $\alpha$ 를 0.99로 설정하였다.

### 3) 평가지표

제안한 모델의 성능을 정량적으로 평가하기 위해 네 가지 보완적 지표를 사용하며, 이 지표들은 영상 분할의 성능을 다양한 측면에서 효과적으로 평가한다.

#### • Dice Score

Dice Score 는 예측 마스크와 정답 마스크 간의 겹치는 정도를 측정하는 지표로, 분할 성능을 대표적으로 평가하는 데 사용된다. 특히 의료영상 분할처럼 배경 대비 관심 영역의 비율이 작은 불균형 데이터에서 안정적으로 성능을 비교할 수 있다는 장점이 있다.

#### • IoU (Intersection over Union)

IoU는 예측 영역과 정답 영역의 교집합을 합집합으로 나눈 값으로, 두 마스크의 일치도를 평가하는 지표이다. Dice Score와 유사하게 분할 성능을 측정하지만, 보다 엄격하게 겹침 정도를 반영하는 특성이 있다.

#### • Precision

Precision은 모델이 관심 영역으로 예측한 픽셀 중 정답인 픽셀의 비율을 의미한다. 즉, False Positive를 얼마나 줄였는지를 평가하는 지표로, 불필요한 영역을 병변 또는 관심 영역으로 잘못 검출하는 정도를 확인할 수 있다.

#### • Recall

Recall은 실제 정답 영역 중 모델이 올바르게 검출한 픽셀의 비율을 의미한다. 즉, False Negative를 얼마나 줄였는지를 평가하는 지표로, 실제 관심 영역을 놓치지 않고 검출하는 능력을 확인할 수 있다.

### 4) 실험 설계

#### • 실험 1: 레이블 데이터 비율 변화에 따른 모델 성능 비교

실험 1에서는 제한된 레이블 데이터 환경에서 제안 모델의 성능을 비교하기 위해, BT-VAE, VAE, BT-U-Net, U-Net 과 함께 최신 segmentation 모델인 MIST를 포함하여 레이블 데이터 비율에 따른 분할 성능 변화를 평가하였다. COVID Dataset을 사용하여 레이블 데이터 비율을 90%에서 10%까지 20% 간격으로 감소시키며 모델을 학습하였고, 지표를 통해 정량적으로 성능을 비교하였다. 또한 레이블 데이터 비율 변화에 따른 예측 변화를 정성적으로 분석하기 위해, 입력 데이터의 분할 결과를 제시하여 레이블 수 감소 상황에서의 예측 안정성과 분할 품질을 비교하였다.

#### • 실험 2: 양성 및 음성 샘플을 포함한 데이터셋에서의 모델 성능 분석

실험 2에서는 양성 및 음성 샘플이 혼재된 보다 복잡한 환경에서 모델 간 성능 차이가 어떻게 나타나는지를 분석하고

자 하였다. 이를 위해 BP Dataset에서 레이블 데이터 비율을 제한한 조건 중, 모델 간 성능 차이가 비교적 뚜렷하게 관찰되는 비율을 대표 조건으로 설정하여 실험을 수행하였다. 해당 조건에서 BT-VAE, VAE, BT-U-Net, U-Net, MIST를 학습하고, 정량적 평가 지표를 통해 분할 성능을 측정하였으며, 각 모델의 결과를 시각적으로 비교하였다. 이를 통해 모델별 예측 마스크의 품질, 경계 표현, 그리고 양성 및 음성 샘플에 대한 안정성을 종합적으로 분석하였다.

### III. 실험 및 결과

#### 3-1 실험 1: 레이블 데이터 비율 변화에 따른 모델 성능 비교

##### 1) 정량적 결과

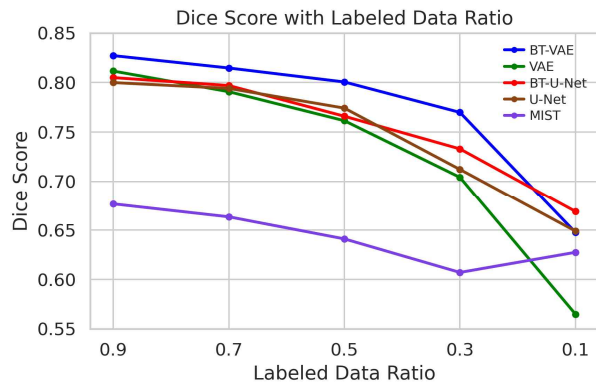


그림 2. COVID Dataset에서의 모델별 레이블 데이터 비율에 따른 Dice Score의 변화

Fig. 2. Changes in dice score according to the labeled data ratio for each model on the COVID dataset

그림 2에서는 레이블 데이터 비율이 감소함에 따라 전반적으로 모든 모델의 성능이 하락하는 경향을 보였으나, 제안한 BT-VAE는 전 구간에서 가장 안정적인 성능을 유지하며 다른 모델 대비 우수한 결과를 나타냈다. 특히 레이블 데이터 비율이 30%인 조건에서 BT-VAE, VAE, BT-U-Net, U-Net, MIST의 Dice Score는 각각 0.7699, 0.7042, 0.7330, 0.7122, 0.6071로 측정되었으며, 이 구간에서 제안 모델이 비교 모델 대비 가장 큰 성능 차이를 보였다. 이는 제한된 레이블 데이터 환경에서도 BT-VAE가 강력한 분할 성능을 유지할 수 있음을 보여준다.

특히, CNN-Transformer 기반의 MIST모델은, 본 실험에서 레이블 데이터 비율이 증가함에도 불구하고 성능 향상이 제한적인 경향을 보였다. 특히 중간 이상의 레이블 비율(50% 이상)에서도 성능 증가 폭이 크지 않거나 일부 구간에서는 성능이 정체되는 현상이 관찰되었다. 이는, Transformer 기반 모델이 일반적으로 대규모 데이터와 충분한 학습 조건에서 높은 성능을 발휘하는 반면, 본 연구와 같이 제한된 레이블

환경에서는 학습 안정성 및 표현 학습 측면에서 한계를 가질 수 있음을 시사한다. 반면, 제안한 BT-VAE는 자기지도 사전학습을 통해 데이터 효율적인 표현을 학습함으로써, 레이블 데이터가 제한된 조건에서도 보다 안정적인 성능을 유지하는 것으로 나타났다.

**2) 정성적 결과**

그림 3은 COVID Dataset에서 각 모델의 분할 결과를 Ground Truth와 오버레이하여 비교한 정성적 분석 결과를 나타낸다. 오버레이 이미지에서 흰색 영역은 Ground Truth를 의미하며, 녹색 영역은 Ground Truth와 모델 예측이 일치하는 정확한 예측 영역을 나타낸다. 반면, 빨간 영역은 Ground Truth에는 존재하지 않지만 모델이 병변으로 잘못 예측한 영역을 나타낸다.

레이블 데이터 비율이 감소함에 따라 비교 모델들은 녹색 영역이 감소하는 경향을 보였으나, 제안한 BT-VAE는 레이블 데이터 수가 줄어든 환경에서도 상대적으로 넓은 녹색 영역을 유지하였다. 이는 제한된 레이블 데이터 조건에서도 제안 모델이 병변 영역을 보다 정확하게 분할하며, 다른 모델 대비 강건한 분할 성능을 보임을 확인할 수 있음을 의미한다.

**3-2 실험 2: 양성 및 음성 샘플을 포함한 데이터셋에서의 모델 성능 분석**

**1) 정량적 결과**

그림 4는 레이블 데이터 비율을 30%로 제한하여 학습한 경우, 각 모델(U-Net, BT-U-Net, MIST, VAE, BT-VAE)의 BP Dataset에 대한 분할 성능을 Dice Score, mIoU, Precision, Recall 지표로 비교한 결과를 나타낸다. 실험 결과, 제안한 BT-VAE는 Dice Score, mIoU, Precision에서 가장 높은 성능을 보였으며, Recall에서는 VAE와 유사한 수준의 성능을 나타냈다. 특히 Precision 지표에서 BT-VAE가 다른 모델 대비 현저히 높은 값을 기록하여, 음성 샘플에 대한 오출을 효과적으로 억제하면서도 병변 영역을 정확하게 분할할 수 있음을 확인하였다. 반면 MIST는 모든 지표에서 전반적으로 낮은 성능을 보였으며, 특히 양성 및 음성 샘플이 함께 포함된 환경에서 예측 결과가 다소 불안정한 경향을 나타냈다. 이는 Transformer 기반 모델이 이러한 복잡한 데이터 환경에서 안정적인 성능을 확보하기 위해서는 보다 충분한 데이터나 추가적인 학습 전략이 필요함을 보여준다.

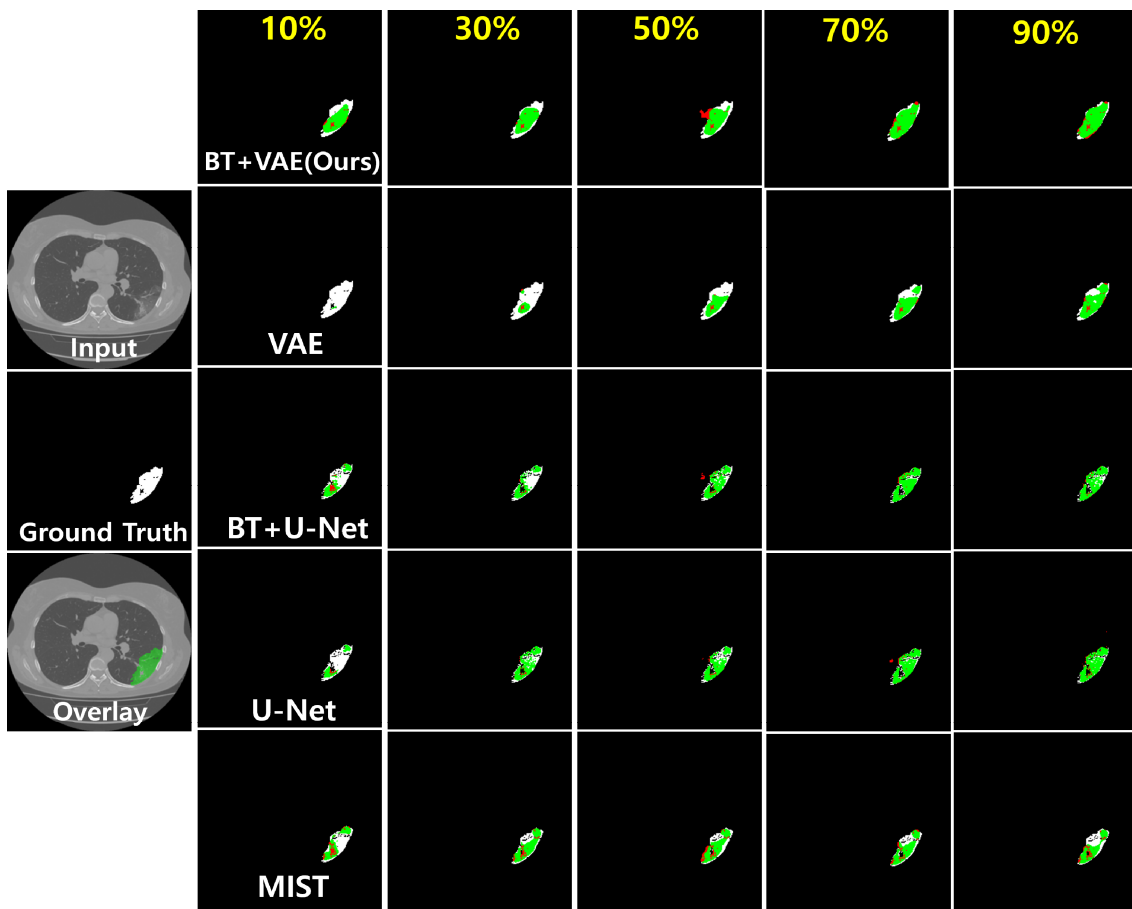


그림 3. COVID Dataset에서 레이블 데이터 비율에 따른 모델별 분할 결과 변화

Fig. 3. Changes in segmentation results for each model according to labeled data ratio on COVID dataset

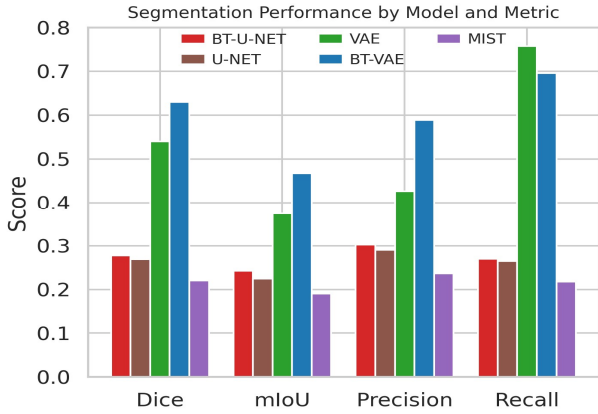


그림 4. BP Dataset에서 레이블 30%로 학습한 경우의 세그멘테이션 성능 지표 비교

Fig. 4. Comparison of segmentation performance metrics on the BP dataset with 30% labeled data

2) 정성적 결과

그림 5는 레이블 데이터 30% 조건에서 학습된 각 모델 (BT-VAE, VAE, BT-U-Net, U-Net, MIST)의 분할 결과를 BP Dataset에 대해 정성적으로 비교한 결과를 나타낸다. BP Dataset은 Negative 데이터가 포함되어 있어 보다 어려운 조건을 제공하며, 이러한 환경에서 모델 간 성능 차이가 뚜렷하게 나타난다. 비교 모델들은 Negative 데이터에서 불필요한 영역을 관심 영역으로 예측하는 오탐지(False Positive)가 증가하거나 예측 결과가 불안정해지는 경향을 보였다. 반면 제안한 BT-VAE는 Negative 데이터가 포함된 상황에서도 오탐지를 상대적으로 억제하면서 관심 영역을 안정적으로 검출하는 경향을 보였다.

IV. 결론

본 연구에서는 Barlow Twins 기법을 생성형 모델인 VAE에 적용한 새로운 학습 프레임워크를 제안하였다. 의료영상 데이터는 특성상 정밀한 레이블링에 많은 전문 인력과 시간이 요구되며, 충분한 레이블 데이터를 확보하는 데 한계가 존재한다. 이러한 문제로 인해 제한된 레이블 데이터 환경에서도 효과적인 학습이 가능한 방법에 대한 필요성이 지속적으로 제기되어 왔다.

본 연구는 이러한 한계를 해결하기 위해, 자기지도 학습 기법을 생성형 모델에 결합함으로써 적은 양의 레이블 데이터 만으로도 안정적인 분할 성능을 달성할 수 있음을 보였다. 점에 의의를 가진다.

제안된 프레임워크의 가장 큰 특징은 기존에 결정론적 모델에 주로 적용되어 온 Barlow Twins 기법을 생성형 모델의 특성에 맞게 변형하여 적용하였다는 점이다. 이를 통해 확률적 잠재 표현을 학습하는 VAE 구조에서도 표현 간 중복을 효과적으로 억제하고, 서로 독립적인 의미 정보를 학습할 수 있도록 하였다. 실험 결과, 제안한 BT-VAE는 기존의 Barlow Twins 기반 U-Net을 포함한 비교 모델 대비 전반적으로 우수한 성능을 보였으며, 특히 Negative 데이터가 포함된 환경과 레이블 데이터가 제한된 조건에서 더욱 강건한 분할 성능을 나타냈다.

그러나 본 연구는 몇 가지 한계점을 가진다. 첫째, 제안된 방법은 실험에 사용된 데이터셋 규모 및 특성에 일정 부분 의존하는 경향을 보인다. 특히 레이블 데이터 비율이 감소함에 따라 모든 모델의 성능이 저하되는 경향이 관찰되었으며, 이는 제안 모델 역시 데이터 규모가 극도로 제한된 환경에서는

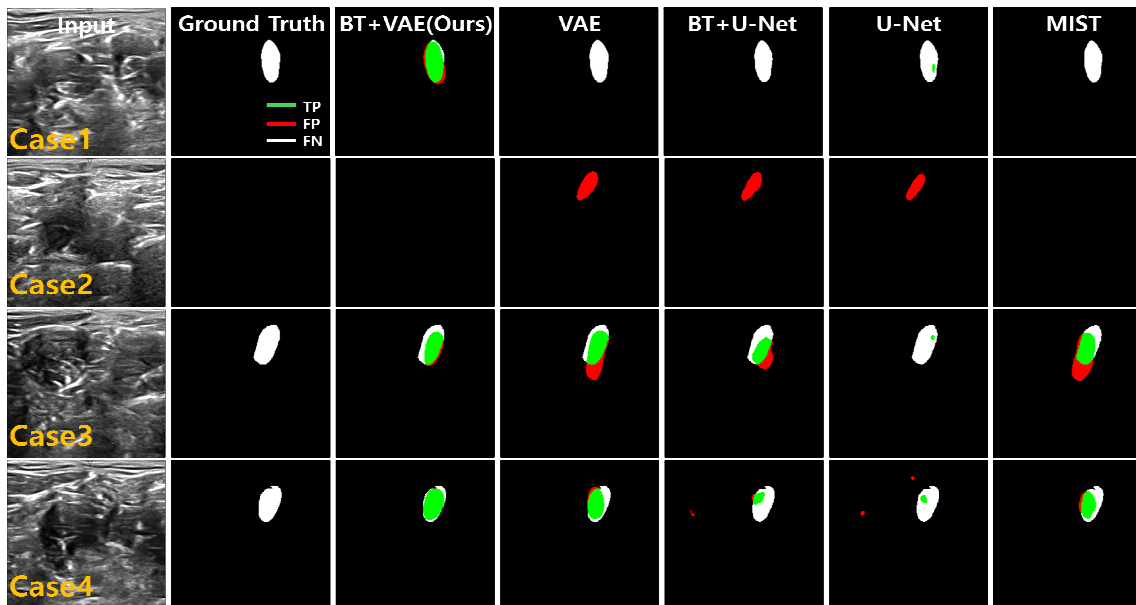


그림 5. BP Dataset에서 레이블 데이터 30% 조건으로 학습한 모델별 분할 결과 비교

Fig. 5. Comparison of segmentation results for each model trained with 30% labeled data on BP dataset

성능 한계를 가질 수 있음을 시사한다. 둘째, 실험은 두 개의 데이터셋(COVID CT, BP Ultrasound)에 한정되어 수행되었기 때문에, 다양한 의료영상 모달리티(예: MRI, X-ray) 및 복잡한 다중 클래스 분할 환경에 대한 일반화 성능은 충분히 검증되지 않았다. 셋째, VAE 기반 구조의 특성상 생성되는 분할 결과가 상대적으로 부드러운 경계를 가지는 경향이 있어, 정밀한 경계 정보가 중요한 과제에서는 추가적인 보완이 필요할 수 있다. 향후 연구에서는 이러한 한계점을 보완하기 위해, 다양한 의료영상 데이터셋과 모달리티에 대한 확장 실험을 수행하고, 데이터 규모 변화에 따른 성능 안정성을 보다 체계적으로 분석할 필요가 있다. 또한 사전학습 단계에서의 증강 방식 다양화 및 잠재 공간 구조 개선을 통해, 보다 일반화된 표현 학습이 가능하도록 확장할 계획이다. 이러한 연구는 실제 임상 환경에서 레이블 데이터 확보의 어려움을 완화하고, 의료영상 분석 시스템의 실용성을 향상시키는 데 기여할 수 있을 것으로 기대된다.

### 감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화 혁신인재양성사업의 지원(IITP-2026-RS-2022-00156287, 33%), 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업의 지원(RS-2025-25398164, 33%), 그리고 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2024-00357917, 34%).

### 참고문헌

[1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciampi, M. Ghafoorian, ... and C. I. Sánchez, "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, Vol. 42, pp. 60-88, December 2017. <https://doi.org/10.1016/j.media.2017.07.005>

[2] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing Imperfect Datasets: A Review of Deep Learning Solutions for Medical Image Segmentation," *Medical Image Analysis*, Vol. 63, 101693, July 2020. <https://doi.org/10.1016/j.media.2020.101693>

[3] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-Supervised Medical Image Segmentation Through Dual-Task Consistency," in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 10, pp. 8801-8809, 2021.

[4] X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang, "Self-Supervised Learning: Generative or

Contrastive," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 35, No. 1, pp. 857-876, January 2023. <https://doi.org/10.1109/TKDE.2021.3090866>

[5] J. Zbontar, L. Jing, I. Misra, Y. LeCun, and S. Deny, "Barlow Twins: Self-Supervised Learning via Redundancy Reduction," in *Proceedings of the 38th International Conference on Machine Learning (ICML)*, Vol. 139, pp. 12310-12320, 2021.

[6] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," in *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278-2324, November 1998. <https://doi.org/10.1109/5.726791>

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, pp. 234-241, 2015. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)

[8] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, Vol. 313, No. 5786, pp. 504-507, July 2006. <https://doi.org/10.1126/science.1127647>

[9] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *Proceedings of the International Conference on Learning Representations (ICLR)*, April 2014. <https://doi.org/10.48550/arXiv.1312.6114>

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... and I. Polosukhin, "Attention Is All You Need," in *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS)*, Long Beach: CA, pp. 6000-6010, 2017

[11] M. M. Rahman, S. Shokouhmand, S. Bhatt, and M. Faezipour, "MIST: Medical Image Segmentation Transformer with Convolutional Attention Mixing (CAM) Decoder," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 403-412, 2024.

[12] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment Anything in Medical Images," *Nature Communications*, Vol. 15, 654, 2024.

[13] Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, ... and D. Ni, "Segment Anything Model for Medical Images?," *Medical Image Analysis*, Vol. 92, 103061, 2024.

[14] R. Jiao, Y. Zhang, L. Ding, B. Xue, J. Zhang, R. Cai, and C. Jin, "Learning with Limited Annotations: A Survey on Deep Semi-Supervised Learning for Medical Image

Segmentation,” *Computers in Biology and Medicine*, Vol. 169, 107840, 2024.

- [15] N. S. Punn and S. Agarwal, “BT-Unet: A Self-Supervised Learning Framework for Biomedical Image Segmentation Using Barlow Twins with U-Net Models,” *Machine Learning*, Vol. 111, pp. 4585-4600, August 2022. <https://doi.org/10.1007/s10994-022-06219-3>
- [16] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation,” in *Proceedings of the 4th International Conference on 3D Vision (3DV)*, Stanford: CA, pp. 565-571, October 2016. <https://doi.org/10.1109/3DV.2016.79>
- [17] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, “A Tutorial on the Cross-Entropy Method,” *Annals of Operations Research*, Vol. 134, pp. 19-67, February 2005. <https://doi.org/10.1007/s10479-005-5724-z>
- [18] S. P. Morozov, A. E. Andreychenko, N. A. Pavlov, A. V. Vladzmyrskyy, N. V. Ledikhova, V. A. Gombolevskiy, ... and V. Y. Chernina, “MosMedData: Chest CT Scans with COVID-19 Related Findings Dataset,” arXiv:2005.06465, 2020. <https://doi.org/10.48550/arXiv.2005.06465>
- [19] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, ... and H. Jian, “COVID-19 CT Lung and Infection Segmentation Dataset,” *Zenodo*, 2020.
- [20] M. Maftouni. COVID-19 CT Scan Lesion Segmentation Dataset [Internet]. Available: <https://www.kaggle.com/datasets/maedemaftouni/covid19-ctscan-lesion-segmentation-dataset>.
- [21] Kaggle. Ultrasound Nerve Segmentation Dataset [Internet]. Available: <https://www.kaggle.com/c/ultrasound-nerve-segmentation/>.

### 박민영 (Minyeong Park)



2023년~현 재: 전남대학교 컴퓨터정보통신공학과 학사과정  
 ※ 관심분야: 딥러닝, 컴퓨터 비전 등

### 박채현 (Chaeheon Park)



2019년~현 재: 전남대학교 컴퓨터정보통신공학과 학사과정  
 ※ 관심분야: 머신러닝, 컴퓨터 비전 등

### 류현승 (Hyunseung Ryu)



2025년 : 전남대학교  
 컴퓨터정보통신공학과 (학사)

2025년~현 재: 전남대학교 지능전자컴퓨터공학과 석사과정  
 ※ 관심분야: 딥러닝, 컴퓨터비전

### 박수형 (Suhung Park)



2008년 : 한양대학교 (공학사)  
 2011년 : 연세대학교 대학원 (이학석사)  
 2015년 : 고려대학교 대학원 (공학박사)

2016년~2017년: 성균관대학교  
 2017년~2020년: University of California, Berkeley  
 2019년~현 재: 전남대학교 전자컴퓨터공학부 부교수  
 ※ 관심분야: 딥러닝, 영상처리