

# 제한된 학습 데이터 환경에서 디퓨전 생성 모델을 활용한 준지도 학습 기반 초음파 영상 생검 바늘 세그멘테이션

전지현<sup>1</sup> · 이성훈<sup>1</sup> · 박수형<sup>2\*</sup><sup>1</sup>전남대학교 지능전자컴퓨터공학과 석사과정<sup>2</sup>전남대학교 전자컴퓨터공학부 부교수

## Semi-Supervised Biopsy Needle Segmentation Using Diffusion Models with Limited Training Data in Ultrasound

Jihyeon Jeon<sup>1</sup> · Sunghoon Lee<sup>1</sup> · Suhyung Park<sup>2\*</sup><sup>1</sup>Master's course, Department of Intelligent Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea<sup>2</sup>Professor, Department of Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea

### [요약]

본 연구는 초음파 유도 생체검사에서 바늘의 정확한 위치 추적을 위한 딥러닝 기반 세그멘테이션 프레임워크를 제안한다. 본 연구에서는 제한된 레이블된 데이터와 고품질의 레이블 없는 데이터의 부족이라는 중요한 문제를 해결하기 위해 프레임 예측 디퓨전 모델인 Masked Conditional Video Diffusion(MCVD)을 활용한 준지도 학습 접근법을 제안하였다. MCVD 모델은 시간적 프레임 예측을 통하여 추가적인 현실적 초음파 시퀀스를 생성하여 레이블 없는 데이터의 부족 문제를 해결하였다. 다음으로, Pseudo Labels 알고리즘을 통해 교사 모델이 레이블 없는 데이터를 통해 생성한 의사 레이블로 학생 모델이 학습하도록 하여 레이블된 데이터와 레이블 없는 데이터를 모두 활용하여 점진적인 성능 향상을 달성하였다. 종합적인 실험을 통해 제안한 프레임워크가 U-Net, Attention U-Net, Swin U-Net 등 기존 세그멘테이션 모델들에 비해 Dice Score, IoU, 바늘 끝점 위치 오차, 궤적 각도 오차를 포함한 다양한 평가 지표에서 가장 우수함을 확인하였다.

### [Abstract]

This study proposes a deep learning-based segmentation framework for accurate needle monitoring in ultrasound-guided biopsy procedures. In this study, we propose a semi-supervised learning approach utilizing a frame-prediction diffusion model, Masked Conditional Video Diffusion (MCVD), to address the critical challenges of limited labeled data and insufficient high-quality unlabeled data. The MCVD model generates additional realistic ultrasound sequences through temporal frame prediction, thus effectively addressing the issue of unlabeled data scarcity. Subsequently, through the pseudo labels algorithm, the teacher model generates pseudo labels from unlabeled data for the student model to learn, where both labeled and unlabeled data are utilized to achieve progressive performance improvement. Comprehensive experiments demonstrate that the proposed framework outperforms existing segmentation models including U-Net, Attention U-Net, and Swin U-Net across multiple evaluation metrics, including the Dice score, IoU, tip-position error, and trajectory-angle error.

**색인어** : 영상 분할, 초음파 생검, 딥러닝, 준지도 학습, 확산 모델**Keyword** : Image Segmentation, Ultrasound-Guided Biopsy, Deep Learning, Semi-Supervised Learning, Diffusion Model<http://dx.doi.org/10.9728/dcs.2025.26.10.2839>

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 04 September 2025; Revised 25 September 2025

Accepted 02 October 2025

\*Corresponding Author; Suhyung Park

Tel: +82-62-530-1797

E-mail: suhyung@jnu.ac.kr

## I. 서론

초음파 유도 생체검사는 현대 정밀의학에서 암 진단 및 조직병리학적 분석을 위한 핵심적인 최소침습적 진단 절차로 자리매김하고 있다[1]. 이 기술은 실시간 초음파 영상 가이드를 하에 목표 조직에 바늘을 정확히 삽입하여 조직 샘플을 채취하는 방법으로, 개복 수술 대비 환자의 이환율을 현저히 감소시키면서도 높은 진단 정확도를 제공한다. 특히 유방, 간, 신장, 전립선 등 다양한 장기의 국소 병변 진단에 있어 그 임상적 유용성이 광범위하게 입증되어 있으며 전 세계적으로 매년 수백만 건의 초음파 유도 생체 검사가 시행되고 있다. 성공적인 조직 채취와 환자의 안전을 위해서는 실시간 바늘 추적 및 정확한 바늘 시각화가 필수적이며 이는 기술의 성공률과 직접적으로 연관된다. 그러나 초음파 영상에서 바늘의 식별은 초음파 영상의 고유한 특성인 저대비, 스펙클 노이즈, 음향 그림자로 인해 바늘과 주변 조직 간의 구별이 어렵다. 특히 바늘이 초음파 빔과 수직이 아닌 각도로 삽입될 때 발생하는 경면 반사 소실은 바늘의 가시성을 크게 저하시킨다. 또한 바늘의 두께가 초음파 파장보다 작을 때 나타나는 comet-tail artifact와 같은 영상 잡음은 정확한 바늘 위치 파악을 더욱 어렵게 만든다. 이러한 물리적 제약으로 인해 숙련된 의료진조차 복잡한 해부학적 구조에서 바늘을 추적하는데 어려움을 겪으며 이는 기술 시간의 연장과 환자 안전성 저하로 이어질 수 있다[2],[3].

최근 딥러닝 기술의 급속한 발전과 함께 의료 영상 분석 분야에서 유망한 성과들이 나타나고 있다. Convolutional Neural Networks를 기반으로 한 U-Net과 같은 아키텍처들은 의료 영상 세그멘테이션에서 뛰어난 성능을 보여주었으며 초음파 바늘 검출 분야에서도 다양한 접근법들이 제안되어 왔다[4]. 또한, Vision Transformer 기반 방법들은 self-attention 메커니즘을 통해 global feature를 효과적으로 캡처하여 기존 CNN 기반 방법 대비 우수한 성능을 달성하였다. 그러나 이러한 지도 학습 기반 딥러닝 방법들은 공통적으로 대량의 고품질 labeled data를 필요로 한다는 근본적인 한계를 가지고 있다. 이는, 의료 영상 분야의 특성상 전문가의 정밀한 주석 작업이 요구되는 labeled data의 확보는 극도로 제한적이며 데이터 수집 및 라벨링 과정에서 발생하는 높은 비용과 시간적 제약이 실제 임상 적용의 주요 장벽으로 작용하고 있다[5].

이러한 데이터 부족 문제를 해결하기 위한 대안으로서 준지도 학습 방법론들이 주목받고 있다[6]. 특히 pseudo-labeling과 같은 기법은 unlabeled data에 대해 모델이 생성한 예측을 pseudo label로 활용하여 학습 데이터를 확장하는 효과적인 해결책으로 제시되고 있다[7]. Pseudo Labels는 교사-학생 구조를 통해 pseudo label의 품질을 점진적으로 개선하여 더 높은 성능을 달성하는 방법론으로, 제한된 labeled data 환경에서의 효과가 입증되었다[8]. 그러나 의

료 영상 분야에서는 고품질의 unlabeled data조차 제한적인 경우가 많아 단순한 준지도 학습만으로는 충분한 성능 향상을 기대하기 어렵다. 이러한 맥락에서 생성 모델을 활용한 데이터 증강이 추가적인 해결책으로 부상하고 있다[9],[10]. 특히 Diffusion model은 높은 품질의 현실적인 의료 영상 생성 능력으로 인해 큰 관심을 받고 있다[11].

이러한 문제들을 해결하기 위해 본 연구에서는 Masked Conditional Video Diffusion(MCVD)을 통하여 의사 데이터를 생성 및 이를 통해 추가적인 pseudo label을 생성 후 교사-학생 메커니즘을 통해 학습하는 준지도 학습 프레임워크를 제안한다[12].

MCVD를 통해 초음파 비디오 시퀀스의 시간적 일관성을 활용하여 현실적인 바늘 삽입 시퀀스를 생성함으로써 unlabeled data를 효과적으로 증강하고 이를 기존 unlabeled data 및 제한된 labeled data와 함께 Pseudo Labels의 교사-학생 학습 메커니즘에 활용한다.

본 연구의 주요 기여도는 다음과 같다:

1. 초음파 기반 생검 바늘 검출을 위해 최초로 Video diffusion model을 활용한 시공간적 데이터 증강 방법론을 제안하여 현실적이고 일관성 있는 바늘 삽입 시퀀스 생성을 가능케 함
2. Pseudo Labels 알고리즘을 의료 영상의 이진 세그멘테이션 문제에 특화하여 적응시키고 최적화함으로써 제한된 데이터 환경에서 효과적인 준지도 학습을 구현
3. Labeled data, unlabeled data, 그리고 생성된 synthetic data를 통합적으로 활용하는 상호 교사-학생 학습 메커니즘을 통해 강건한 바늘 검출 성능을 달성

## II. 연구 방법

### 2-1 연구방법론 및 준지도 학습 프레임워크 구조

#### 1) 학습 프레임워크 구조

본 연구는 제한된 labeled data 환경에서 효과적인 초음파 바늘 세그멘테이션을 달성하기 위해 two-step 접근법을 제안한다. 첫 번째 단계에서는 Masked Conditional Video Diffusion(MCVD) 모델을 활용하여 시공간적 일관성을 보장하는 데이터 증강을 수행하며 두 번째 단계에서는 Pseudo Labels 알고리즘을 활용한 준지도 학습을 진행한다(그림 1). 이러한 접근법은 기존 unlabeled data와 생성된 augmented data, 그리고 제한된 labeled data를 통합적으로 활용하여 강건한 바늘 검출 성능을 달성한다. 이를 위해 전체 시스템은 크게 세 가지 주요 구성요소로 이루어져 있다. 첫째, MCVD 기반 데이터 생성 모듈은 초기 2개의 프레임을 조건으로 하여 후속 5개 프레임을 생성하는 조건부 확산 모델로 구현된다.

둘째, U-Net 기반의 교사-학생 네트워크는 서로 다른 초

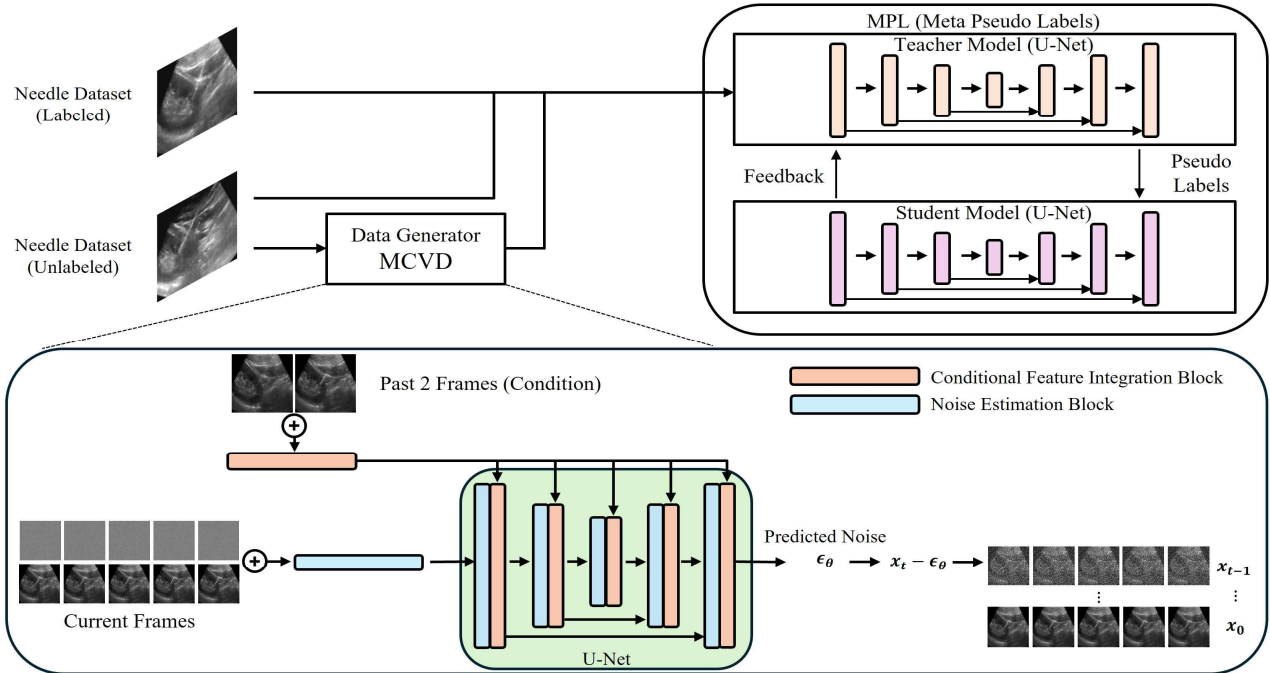


그림 1. 제안된 초음파 바늘 세그멘테이션 시스템의 전체 아키텍처. MCVD 모델을 통한 데이터 증강과 Pseudo Labels를 활용한 교사-학생 상호 학습을 결합한 프레임워크

Fig. 1. Overall architecture of the proposed ultrasound needle segmentation system. Framework combining data augmentation through MCVD model and teacher-student mutual learning using Pseudo Labels

기회를 가진 동일한 구조의 세그멘테이션 모델로 구성된다. 셋째, 메타 학습 메커니즘은 교사 모델의 pseudo label 품질을 학생 모델의 성능 변화를 통해 평가하고 개선하는 피드백 루프를 제공한다.

2-2 MCVD 모델 활용 및 적용

1) 데이터 증강을 위한 생성형 확산 모델 구조

MCVD 모델은 초음파 영상 시퀀스의 시간적 연속성을 보존하면서 현실적인 바늘 삽입 과정을 생성하기 위해 설계된 조건부 확산 모델이다[12]. 모델의 입력은 연속된 7개의 프레임으로 구성되며 처음 2개는 조건 프레임  $x_{1:2}$ 로, 나머지 5개는 생성 대상 프레임  $x_{3:7}$ 로 정의된다.

확산 모델의 순방향 과정은 생성 대상 프레임에 점진적으로 가우시안 노이즈를 추가하는 과정으로 다음과 같이 정의된다(수식 1):

$$q(x_{3:7}^t | x_{3:7}^0) = N(x_{3:7}^t; \sqrt{\bar{\alpha}} x_{3:7}^0, (1 - \bar{\alpha}_t)I) \tag{1}$$

여기서  $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$ 이고,  $\alpha_t = 1 - \beta_t$ 는 미리 정의된 노이즈 스케줄에 따라 결정된다.

역방향 과정은 노이즈가 추가된 프레임으로부터 조건 프레임을 참조하여 원본 프레임을 복원하는 과정이다. 모델은 각 시간 스텝에서 추가된 노이즈를 예측하도록 학습되며 이를

통해 최종적으로 조건 프레임과 일관성 있는 미래 프레임들을 생성할 수 있게 된다(수식 2):

$$p_\theta(x_{3:7}^{t-1} | x_{3:7}^t, x_{1:2}) = N(x_{3:7}^{t-1}; \mu_\theta(x_{3:7}^t, x_{1:2}, t), \sum_\theta(t)) \tag{2}$$

여기서  $p_\theta(x_{3:7}^{t-1} | x_{3:7}^t, x_{1:2})$ 는 조건 프레임  $x_{1:2}$ 와 특정 시간 스텝의 노이즈가 추가된 후속 프레임  $x_{3:7}^t$ 가 주어졌을 때 이전 시간 스텝의 프레임  $x_{3:7}^{t-1}$ 에 대한 조건부 확률 분포를 나타낸다.  $\mu_\theta(x_{3:7}^t, x_{1:2}, t)$ 는 신경망 파라미터  $\theta$ 로 매개화된 평균 함수로, 노이즈가 추가된 후속 프레임과 조건 프레임, 그리고 시간 스텝 정보를 입력으로 받아 이전 시간 스텝의 평균을 예측한다.  $\sum_\theta(t)$ 는 시간 스텝  $t$ 에 따른 분산 스케줄로,

각 denoising step에서 추가할 노이즈의 크기를 결정한다. 이러한 역방향 과정을 통해 순수한 가우시안 노이즈로부터 시작하여 점진적으로 현실적인 비디오 프레임을 생성할 수 있다.

2) U-Net 기반 노이즈 추정 네트워크

MCVD의 핵심인 노이즈 추정 네트워크는 3D U-Net 구조를 기반으로 설계되어 시공간적 특징을 효과적으로 처리할 수 있도록 구성된다[13]. 네트워크는 시간 임베딩(time

embedding)  $\gamma(t)$ 를 포함하여 각 확산 단계에서 적응적으로 동작한다(수식 3):

$$\epsilon_{\theta}(x_t, t) = UNet(x_t, \gamma(t)) \tag{3}$$

손실 함수는 실제 노이즈와 예측된 노이즈 간의 MSE를 최소화하여, 조건부 확산 과정에서 시간적 일관성을 유지하며 현실적인 프레임을 생성할 수 있도록 한다(수식 4):

$$E[\|\epsilon - \epsilon_{\theta}(\sqrt{a_t} x_{3:7}^0 + \sqrt{1-a_t} \epsilon | x_{1:2}, t)\|^2] \tag{4}$$

5프레임보다 더 긴 데이터를 생성하고자 할 때는 자동회귀 방식을 사용하여, 이전에 예측한 5프레임 중 마지막 2프레임을 다음 예측의 조건으로 사용하여 이 과정을 반복해 최종적으로 원하는 길이의 미래 프레임들을 생성하게 된다.

예를 들어, 초기 조건 프레임  $x_{1:2}$ 로부터 첫 번째 5프레임  $x_{3:7}$ 을 생성한 후, 생성된 프레임 중 마지막 2개  $x_{6:7}$ 을 새로운 조건으로 사용하여 다음 5프레임을 생성한다(수식 5):

$$x_{8:12} \sim p_{\theta}(x_{8:12} | x_{6:7}) \tag{5}$$

이러한 방식을 반복함으로써 총 28개의 연속된 프레임을 생성할 수 있으며 각 단계에서 이전 프레임의 정보가 자연스럽게 전달되어 전체 시퀀스의 시간적 일관성이 유지된다.

• 데이터셋 확장

최종적으로 MCVD 모델을 통해 추가로 생성된 1,106개의 데이터를 기존 데이터와 합쳐 총 3,759개의 확장된 unlabeled dataset을 구성하였다.

2-3 Pseudo Labels 모델 설계 및 구현

1) Pseudo Labels 도입 및 적용

Pseudo Labels의 프레임워크는 교사-학생 모델 모두 U-Net 구조를 기반으로 구성된다. 두 모델은 서로 다른 초기 가중치를 가지며, 교사 모델은 unlabeled data에 대한 pseudo label을 생성하고, 학생 모델은 이를 활용하여 학습한다. 이는, 학생 모델의 성능 변화를 통해 교사 모델에게 피드백을 제공하고 이러한 피드백 과정을 통해 교사 모델은 점진적으로 더 정확하고 신뢰할 수 있는 pseudo label을 생성하게 되며 이는 다시 학생 모델의 성능 향상으로 이어지는 선순환 구조로 설계하였다.

2) 피드백 메커니즘

매 학습 스텝마다 학생 모델이 labeled data로 손실을 측정한다. 현재 스텝에서 진행한 손실 측정 값을 new loss로 두고, 이전 스텝의 손실을 old loss로 두어 두 값의 차이를 계

산한다. 이 값이 음수라면 현재 손실이 더 낮다는 의미이기 때문에 교사 모델이 올바른 pseudo label을 만들고 있다고 피드백을 주고, 반대로 양수 값이 나오면 만들고 있는 pseudo label이 잘못된 방향이라는 피드백을 주어 상호 학습을 하도록 한다(수식 6). 각 학습 스텝에서 다음과 같은 과정을 거친다: 1) 먼저 학생 모델이 현재 교사 모델의 pseudo label을 사용하여 한 스텝 업데이트를 수행한다, 2) 업데이트된 학생 모델이 성능을 labeled data에서 평가하여 피드백 신호를 생성한다:

$$L_{MPL} = (L_{student}^{new} - L_{student}^{old}) \cdot DiceBCELoss(\hat{y}_u^t, y_u^{pseudo}) \tag{6}$$

여기서  $L_{student}^{new}$  와  $L_{student}^{old}$  는 각각 new loss와 old loss로 각각 현재 스텝에서 업데이트된 학생 모델과 업데이트 이전 학생 모델이 동일한 labeled data에 대해 계산한 손실을 나타낸다. 이들의 차이는 교사 모델의 pseudo label이 학생 모델 성능에 미친 영향을 측정하는 피드백 신호로 사용된다.  $\hat{y}_u^t$ 는 교사 모델이 unlabeled data에 대해 예측한 로짓 출력이며,  $y_u^{pseudo}$ 는 이로부터 생성된 이전 pseudo label이다. 이러한 구조를 통해 학생 모델의 성능 변화가 교사 모델의 pseudo label 생성 시 품질 개선에 직접적으로 반영된다.

세그멘테이션 작업의 특성과 클래스 불균형 문제를 고려하여 Dice Loss 와 Binary Cross-Entropy Loss를 결합하여 다음과 같은 손실 함수를 사용한다(수식 7)-(수식 9):

$$L_{teacher} = L_{supervised} + L_{MPL} \tag{7}$$

$$L_{supervised} = DiceBCELoss(\hat{y}_i^t, y_i^t) \tag{8}$$

$$L_{student} = DiceBCELoss(\hat{y}_u^s, y_u^{pseudo}) \tag{9}$$

여기서  $L_{supervised}$ 는 labeled data에 대한 교사 모델의 예측에 대한 손실 함수로 교사 모델의 전체 손실 함수의 일부를 구성한다.  $L_{student}$ 는 교사 모델이 생성한 pseudo label과 unlabeled에 대한 학생 모델의 예측으로 손실 함수가 구성된다.

2-4 데이터의 통합 및 학습 방법

전체 학습 과정은 기존 unlabeled data, MCVD로 생성된 augmented data, 그리고 제한된 labeled data를 통합적으로 활용한다. 최종적으로 MCVD 모델을 통해 생성된 1,106개의 데이터를 기존 unlabeled data와 합쳐 총 3,759개의 확장된 데이터셋을 구성하였다.

학습은 다음과 같은 스케줄로 진행된다. 교사와 학생 모델

을 랜덤 초기화한 상태에서 Pseudo Labels 알고리즘을 적용하여 교사와 학생 모델을 동시에 최적화한다. 각 에포크에서 배치는 labeled data와 unlabeled data(original + augmented)의 비율을 1:4로 구성하여 데이터 불균형을 완화한다.

### III. 실험 및 결과

#### 3-1 실험 환경 및 평가 지표

##### 1) 데이터셋 구성

본 연구에서 사용한 데이터셋은 다음과 같이 구성되었다.

##### • Labeled data

원본 Labeled data 이미지를 회전 증강(Rotation augmentation)을 통해 확장하였다. 각 원본 이미지당 7배 증강을 적용하여 총 1,162개의 labeled data를 구성하였다.

##### • Unlabeled data

Unlabeled data는 두 가지 방법으로 구성하였다. 먼저 기존 49,800개의 초음파 영상 중 바늘 가시성이 우수한 2,653개를 선별하였다. 또한 MCVD 모델을 통해 현실적인 바늘 삽입 시퀀스 1,106개를 추가로 생성하였다. 이를 통해 Unlabeled data를 구성하였다.

##### • 전체 데이터셋

최종적으로 labeled data 1,162개와 unlabeled data 3,759개로 구성된 총 4,921개의 데이터셋을 사용하여 실험을 수행하였다.

##### 2) 평가 지표

##### • Dice Score

예측된 바늘 영역과 실제 바늘 영역 간의 중복도를 측정하는 대표적인 세그멘테이션 평가 지표이다(수식 10). 두 영역의 교집합에 2를 곱한 값을 두 영역의 크기 합으로 나누어 계산된다. 0에서 1 사이의 값을 가지며, 1에 가까울수록 예측이 정확함을 의미한다. 의료 영상 세그멘테이션에서 널리 사용되며, 클래스 불균형 상황에서도 안정적인 평가가 가능하다.

##### • IoU(Intersection over Union)

예측된 바늘 영역과 실제 바늘 영역의 교집합을 합집합으로 나눈 비율로 계산된다(수식 11). 객체 검출과 세그멘테이션 분야에서 표준 평가 지표로 사용된다. Dice Score와 수학적으로 연관되어 있으며, 일반적으로 Dice Score보다 낮은 값을 보인다.

##### • Accuracy

전체 픽셀 중에서 올바르게 분류된 픽셀의 비율을 나타내는 기본적인 분류 성능 지표이다(수식 12). 직관적이고 이해하기 쉬운 장점이 있으나, 클래스 불균형 문제에 취약하다. 본 연구의 초음파 바늘 데이터셋의 경우 바늘 영역(양성 클래스)이 전체 픽셀의 1% 미만을 차지하는 극심한 클래스 불균형이 존재한다. 이러한 환경에서는 모든 픽셀을 배경으로 예측해도 99% 이상의 높은 Accuracy를 얻을 수 있어, 실제 세그멘테이션 성능을 제대로 반영하지 못한다. 따라서 Dice Score와 IoU가 모델의 실질적 성능을 더 신뢰성 있게 평가한다.

##### • Tip-position Error

예측된 바늘 끝점과 실제 바늘 끝점 간의 유클리드 거리를 픽셀 단위로 측정된 값이다. 값이 낮을수록 바늘 끝점 예측이 정확함을 의미한다.

##### • Trajectory Angle Error

예측된 바늘의 삽입 궤적과 실제 궤적 간의 각도 차이(도(degree)) 단위로 측정된 값이다. 바늘의 삽입 방향과 각도는 목표 조직에 정확히 도달하기 위해 매우 중요하며, 이 값이 낮을수록 예측된 바늘 궤적이 실제와 일치함을 나타낸다.

$$Dice\ Score = \frac{2(A \cap B)}{|A| + |B|} \quad (10)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (11)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

##### 3) 비교 모델

제안한 방법의 우수성을 입증하기 위해 다음 3가지 기준 세그멘테이션 모델과 비교하였다.

##### • U-Net

Ronneberger 등이 2015년에 제안한 의료 영상 세그멘테이션의 표준 모델로, 인코더-디코더 구조와 스킵 연결을 통해 세밀한 공간적 정보를 보존하면서도 효과적인 특징 추출을 가능하게 한다[4]. 의료 영상 세그멘테이션에서 60-70% 수준의 Dice Score를 달성하여 널리 사용되고 있으며, 본 연구에서도 baseline 모델로 활용하였다.

##### • Attention U-Net

Oktay 등이 2018년에 제안한 모델로, 기존 U-Net에 어텐션 메커니즘을 추가하여 중요한 특징에 집중할 수 있도록 개선한 모델이다[14]. 어텐션 게이트를 통해 관련성이 높은

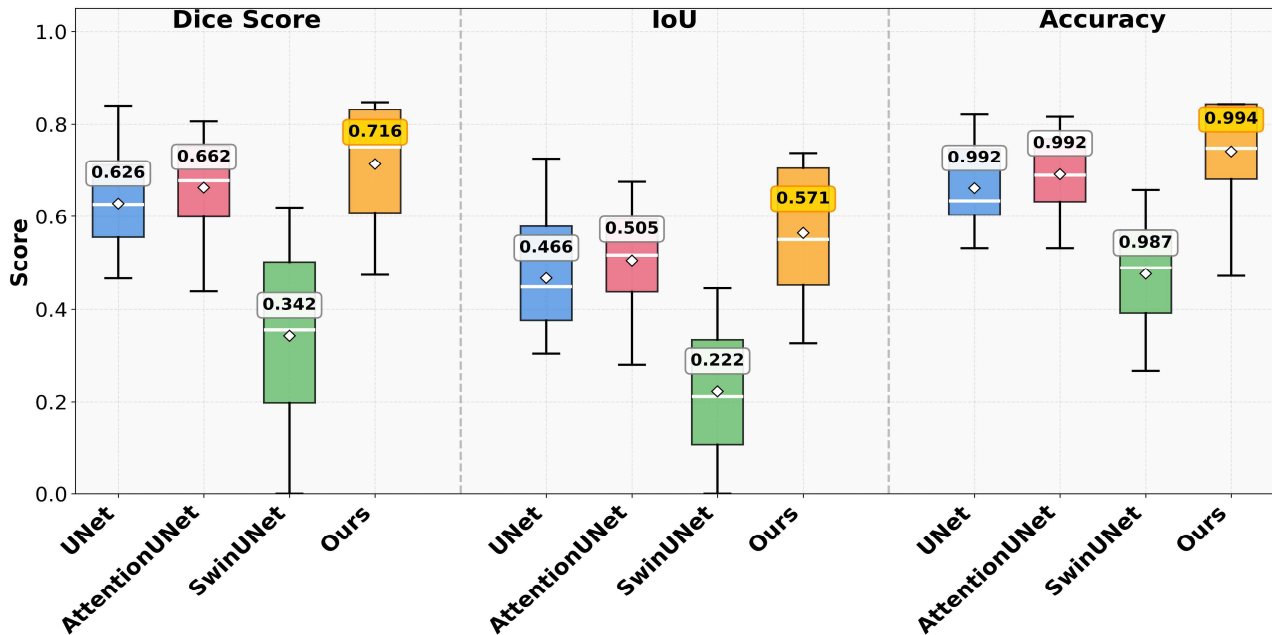


그림 2. 제안한 프레임워크와 비교 모델들의 주요 성능 지표별 박스플롯 결과(Dice Score, IoU, Accuracy). 제안한 프레임워크가 가장 높은 성능을 보임

Fig. 2. Box plot comparison results of key performance metrics between the proposed framework and comparison models (Dice Score, IoU, Accuracy). The proposed framework shows the highest performance

특징만을 선택적으로 전달함으로써 기존 U-Net 대비 5-10%의 성능 향상을 이루었다. 특히 복잡한 해부학적 구조를 가진 의료 영상에서 우수한 성능을 보여준다.

• Swin U-Net

Cao 등이 2022년에 제안한 Transformer 기반의 최신 세그멘테이션 모델로, Swin Transformer를 인코더로 사용하고 patch expanding layer를 디코더로 활용한다[15]. 전역적 특징을 포착하는 능력이 뛰어나지만, Transformer의 특성상 대량의 학습 데이터를 필요로 하여 제한된 데이터 환경에서는 성능 저하가 발생할 수 있다.

4) 실험 설정 및 비교 조건

공정한 성능 비교를 위해 다음과 같이 실험을 설계하였다.

• 기존 지도 학습 모델들

비교 모델인 U-Net, Attention U-Net, Swin U-Net은 순수 지도 학습 방식으로 학습하였다. 이 모델들은 회전 증강을 거친 labeled data 1,162개만을 사용하여 학습을 수행하였다. 학습률 0.0001, 배치 크기 6, 100 에포크로 설정하였다.

• 제안 프레임워크

제안한 프레임워크는 준지도 학습 방식으로 학습하였으며, 회전 증강을 통해 얻은 labeled data 1,162개와 전체 unlabeled data에서 선별한 것과 MCVD 모델을 통해 증강

하여 얻은 총 3,759개의 unlabeled data를 모두 활용하였다. 교사-학생 구조를 통한 상호 학습을 수행하여 unlabeled data로부터 추가적인 학습 정보를 획득하였다. 교사 모델의 학습률은 0.001, 학생 모델의 학습률은 0.0001로 설정하여 교사가 빠른 수렴을 통해 고품질의 pseudo label을 생성하도록 하였다(배치 크기 6, 30,000스텝). 이러한 실험 설계를 통해 제안한 방법론의 성능 향상이 단순한 데이터 증가가 아닌 준지도 학습 알고리즘의 효과임을 검증하고자 하였다.

3-2 실험 결과

표 1. 제안한 프레임워크와 비교 모델들의 정량적 성능 비교 결과 (Dice Score, IoU, Accuracy)

Table 1. Quantitative performance comparison results between the proposed framework and comparison models (Dice Score, IoU, Accuracy)

	U-Net	Attention U-Net	Swin U-Net	Ours
Dice Score	62.65 ±10.91	66.25 ±11.13	34.19 ±18.99	71.61 ±12.31
IoU	46.57 ±12.02	50.54 ±12.23	22.20 ±13.92	57.13 ±14.18
Accuracy	99.15 ±0.22	99.23 ±0.23	98.69 ±0.29	99.36 ±0.27

1) 정량적 평가 결과

Test dataset에 대한 실험 결과는 다음과 같다(표 1, 그림 2). 지도 학습 모델 U-Net, Attention U-Net, Swin U-Net

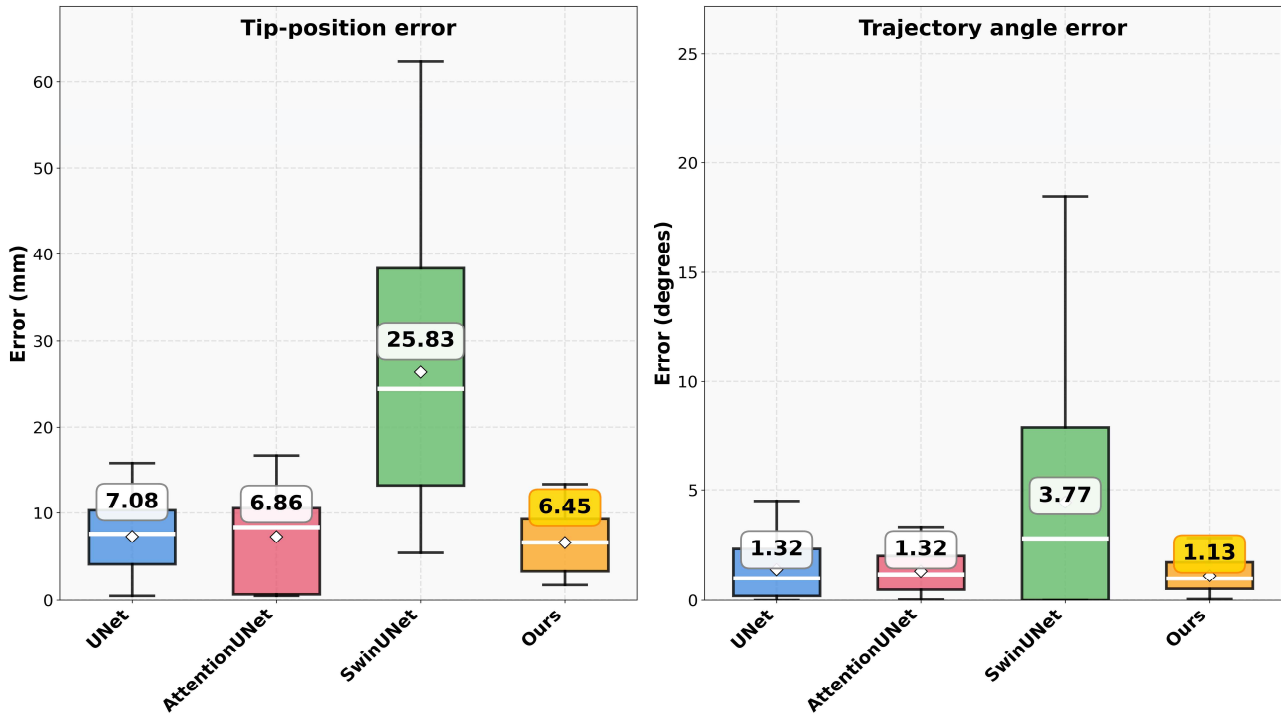


그림 3. 제안한 프레임워크와 비교 모델들의 주요 성능 지표별 박스플롯 결과(Tip-position error, Trajectory angle error). 제안한 프레임워크가 가장 높은 성능을 보임

Fig. 3. Box plot comparison results of key performance metrics between the proposed framework and comparison models (Tip-position error, Trajectory angle error). The proposed framework shows the highest performance

은 순서대로 Dice score 62.65%, 66.25%, 34.19%의 성능을 보였다. 이 중, Swin U-Net과 같은 Transformer 기반 모델은 다른 모델에 비해 훨씬 많은 학습 데이터를 필요로 하기 때문에, 제한된 데이터 환경에서 낮은 성능을 나타낸 것으로 해석된다. 반면, 제안 프레임워크는 71.61%의 Dice score로 비교 모델들의 성능을 능가했다. 비교 모델 중 최고 성능인 Attention U-Net 대비 8% 향상(66.25%→71.61%)을 이루었으며, IoU에서는 13% 향상(50.54%→57.13%)을 보였다. 특히 임상적으로 중요한 Tip-position error는 6%(6.86→6.45), Trajectory angle error는 14%(1.32→1.13) 개선되었다(표 2, 그림 3). 이러한 성능 향상은 두 가지 핵심 요소의 시너지 효과로 분석된다. 첫째, MCVD를 통한 unlabeled data의 데이터 증강으로 부족한 unlabeled data

문제를 해결하였다. 둘째, Pseudo Labels의 교사-학생 상호 학습 메커니즘이 labeled data와 unlabeled data를 모두 활용하여 점진적으로 성능을 개선할 수 있게 했다.

### 2) 정성적 평가 결과

본 연구의 정성적 분석에서는 제안 프레임워크와 비교 모델의 성능을 Ground Truth와 비교하여 주요 차이점을 도출하였다(그림 4). 이미지와 그에 대한 Ground Truth, 모델의 예측을 오버레이한 결과를 통해 각 모델의 예측 차이를 더 자세히 볼 수 있다(그림 5). 오버레이 이미지의 빨간 영역은 Ground Truth, 파란 영역은 모델의 예측, 녹색 영역은 Ground Truth와 모델의 예측이 겹치는 정확한 예측 부분을 나타낸다. 특히 제안 프레임워크가 다른 모델들에 비해 Ground Truth와 가장 높은 일치도를 보였으며 바늘의 경계와 형태를 더욱 정확하게 검출하는 것을 확인할 수 있다. U-Net과 Attention U-Net은 바늘의 전체적인 방향은 감지하지만 세부적인 경계 부분에서 상대적으로 부정확한 예측을 보였고 Swin U-Net은 바늘을 완전히 놓치는 문제를 보였다. 반면 제안 프레임워크는 오버레이 결과에서 녹색 영역이 가장 많이 관찰되며 파란 영역과 빨간 영역이 상대적으로 적어 우수한 세그멘테이션 성능을 시각적으로 입증한다.

### 3) 성능 향상 요인 분석

제안 방법론의 성능 향상은 두 핵심 구성요소의 상호 보완

표 2. 제안한 프레임워크와 비교 모델들의 정량적 성능 비교 결과 (Tip-position error, Trajectory angle error)

Table 2. Quantitative performance comparison results between the proposed framework and comparison models (Tip-position error, Trajectory angle error)

	U-Net	Attention U-Net	Swin U-Net	Ours
Tip-position error	7.08 ±4.53	6.86 ±5.58	25.83 ±17.07	6.45 ±3.91
Trajectory angle error	1.32 ±1.47	1.32 ±0.97	3.77 ±5.71	1.13 ±1.47

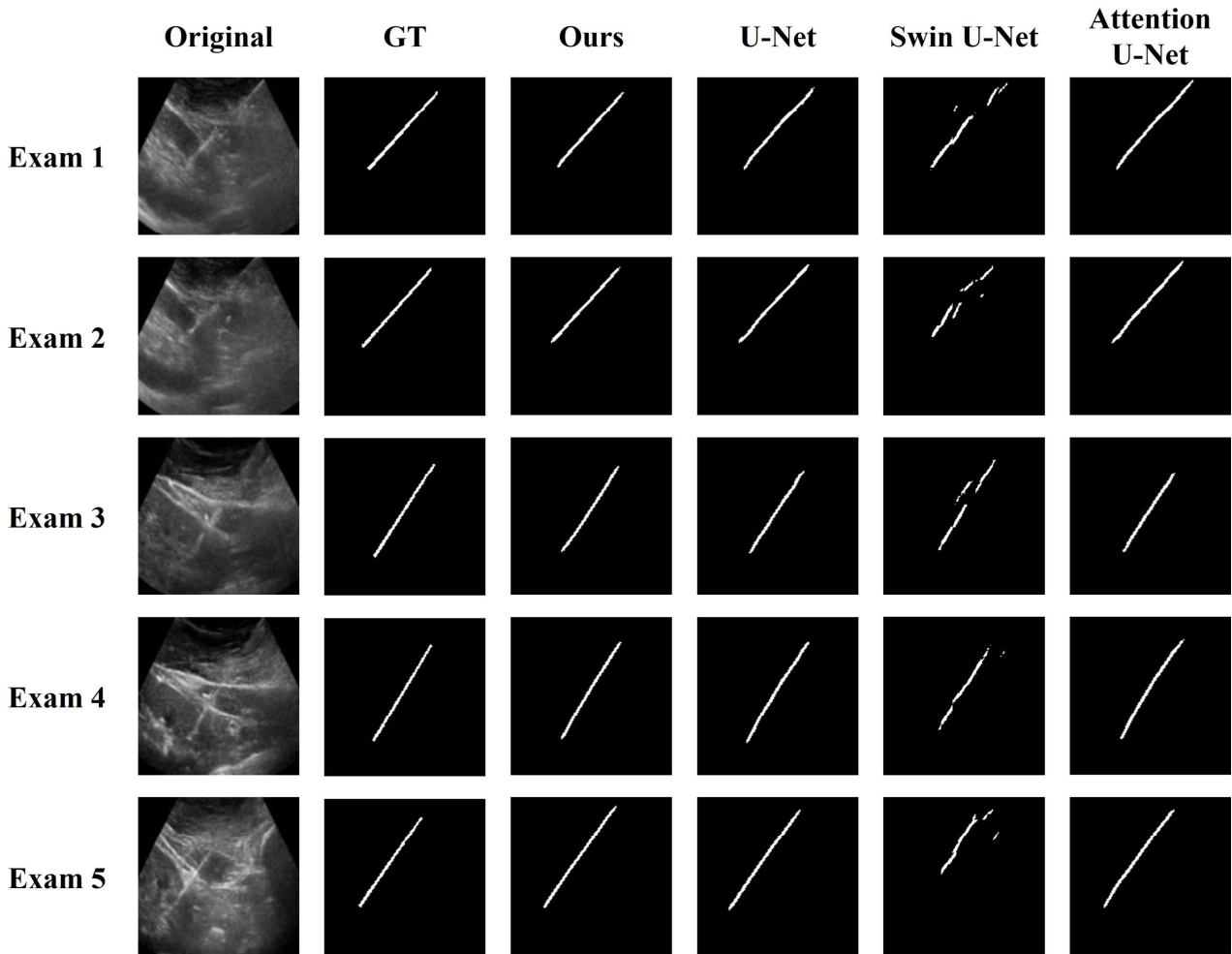


그림 4. 제안한 프레임워크와 비교 모델과의 정성적 비교 결과. 제안한 프레임워크가 비교 모델 대비 Ground Truth와 가장 유사한 바늘 세그멘테이션 결과를 보임

Fig. 4. Qualitative comparison results between the proposed framework and comparison models. The proposed framework shows needle segmentation results most similar to Ground Truth compared to comparison models

적 효과에서 기인한다. 첫째, MCVD 모델이 시간적 일관성을 유지하는 현실적인 초음파 시퀀스를 생성하여 학습 데이터의 질과 양을 개선하였다. 둘째, Pseudo Labels 알고리즘으로 교사-학생 상호 학습을 통해 unlabeled data의 잠재 정보를 효과적으로 활용했다. 이러한 두 요소의 시너지를 통한 적응적 pseudo label 생성 과정이 점진적 성능 향상을 가능하게 했으며, 특히 바늘 끝점 정확도 개선은 다양한 삽입 각도 학습과 정밀한 pseudo-labeling의 결합 효과로 해석된다.

#### IV. 결 론

본 연구에서는 초음파 유도 생체검사에서 바늘의 정확한 위치 추적을 위한 딥러닝 기반 세그멘테이션 프레임워크를 제안하였다. 의료 데이터 특성상 극히 제한된 labeled data와 충분하지 않은 unlabeled data 문제를 프레임 예측 디퓨전

모델(MCVD)로부터 생성된 Pseudo Labels를 결합하여 해결하였다는 점에서 중요한 의미를 가진다. 제안된 프레임워크의 가장 큰 의미는 의료 데이터가 부족한 환경에서 데이터 확보와 세그멘테이션 성능 개선을 동시에 달성했다는 점이다. MCVD 모델은 시간적 프레임 예측을 통해 현실적인 초음파 바늘 삽입 시퀀스를 생성하여 unlabeled data의 부족 문제를 해결했다. 이를 통해 생성된 고품질의 합성 unlabeled data는 준지도 학습의 효과를 극대화했다. Pseudo Labels의 도입은 제한된 labeled data 환경에서 unlabeled data를 효과적으로 활용하는 새로운 접근법을 제시하였다. 교사-학생 모델 구조를 통한 상호 학습 메커니즘은 MCVD가 생성한 unlabeled data를 포함하여 대량의 unlabeled data로부터 유용한 정보를 추출하고, 적응적 pseudo label 생성을 통해 점진적 성능 향상을 가능하게 했다. 종합적인 실험을 통해 제안한 방법이 다른 지도 학습 모델들을 다양한 평가 지표에서 능가함을 확인했다. 이러한 성능 향상은 초음파 유도 시술의

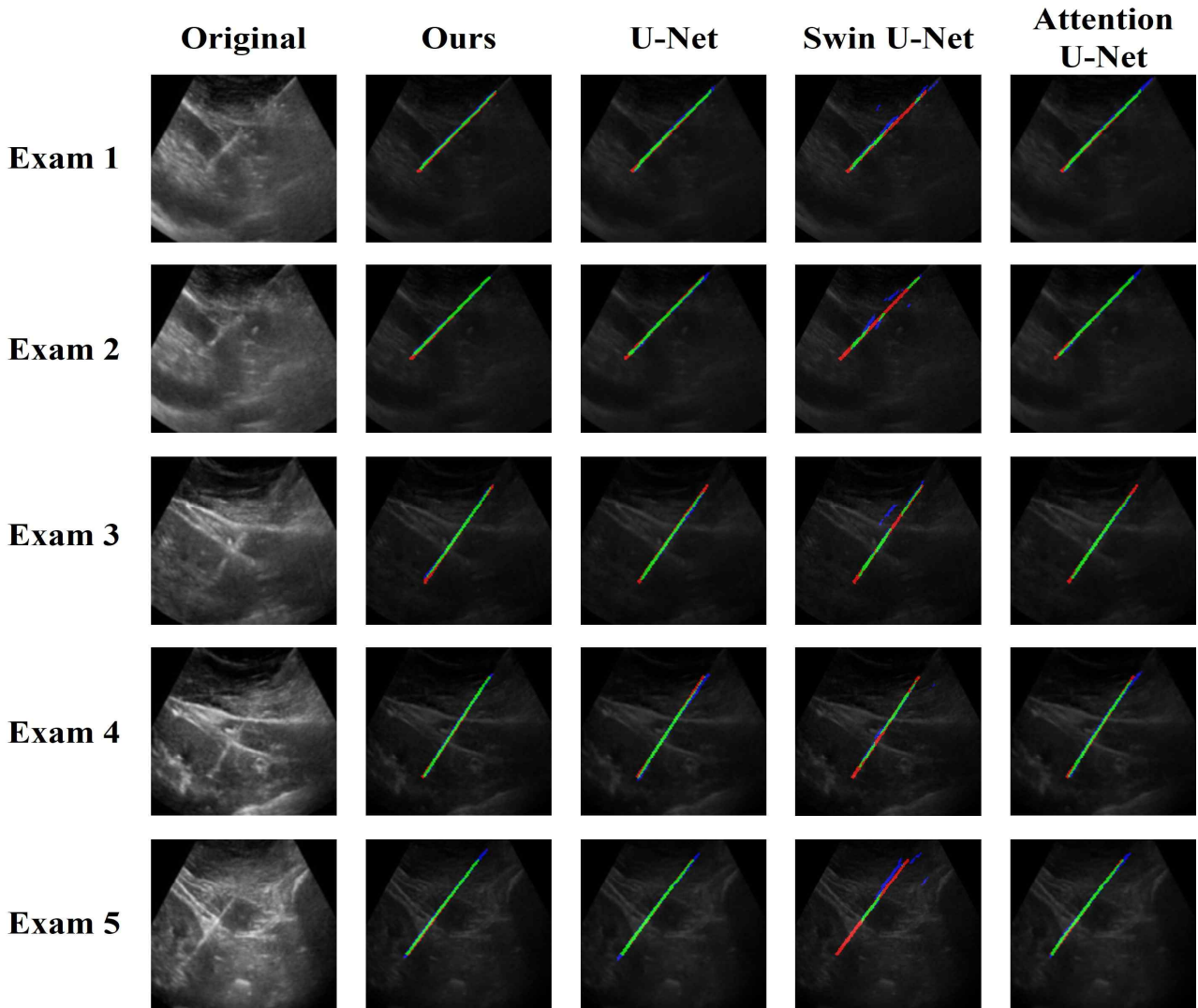


그림 5. 각 모델들의 결과와 원본 이미지와 Ground Truth를 오버레이한 결과. 파란 영역은 모델의 예측, 빨간 영역은 Ground Truth, 녹색 영역은 일치하는 부분을 나타냄

Fig. 5. Overlay results of each model's output with original images and Ground Truth. Blue regions represent model predictions, red regions represent Ground Truth, and green regions represent overlapping areas

정확도 향상과 환자 안전성 증대에 기여할 것으로 기대되며, 제한된 의료 데이터 환경에서의 효과적인 딥러닝 모델 개발에 새로운 방향을 제시한다.

### 감사의 글

이 논문은 1) 정부(과학기술정보통신부)의 재원으로 정보통신 기획평가원-지역지능화혁신인재양성사업(IITP-2025-RS-2022-00156287, 25%), 2) 학석사연계 ICT핵심인재양성사업의 연구결과(RS-2022-00156385, 25%), 3) 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (RS-2025-25398164, 25%). 그리고 4) 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2024-00357917, 25%).

### 참고문헌

- [1] G. A. Chapman, D. Johnson, and A. R. Bodenham, "Visualisation of Needle Position Using Ultrasonography," *Anaesthesia*, Vol. 61, No. 2, pp. 148-158, 2006. <https://doi.org/10.1111/j.1365-2044.2005.04475.x>
- [2] K. J. Chin, A. Perlas, V. W. S. Chan, and R. Brull, "Needle Visualization in Ultrasound-Guided Regional Anesthesia: Challenges and Solutions," *Regional Anesthesia and Pain Medicine*, Vol. 33, No. 6, pp. 532-544, 2008. <https://doi.org/10.1016/j.rapm.2008.06.009>
- [3] M. B. Rominger, K. Martini, E. Dappa, G. Puipe, V. Klingmüller, T. Frauenfelder, and S. J. Sanabria, "Ultrasound Needle Visibility in Contrast Mode Imaging:

- An In Vitro and Ex Vivo Study,” *Ultrasound in Medicine & Biology*, Vol. 43, No. 10, pp. 2355-2364, 2017. <https://doi.org/10.1055/s-0043-101511>
- [4] 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*, 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)
- [5] X. Zheng, C. Fu, H. Xie, J. Chen, X. Wang, and C.-W. Sham, “Uncertainty-Aware Deep Co-Training for Semi-Supervised Medical Image Segmentation,” *Computers in Biology and Medicine*, Vol. 149, 106051, 2022. <https://doi.org/10.1016/j.compbiomed.2022.106051>
- [6] 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, 1077840, 2022. <https://doi.org/10.1016/j.compbiomed.2023.107840>
- [7] D.-H. Lee, “Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks,” in *Proceedings of the International Conference on Challenges in Representation Learning Workshop*, Atlanta: GA, pp. 1-6, 2013.
- [8] H. Pham, Z. Dai, Q. Xie, and Q. V. Le, “Meta Pseudo Labels,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville: TN, pp. 11552-11563, 2021. <https://doi.org/10.1109/CVPR4643.7.2021.01139>
- [9] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof, “Diffusion Models in Medical Imaging: A Comprehensive Survey,” *Medical Image Analysis*, Vol. 88, 102846, 2023. <https://doi.org/10.1016/j.media.2023.102846>
- [10] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, “Data Augmentation Using Generative Adversarial Networks (CycleGAN) to Improve Generalizability in CT Segmentation Tasks,” *Scientific Reports*, Vol. 9, No. 1, 16884, 2019. <https://doi.org/10.1038/s41598-019-52737-x>
- [11] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, “Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review,” *Journal of Imaging*, Vol. 9, No. 4, 81, 2023. <https://doi.org/10.3390/jimaging9040081>
- [12] V. Voleti, A. Jolicœur-Martineau, and C. Pal, “MCVD: Masked Conditional Video Diffusion for Prediction, Generation, and Interpolation,” in *Proceedings of the Conference on Neural Information Processing Systems*, New Orleans: LA, pp. 23371-23385, 2022. <https://doi.org/10.48550/arXiv.2205.09853>
- [13] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” in *Proceedings of the Conference on Neural Information Processing Systems*, Virtual Event, pp. 6840-6851, 2020. <https://doi.org/10.48550/arXiv.2006.11239>
- [14] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, ... and D. Rueckert, “Attention U-Net: Learning Where to Look for the Pancreas,” in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Granada, Spain, pp. 423-431, 2018. <https://doi.org/10.48550/arXiv.1804.03999>
- [15] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, “Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation,” in *Proceedings of the European Conference on Computer Vision*, Tel Aviv: Israel, pp. 205-218, 2022. [https://doi.org/10.1007/978-3-031-25066-8\\_9](https://doi.org/10.1007/978-3-031-25066-8_9)



전지현 (Jihyeon Jeon)

2025년 : 전남대학교 컴퓨터정보통신 공학과 (학사)

2025년~현 재: 전남대학교 지능전자컴퓨터공학과 석사과정  
※ 관심분야 : 컴퓨터 비전, 딥러닝 등



이성훈 (Sunghoon Lee)

2025년 : 전남대학교 컴퓨터정보통신 공학과 (학사)

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



박수형 (Suhung Park)

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

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