

Dual Encoder와 DDPM을 활용한 위성 영상 내 건물 변화 탐지 성능 향상

류현승¹ · 전지현¹ · 박수형^{2*}

¹전남대학교 컴퓨터정보통신공학과 학사과정

²전남대학교 지능전자컴퓨터공학부 교수

Improved Performance of Building Change Detection in Satellite Images Using Dual Encoder and Denoising Diffusion Probabilistic Models

Hyunseung Ryu¹ · Jihyeon Jeon¹ · Suhyung Park^{2*}

¹Bachelor's Course, Department of Computer Engineering, Chonnam National University, Gwangju 61186, Korea

²Professor, Department of Intelligent Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea

[요약]

본 연구는 위성영상 기반 건물 변화 탐지의 성능을 개선하기 위해 Dual Encoder와 DDPM(Denoising Diffusion Probabilistic Models)을 결합한 새로운 방법을 제안한다. 기존의 변화 탐지 방식이 가지는 경계 불명확성과 환경 변화에 대한 민감성 문제를 해결하기 위해, Dual Encoder를 통해 시기별 특징을 독립적으로 추출하고 Feature Fusion을 통해 통합하는 구조를 도입했다. 또한 DDPM의 점진적 노이즈 제거 과정을 활용하여 더욱 정교한 Segmentation Map 생성이 가능하도록 했다. LEVIR-CD 데이터셋을 활용한 실험을 통해 기존 방식 대비 현저한 성능 향상을 확인했으며, 특히 복잡한 도시 환경에서의 작은 규모 변화 탐지에서 뛰어난 성능을 보였다. 제안된 방법은 도시 계획, 재난 평가, 불법 건축물 감시 등 다양한 실제 응용 분야에서 높은 활용 가능성을 보여주며, 자동화된 건물 변화 탐지의 신뢰성과 정확도를 크게 향상시켰다.

[Abstract]

This study proposes a novel approach combining dual encoder and denoising diffusion probabilistic models (DDPM) to enhance building change detection in satellite imagery. Traditional change detection methods face challenges in terms of boundary ambiguity and sensitivity to environmental changes. To address these limitations, we introduce a dual encoder architecture that independently extracts temporal features from bitemporal images and integrates them via feature fusion. We utilize the gradual denoising process of DDPM to generate more precise segmentation maps, focusing particularly on complex urban environments. Experiments on the LEVIR-CD dataset demonstrate significant performance improvements over existing methods, especially in the detection of small-scale changes in complex urban environments. The proposed method shows great potential for practical applications in urban planning, disaster assessment, and illegal construction monitoring, and it offers a more reliable and accurate approach to automated building change detection.

색인어 : 위성영상, 건물 변화 탐지, 딥러닝, 영상 분할, 생성 모델

Keyword : Satellite Imagery, Building Change Detection, Deep Learning, Image Segmentation, Generative Model

<http://dx.doi.org/10.9728/dcs.2025.26.2.489>



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 18 January 2025; **Revised** 19 February 2025

Accepted 24 February 2025

***Corresponding Author; Suhyung Park**

Tel: +82-62-530-1797

E-mail: suhyung@jnu.ac.kr

1. 서론

1-1 연구배경

위성영상은 광범위한 지역에 대한 관측을 가능하게 하여 도시 계획, 재난 평가, 환경 조사 등 다양한 분야에서 활용되고 있다[1]. 특히 고해상도 위성의 발달로 건물 단위의 변화까지 관측할 수 있게 되면서 위성영상을 활용한 건물 변화 탐지(Building Change Detection)가 중요한 연구 주제로 부상하고 있다. 건물 변화 탐지는 서로 다른 시점에 촬영된 위성 영상들을 비교하여 건물의 신축, 철거, 증축 등의 변화를 자동으로 식별하는 기술로, 도시 발전의 모니터링과 계획 수립에 핵심적인 역할을 한다.

위성영상 기반의 건물 변화 탐지는 여러 가지 도전적인 과제를 안고 있다. 위성영상은 촬영 고도가 높아 건물이 영상에서 차지하는 비중이 매우 작고 촬영 각도나 기상 조건에 따라 같은 건물도 다르게 보일 수 있다[2]. 특히 딥러닝을 이용한 건물 변화 탐지 시에는 Bi-Temporal 영상 간의 기후 변화, 전처리 보정 등으로 인한 도메인 차이가 존재한다[3]. 이러한 차이는 관심 대상(예: 건물)과 배경에 서로 다른 정도로 영향을 미치며 Siamese 네트워크 사용과 함께 “관심 있는 변화”를 모델이 이해하는 것을 모호하게 만든다[4]. 또한 계절적 변화로 인한 식생의 차이, 그림자의 변화 등 실제 건물의 변화와 무관한 요소들이 변화 탐지를 더욱 어렵게 만든다.

1-2 관련 연구 분석

위성영상 기반 건물 변화 탐지 연구는 크게 두 가지 방향으로 발전해왔다. 첫째, CNN 기반의 Siamese 네트워크를 활용한 접근법이다[8]. 이 방식은 두 시기의 이미지를 동일한 네트워크로 처리하여 특징을 추출하지만, 시기별 특성을 독립적으로 고려하지 못하는 한계가 있다.

둘째, 트랜스포머 기반의 접근법 중 좋은 성능을 보이는 모델은 BAN(ChangeFormer)[17]과 RSP-Swin-T[18]가 대표적이다. 이들은 self-attention 메커니즘을 통해 장거리 의존성을 효과적으로 포착할 수 있지만 작은 객체의 변화를 놓치는 경향이 있다.

U-Net은 의미론적 세그멘테이션 분야에서 가장 널리 사용되는 딥러닝 아키텍처로 인코더-디코더 구조를 통해 이미지의 특징을 추출하고 이를 바탕으로 세그멘테이션 맵을 생성한다[14]. 인코더는 합성곱 연산을 통해 이미지의 특징을 점진적으로 추출하며 Skip Connection을 통해 이 특징들을 디코더로 전달한다. 그러나 U-Net은 복잡한 세분화 작업에서 개념적인 한계를 보이는데 특히 모호한 시각적 특징이나 낮은 대비를 가진 영역을 처리할 때 문제가 된다[5]. 이는 U-Net이 제한된 수용영역으로 인해 충분한 문맥 정보를 캡처하지 못하기 때문이다.

더욱이 U-Net 디코더의 업샘플링(Up-Sampling) 과정에서는 고수준의 의미론적 표현에서 손실된 공간 정보를 복구하는데 한계가 있다[6]. 특히 저해상도의 특징 맵을 순차적으로 업샘플링하여 원본 크기로 확대하는 과정에서 세그멘테이션 결과의 객체 경계가 부정확하고 뭉뚱해지는 문제가 발생한다. 이는 건물의 정확한 형태와 경계를 파악해야 하는 건물 변화 탐지 작업에서 심각한 성능 저하를 초래할 수 있다.

1-3 연구 목표 및 접근 방법

본 연구에서는 DDPM(Denoising Diffusion Probabilistic Models)을 활용한 새로운 접근 방식을 제안한다[7]. DDPM은 점진적인 노이즈 제거 과정을 통해 고품질의 이미지를 생성할 수 있는 생성 모델이다. 기존의 생성 모델들과 비교했을 때 DDPM은 단순하면서도 높은 품질의 샘플을 생성할 수 있으며 특히 세부 디테일의 표현에서 뛰어난 성능을 보인다.

이를 통해 본 연구는 기존 변화 탐지의 한계점들을 극복하고 더 효율적이고 정확한 건물 변화 탐지 시스템을 구현하고자 한다. 특히 넓은 범위의 문맥 정보를 효과적으로 활용하고 Bi-Temporal 특징들 간의 상호작용을 정교하게 모델링하며 레이블링의 효율성을 높이는 데 초점을 맞추었다.

II. 연구 방법

2-1 모델 구조

본 연구의 핵심은 Dual Encoder 구조와 DDPM(Denoising Diffusion Probabilistic Models) 생성 모델을 결합하여 고품질의 Segmentation Map을 생성하는 새로운 방법을 개발하는 것이다. 이 접근법은 기존 Up-Sampling이 가지는 경계 불명확성 문제를 해결하고자 하는 목적으로 고안되었다(그림 1). 특히 본 연구의 목표인 작은 규모의 변화도 놓치지 않아야 한다는 점에서 이 접근 방식은 큰 의미를 가진다.

1) Dual Encoder

위성영상에서의 두 시기의 특징을 독립적으로 추출하기 위해 Siamese 네트워크 구조를 채택한다. 각 시점의 이미지가 독립적인 인코더를 통과하면서 각각의 고유한 특성을 보존하면서도 시간에 따른 변화를 효과적으로 포착할 수 있다. 이는 시기에 따른 밝기 변화나 구조적 차이가 존재할 수 있는 위성 이미지에서 특히 중요하다. 예를 들어, 계절의 변화, 기상 조건의 차이, 또는 시간대에 따른 그림자 변화 등으로 인해 같은 지역이라도 다른 시기의 이미지는 상당히 다른 특성을 보일 수 있다.

각각의 인코더는 입력 이미지로부터 다중 스케일의 특징을 추출하며 이 과정에서 ResNet 기반의 Backbone 네트워크를

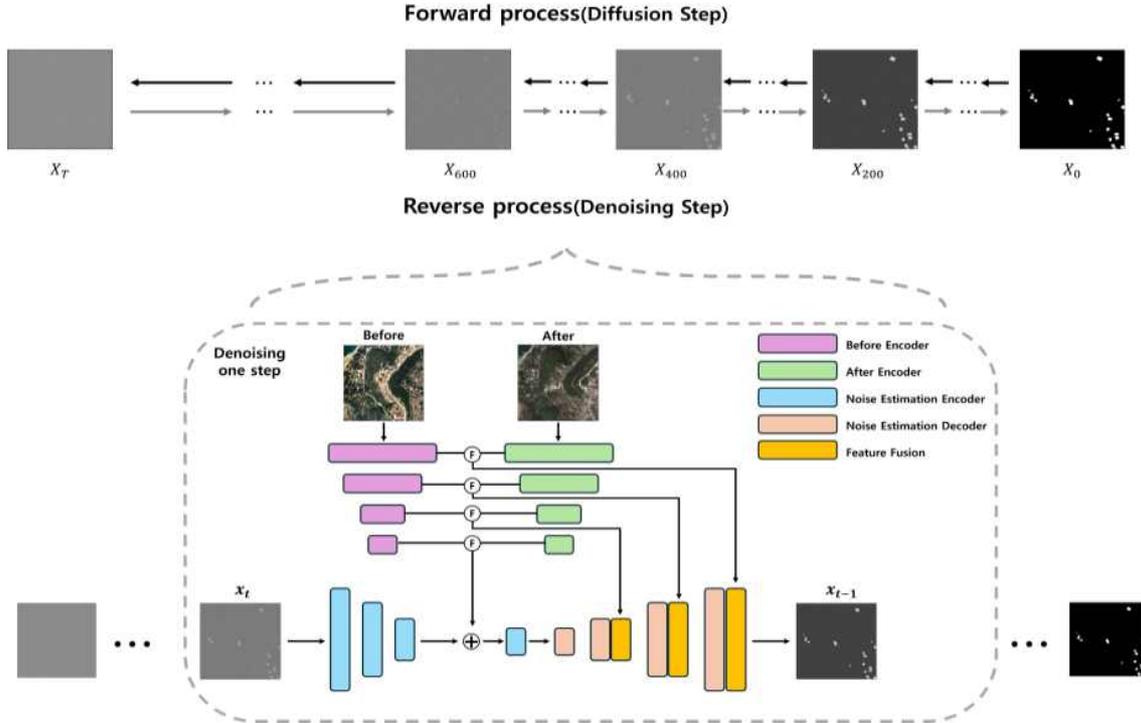


그림 1. Dual Encoder와 DDPM을 결합한 전체 모델 구조. Dual Encoder에서 추출된 Feature들은 Feature Fusion을 통해 DDPM의 Condition으로 동작

Fig. 1. Overall model structure combining Dual Encoder and DDPM. Features extracted from the Dual Encoder operate as conditions for DDPM through Feature Fusion

사용한다. 각 단계별로 추출되는 특징들은 Skip Connection을 통해 이후 단계로 전달되어 다양한 수준의 정보가 보존될 수 있도록 한다. 이러한 시기에 따른 특징의 독립적 추출을 통해 실제 건물 변화와 환경적 변화를 더 정확히 구분할 수 있다.

또한, Dual Encoder 구조는 폭설, 폭우 등의 계절적 요인으로 인한 경계선 추출 방해 문제를 완화하는 데에도 효과적이다. 단일 인코더 방식에서는 변화 이전(Before)과 변화 이후(After) 이미지를 단순히 하나의 입력으로 결합하여 처리하기 때문에 강설이나 강우로 인해 발생하는 전역적인 픽셀 값 변화가 모델이 학습하는 특징에 혼선을 줄 수 있다. 반면, Dual Encoder 구조는 각 시점의 이미지를 개별적으로 인코딩하여 기후 변화로 인한 노이즈와 실제 건물 변화의 차이를 구분하는 능력을 향상시킨다. 특히, 계절적 변화에 강한 특징 학습을 위해 각 인코더는 다중 스케일 정보를 활용하여 저 해상도에서는 적설, 강수로 인한 반사율 변화 등의 전역적인 환경 변화를 억제하고 고해상도에서는 미세한 건물 경계 정보를 강조하는 방식으로 동작하여 눈 덮인 도로, 비로 인해 어두워진 건물 외벽 등 계절적 요인이 포함된 영역에서도 정확한 변화 탐지가 가능하도록 한다.

2) Feature Fusion

Dual Encoder에서 추출된 두 개의 Feature Map을

CBAM(Convolutional Block Attention Module) 형식의 Self-Attention 메커니즘을 통해 통합한다. CBAM은 채널 Attention과 공간 Attention을 순차적으로 적용하여 중요한 특징을 강조하고 불필요한 정보를 억제한다[12].

CBAM의 채널 Attention 모듈은 Feature Map의 채널 간 관계를 활용하여 '무엇'에 주목할지를 결정한다. 이를 위해 Average-Pooling과 Max-Pooling을 모두 활용하여 공간 정보를 압축하고 이를 Shared Network를 통해 처리하여 채널별 가중치를 생성한다. 공간 Attention 모듈은 '어디'에 주목할지를 결정하는데 채널 축을 따라 Pooling 연산을 수행하여 공간적 Attention Map을 생성한다.

3) DDPM 생성 모델

DDPM은 점진적인 노이즈 제거 과정을 통해 고품질의 Segmentation Map을 생성하는 생성 모델이다. 모델은 1000 개의 타임스텝 동안 이미지에 점진적으로 노이즈를 추가하는 Forward Process와 이를 제거하는 Reverse Process로 구성된다. Forward Process에서는 각 단계마다 작은 양의 가우시안 노이즈가 이미지에 추가되며 Reverse Process에서는 학습된 모델이 이 노이즈를 제거하는 방향으로 동작한다.

이 과정은 조건부 확률로 표현되며 여기서 사용되는 β_t 는 시간에 따른 노이즈의 크기를 조절하는 파라미터이다(식 1). 이러한 Forward Process의 특별한 점은 임의의 시점 t 에서

의 x_t 를 한 번에 계산할 수 있다는 것이다. Reparameterization Trick을 사용하여 표현할 수 있으며(식 2), 이를 통해 학습 과정을 효율적으로 만들 수 있다.

본 연구의 DDPM 구조는 U-Net을 기반으로 하며, 특히 Residual Block과 Attention Layer를 포함하는 개선된 구조를 사용한다. 각 Residual Block은 Group Normalization, SiLU Activation, 그리고 2D Convolution Layer로 구성되며 시간 정보(Time Embedding)가 첫 번째 Convolution Block의 출력에 더해지는 구조를 가진다. 특히 16×16 과 8×8 Resolution에서는 Residual Block 다음에 Attention Layer를 추가하여 더 효과적인 특징 추출이 가능하도록 한다[13].

본 연구에서는 Before Image, After Image를 Condition으로 활용하는 방식을 채택한다. 특히 현재 Step의 Noisy Segmentation 정보를 Condition Encoding에 통합하여 상호보완적인 정보를 활용한다. 이는 Before Image, After Image가 가진 정확한 Target 정보와 현재 Step의 Segmentation Map이 가진 강화된 Target 영역 정보를 결합함으로써 더 정확한 Segmentation을 가능하게 한다[14].

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t) \tag{1}$$

$$x_t = \sqrt{\alpha_t}x_0 + \sqrt{1-\alpha_t}\epsilon, \epsilon \sim N(0, I) \tag{2}$$

4) 모델의 노이즈(ϵ) 추정 과정

Reverse Process는 Forward Process의 반대 과정으로, 노이즈가 포함된 이미지에서 점진적으로 노이즈를 제거하는 과정이다. Wu et al.[15]은 이 과정을 조건부 생성 과정으로 확장하여 Before Image, After Image의 정보를 함께 활용하는 방식을 제안했다(식 3). 이는 변화 탐지 태스크에 특화된 접근 방식으로 시간적 정보를 효과적으로 활용할 수 있게 한다.

여기서 α_t 는 노이즈 스케줄링 파라미터이며 $\bar{\alpha}_t$ 는 누적 곱을 나타낸다. 본 연구에서는 Linear Noise Schedule을 사용하며 β_1 부터 β_T 까지 선형적으로 증가하는 값을 사용한다. 이는 초기에는 작은 노이즈를 제거하고 점차 큰 노이즈를 제거하는 방식으로 안정적인 학습을 가능하게 한다.

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}}\epsilon_\theta(x_t, Before, After, t)) \tag{3}$$

2-2 데이터셋

본 연구에서는 LEVIR-CD 데이터셋을 활용하여 모델의 훈련 및 테스트를 수행하였다[16]. LEVIR-CD는 2~4년 간격으로 촬영된 고해상도(1024×1024) 위성 이미지 637쌍으로 구성되어 있으며, 각 쌍은 변화 이전(Before) 이미지, 변화 이후(After) 이미지, 그리고 변화 영역을 나타내는 레이블로 구성된다(그림 2).



그림 2. Before, After, Label 쌍으로 이루어진 LEVIR-CD Dataset

Fig. 2. LEVIR-CD Dataset consisting of Before, After, and Label pairs

데이터셋의 보정 과정은 다음과 같은 단계로 수행되었다. 먼저, 원격 감지 이미지 해석 전문가들이 상세한 주석 처리 사양에 따라 일관된 보정을 수행하였다. 각 이미지 쌍은 한 명의 전문가가 1차 보정을 수행하고, 다른 전문가가 이중 확인하는 방식으로 진행되어 높은 품질의 주석을 확보하였다. 전체 데이터셋의 보정에는 약 120인일(Person-days)이 소요되었으며, 각 이미지 쌍은 평균 50개의 변화 건물을 포함하고 있다. 데이터셋의 특징적인 점은 실제 위성 영상에서 발생할 수 있는 정합(registration) 오차를 포함하고 있다는 것이다. 이는 완벽한 정합을 목표로 하기보다는 실제 응용 환경에서 마주할 수 있는 현실적인 문제들을 반영한 것으로, 더욱 실용적인 변화 탐지 알고리즘 개발을 가능하게 한다.

각 이미지는 RGB 채널의 고해상도 항공 사진으로, 0.5 m/pixel의 상세한 지상 정보를 제공하며 1024×1024 픽셀

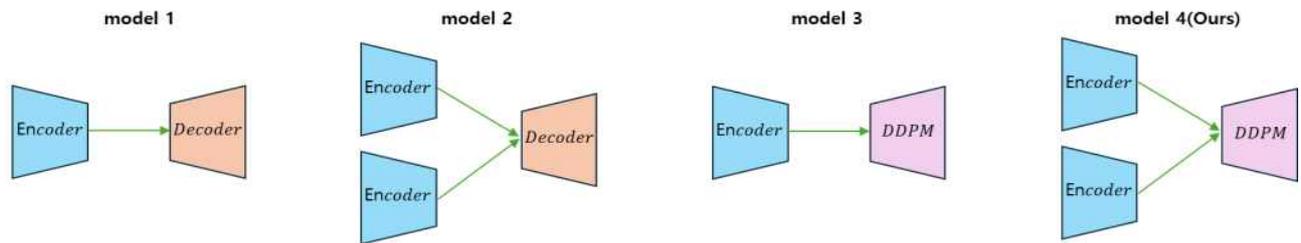


그림 3. 비교 모델 구조(모델 1: Single Encoder U-Net / 모델 2: Dual Encoder U-Net / 모델 3: Single Encoder DDPM / 모델4 : Dual Encoder DDPM(ours))

Fig. 3. Comparison of model architectures (Model 1: Single Encoder U-net / Model 2: Dual Encoder U-net / Model 3: Single Encoder DDPM / Model 4: Dual Encoder DDPM (ours))

크기로 구성되어 있다. 레이블은 변화된 영역을 나타내는 이진 마스크(변화 있음: 1, 변화 없음: 0)로 제공된다.

데이터셋은 Train, Validation, Test Set을 7:1:2의 비율로 분할하여, 각각 446쌍(70%), 64쌍(10%), 127쌍(20%)으로 구성하였다. 데이터셋에는 구조물의 신축, 철거, 확장 등 다양한 변화 유형과 계절 변화, 그림자, 조도 차이 등 환경적 변화가 포함되어 있어, 실제 변화 탐지 상황에서 발생할 수 있는 복잡하고 세밀한 변화 패턴을 효과적으로 시뮬레이션할 수 있다.

2-3 실험 설계

1) 실험1: Single Encoder와 Dual Encoder 비교

첫 번째 실험은 제안하는 Dual Encoder 구조의 효과성을 검증하기 위해 Single Encoder(Model 1, 3)와 Dual Encoder(Model 2, 4) 구조를 비교한다(그림 3, 표 1). Single Encoder는 이전/이후 이미지를 하나의 입력으로 연결(Concatenation)하여 단일 경로로 특징을 추출하는 반면, 제안하는 Dual Encoder는 각 시기의 이미지에서 특징을 독립적으로 추출한 후 Feature Fusion을 통해 통합한다. 이러한 구조적 차이가 시간적 특징 추출에 미치는 영향을 분석하기 위해 기본 U-Net 구조를 Single Encoder의 Baseline으로 사용한다.

이 실험을 통해 시기별 독립적 특징 추출이 환경 변화와 구조물 변화의 구분에 미치는 영향을 분석한다.

표 1. 각 구성요소(Dual Encoder, DDPM) 평가를 위한 실험과 Hyperparameters 설계. Single Encoder는 Before와 After가 Concatenation 되어 입력

Table 1. Experimental design to evaluate each component (Dual Encoder, DDPM). Single Encoder receives Before and After images as Concatenated input

	Optimizer	Epochs	Batch size	Learning rate
Model 1	Adam	100	6	0.0001
Model 2	Adam	100	6	0.0001
Model 3	Adam	1,800	8	0.00005
Model 4 (Ours)	Adam	1,800	8	0.00005

2) 실험2: U-Net과 DDPM 비교

두 번째 실험은 기존 U-Net의 Up-sampling이 가지는 한계를 DDPM이 얼마나 효과적으로 개선하는지 평가하기 위해 U-Net(Model 1, 2)과 DDPM(Model 3, 4) 구조를 비교한다. U-Net의 Up-sampling은 저해상도 특징 맵을 순차적으로 확대하는 과정에서 객체의 경계가 부정확해지는 한계가 있다. 반면 DDPM은 점진적 노이즈 제거를 통해 세부 특징을 복원하는 방식으로 보다 선명한 경계를 가진 Segmentation Map 생성이 가능하다. 이 실험을 통해 U-Net이 가지는 경계 불명확성 문제를 DDPM이 얼마나 개선하는지 그리고 이러한 개선이 실제 변화 탐지 성능에 미치는 영향을 분석한다.

2-4 평가지표

본 연구에서는 제안된 모델의 성능 평가를 위해 OA (Overall Accuracy), IoU(Intersection over Union), F1 Score를 평가 지표로 사용한다. 구조물 변화 탐지는 각 픽셀이 변화 있음 (1) 또는 변화 없음 (0)으로 분류되는 이진 분류 문제이다. 변화 탐지에서는 위양성(False Positive)과 위음성(False Negative)이 분석 정확도에 영향을 미치므로 각각적인 성능 평가가 요구된다.

OA(Overall Accuracy)(식 4)는 전체 픽셀 중 올바르게 분류된 픽셀의 비율을 나타낸다. True Positive, True Negative, False Positive, False Negative를 기반으로 계산되며 계산이 단순하고 직관적이다. 그러나 변화가 적은 지역에서 대부분의 픽셀을 '변화 없음'으로 분류하는 경우에도 높은 정확도를 나타낼 수 있어 단독 지표로는 한계가 있다.

IoU(Intersection over Union)(식 5)는 예측된 변화 영역과 실제 변화 영역의 교집합을 합집합으로 나눈 값으로, 변화 영역의 위치와 형태를 평가한다. 구조적 변화에서 경계 탐지의 정확성을 측정하는 지표로 활용된다. 단, 변화 영역의 크기가 극단적인 경우 전체 성능 반영에 제한이 있다[17].

F1 score(식 6)는 정밀도(Precision)와 재현율(Recall)의 조화 평균이며 클래스 불균형 문제에서 안정적인 평가가 가능하다. 변화 영역의 크기와 관계없이 위양성과 위음성을 고려한 평가가 가능하여 정밀한 변화 탐지가 요구되는 상황에서 주요 평가 지표로 사용된다[18].

이러한 평가 지표들을 통해 Dual Branch 구조와 DDPM 도입으로 인한 성능 변화를 분석하고 각 구성 요소가 모델 성능에 미치는 영향을 검증한다.

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$IoU = \frac{|Prediction \cap Ground Truth|}{|Prediction \cup Ground Truth|} \quad (5)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (6)$$

III. 연구 결과

3-1 정량적 분석 결과

1) 실험1: Single Encoder와 Dual Encoder 비교 결과

Single Encoder 기반 모델(Single Encoder U-Net: 86.33%, Single Encoder DDPM: 91.63%)의 평균 F1 Score는 88.98%인 반면, Dual Encoder 기반 모델(Dual Encoder U-Net: 88.97%, Dual Encoder DDPM: 94.38%)은 91.67%의 F1 Score를 나타내었다(표 2, 그림 4). 이러한 성능 차이

는 Dual Encoder가 각 시점 이미지의 특성을 독립적으로 추출하고 Feature Fusion을 통해 통합하는 구조에 기인한다. 이는 환경적 변화와 구조적 변화를 효과적으로 구분할 수 있는 능력을 강화하였다.

표 2. Dual Branch, DDPM 각 구조의 효과를 정량적으로 평가하기 위한 실험 결과표

Table 2. Experimental results table for quantitative evaluation of the effects of Dual Branch and DDPM architectures

	OA	IoU	F1
Model 1	98.68 ± 01.18	76.27 ± 10.69	86.33 ± 06.59
Model 2	98.80 ± 01.04	77.25 ± 10.39	88.97 ± 06.41
Model 3	99.44 ± 00.05	86.44 ± 02.04	91.63 ± 01.15
Model 4 (Ours)	99.45 ± 00.05	89.26 ± 02.16	94.38 ± 01.20

2) 실험2: U-Net과 DDPM 비교 결과

U-Net 모델(Single Encoder U-Net: 86.33%, Dual Encoder U-Net: 88.97%)의 평균 F1 Score는 87.65%인 반면, DDPM 기반 모델(Single Encoder DDPM: 91.63%, Dual Encoder DDPM: 94.38%)은 93.00%의 F1 Score를 기록하였다(표 2, 그림 4). 이 차이는 DDPM의 점진적 노이즈 제거 과정과 확률적 생성 모델의 특성에 기인하여 다중 스케일 특징의 복원과 경계 영역의 세분화에서 우수한 성능을 보이는 것으로 나타났다.

3-2 정성적 분석 결과

본 연구의 정성적 분석에서는 제안된 모델들의 성능을 Ground Truth와 비교하여 주요 차이점을 도출하였다(그림 5).

1) 실험1: Single Encoder와 Dual Encoder 비교 결과

Single Encoder 기반 모델과 Dual Encoder 기반 모델을 Ground Truth와 비교한 결과, Dual Encoder 기반 모델이

복잡한 형태의 구조물 변화 탐지에서 높은 정확도를 보였다.

특히 Image 1과 5에서 볼 수 있듯이, 불규칙한 형태의 건물과 군집 구조에서 Dual Encoder는 경계를 더 정확하게 포착하였다.

Ground Truth와의 비교를 통해, Dual Encoder가 시간적 특징을 효과적으로 추출하고 정보 통합을 통해 구조적 변화를 보다 정확하게 인식함으로써 오탐지율을 줄이는 데 기여하고 있음을 확인하였다. 또한 Image 3과 4에서 확인할 수 있듯이, 계절 변화나 그림자와 같은 환경적 변화가 있는 경우에도 Dual Encoder는 실제 건물 변화만을 효과적으로 탐지하는 능력을 보여주었다. 이는 각 시점의 이미지에서 독립적으로 특징을 추출하는 구조가 환경 변화에 의한 영향을 최소화하는데 기여했음을 시사한다.

2) 실험2: U-Net과 DDPM 비교 결과

U-Net은 Ground Truth와 비교했을 때 객체 경계가 불분명하게 나타나는 경향을 보였다. Image 2와 4에서 관찰되듯이, 이는 경계선이 흐릿하게 표현되거나 중요한 구조적 변화가 명확하게 드러나지 않는 문제를 초래하였다. 특히 건물의 복잡한 외곽선이나 작은 규모의 변화를 포착하는 데 있어 한계를 보였다. 반면, DDPM은 Ground Truth와의 비교에서 경계선 생성 및 구조적 변화 탐지에 있어 높은 정확도를 나타내었다. Image 5에서 볼 수 있듯이, DDPM의 이미지 생성 과정이 U-Net에 비해 세부적인 경계 영역을 보다 효과적으로 처리함으로써 Ground Truth와의 일치도가 향상된 결과를 가져왔다. 이는 DDPM의 점진적 노이즈 제거 과정이 건물의 세부 구조와 경계를 더 정교하게 복원할 수 있게 해주었기 때문으로 분석된다. 또한 Image 1과 3에서 확인할 수 있듯이, 넓은 면적의 변화나 복잡한 건물 구조에서도 DDPM은 안정적인 탐지 성능을 보여주었다.

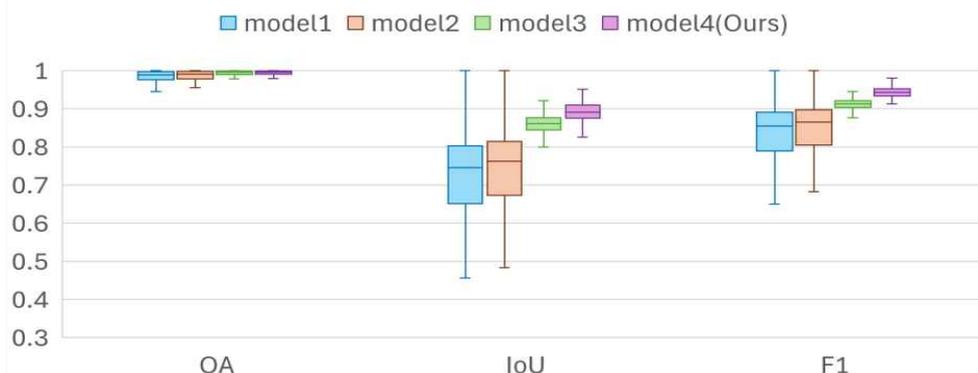


그림 4. 모델 구조에 따른 성능 지표(OA, IoU, F1) 비교. Dual Encoder와 DDPM 기반 모델이 다른 모델 대비 우수한 성능을 보임. Fig. 4. Performance metric (OA, IoU, F1) comparison graph by model architectures. Models based on Dual Encoder and DDPM demonstrate superior performance

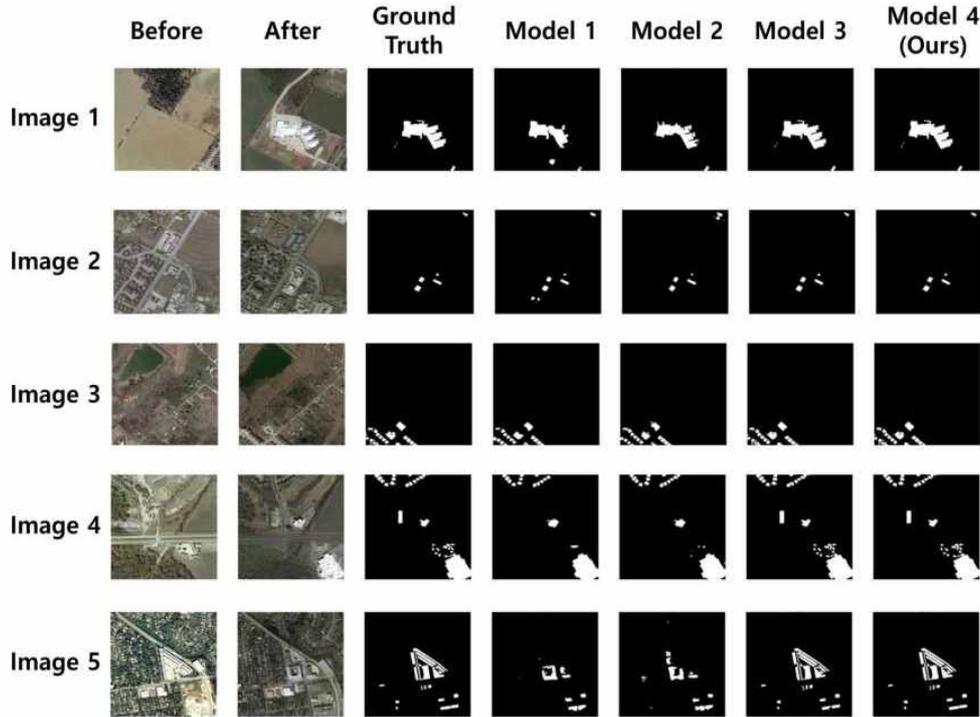


그림 5. 건물 변화 탐지 모델별 결과에 대한 정성적 비교 결과. Dual Encoder-DDPM 모델이 Ground Truth와 가장 유사한 결과를 보임.

Fig. 5. Qualitative comparison results of building change detection models. The Dual Encoder-DDPM model yields results most similar to the Ground Truth.

3-3 기존 연구와의 비교

최신 트랜스포머 기반 변화 탐지 모델들과의 비교를 통해 제안 모델의 성능을 분석하였다(그림 6, 표 3). BAN(ChangeFormer)와 RSP-Swin-T 모델은 각각 다른 접근 방식으로 변화 탐지 문제를 해결하고자 하였으며, 두 모델 모두 90% 이상의 F1 Score를 달성하며 우수한 성능을 보였다.

본 연구에서 제안하는 Dual Encoder DDPM 모델은 이러한 최신 모델들과 비교하여 모든 평가 지표에서 향상된 성능을 보여주었다. 특히 IoU 지표에서 89.26%(±2.16)를 기록하여 BAN의 85.20%(±6.83)와 RSP-Swin-T의 85.94%(±2.04) 대비 유의미한 성능 향상을 달성하였다. 이는 제안된 모델이 건물 변화의 경계를 더욱 정확하게 포착할 수 있음을 시사한다.



그림 6. 트랜스포머 기반 변화 탐지 모델들의 성능 지표(OA, IoU, F1) 비교를 통해 제안 모델의 우수한 정확도와 안정성을 보임.

Fig. 6. Comparison of transformer-based change detection models' performance metrics (OA, IoU, F1) demonstrating our model's superior accuracy and stability

표 3. 트랜스포머 기반 변화 탐지 모델들의 지표별 평균 및 표준 편차를 나타낸 정량적 비교 결과

Table 3. Quantitative comparison of transformer-based change detection models showing mean and standard deviation for each metric

	OA	IoU	F1
BAN (ChangeFormer)	99.24 ± 00.67	85.20 ± 06.83	91.13 ± 03.98
RSP-Swin-T	99.44 ± 00.55	85.94 ± 02.04	91.86 ± 01.26
Ours	99.45 ± 00.05	89.26 ± 02.16	94.38 ± 01.20

또한 F1 Score에서도 94.38%(±1.20)를 기록하여 기존 모델들(BAN:91.13%±3.98, RSP-Swin-T:91.86%±1.26)을 상회하는 결과를 보였다. 특히 주목할 만한 점은 모든 평가 지표에서 제안 모델의 표준편차가 일관되게 낮게 나타났다는 것이다. 이는 Dual Encoder와 DDPM의 결합이 변화 탐지의 정확성뿐만 아니라 안정성 측면에서도 개선 효과가 있음을 입증한다.

정성적 비교 결과(그림 7)에서도 제안 모델의 우수성이 명확하게 드러났다. Image 1의 경우, 기존 모델들이 건물의 일부 영역을 놓치거나 과대 추정하는 경향을 보인 반면, 제안 모델은 Ground Truth와 가장 유사한 경계를 포착하였다. Image 2와 Image3에서는 소규모 건물들의 변화를 감지하는데 있어 제안 모델이 더 정확한 성능을 보였으며 특히 Image 3의 경우 복잡한 건물 구조에서도 안정적인 탐지 결과를 보여주었다. Image 4의 경우에는 환경적 변화(그림자, 식생)가 존재하는 상황에서도 실제 건물 변화만을 효과적으로 구분해내는 것을 확인할 수

있다. 이러한 결과들은 제안 모델이 다양한 환경 조건과 건물 구조에서 일관되게 우수한 성능을 유지함을 보여준다.

IV. 결 론

본 연구에서는 위성영상 기반 건물 변화 탐지를 위해 Dual Encoder 구조와 DDPM 생성모델을 결합한 새로운 방법론을 제안하였다. 이는 기존 변화 탐지 방식의 주요 한계점이었던 경계 불명확성 문제와 환경 변화에 대한 민감성을 효과적으로 개선하였다는 점에서 중요한 의미를 가진다.

제안된 모델의 가장 큰 의의는 건물 변화 탐지의 정확성과 신뢰성을 동시에 향상시켰다는 점이다. 특히 Dual Encoder 구조는 계절 변화나 그림자와 같은 환경적 변화와 실제 건물의 구조적 변화를 효과적으로 구분함으로써 오탐지율을 크게 낮출 수 있었다. 이는 도시 계획이나 불법 건축물 감시와 같은 실제 응용 분야에서 매우 중요한 의미를 가진다.

DDPM의 도입은 건물 변화 탐지에 있어 새로운 패러다임을 제시했다고 볼 수 있다. 기존의 직접적인 특징 맵 변환 방식에서 벗어나 확률적 생성 모델을 통한 접근은 더욱 정교한 변화 탐지를 가능하게 했다. 이는 특히 복잡한 도시 환경에서 작은 규모의 건물 변화나 부분적 변형을 포착하는 데 큰 강점을 보였다.

그러나 본 연구는 몇 가지 중요한 한계점을 가지고 있다. 가장 큰 문제는 추론 시간이 60초 이상 소요되어 실시간 모니터링이 어렵다는 점이다. 이를 해결하기 위해 DDPM 기반 모델을

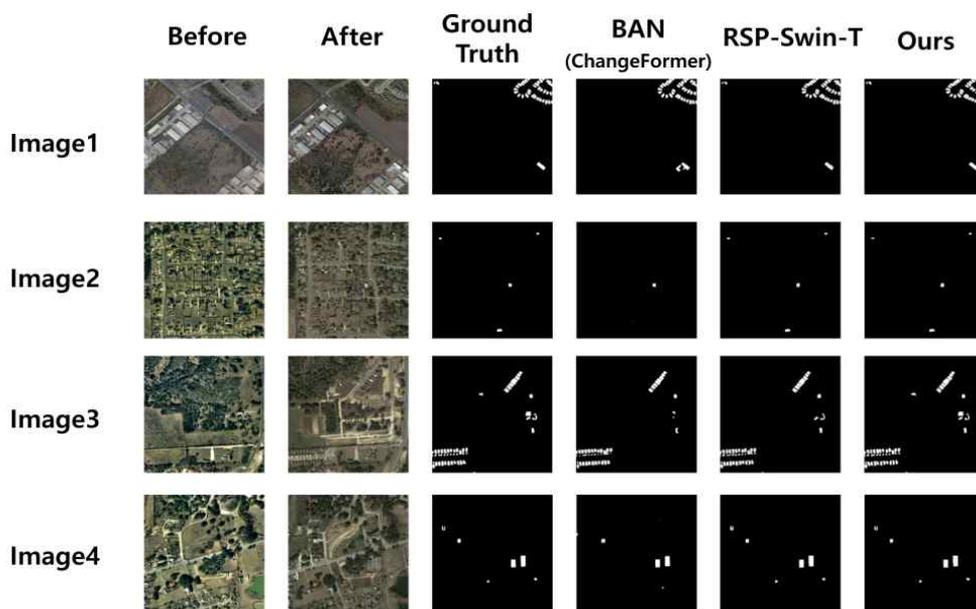


그림 7. 트랜스포머 기반 기존 연구 모델들과의 정성적 비교 결과. 제안 모델이 기존 모델들 대비 Ground Truth와 가장 유사한 변화 탐지 결과를 보임

Fig. 7. Qualitative comparison results of transformer-based building change detection models. Our proposed model demonstrates detection results most similar to the Ground Truth compared to existing models

표 4. DDPM과 Stable Diffusion 모델의 평균 성능 지표 및 평균 추론 시간 비교 분석

Table 4. Comparative analysis of average performance metrics and average inference time between DDPM and Stable Diffusion models

	OA	IoU	F1	Inference Time(sec)
Dual Encoder DDPM(Ours)	99.45	89.26	94.38	68.25
Dual Encoder Stable Diffusion	97.84	60.48	77.95	0.23

Stable Diffusion Model로 대체한 결과, 추론 시간이 0.2초대로 단축되어 실시간 모니터링의 가능성을 확인할 수 있었다(표 4). 다만 Overall Accuracy, IoU, F1 score 등 전반적인 성능 지표가 하락하는 trade-off가 발생했다. Stable Diffusion Model에서는 latent space에서 diffusion-denoising을 수행하고 단 1회의 denoising 과정으로 결과를 생성하는 방식을 통해 이러한 속도 향상이 가능했다[19].

이러한 한계를 극복하기 위한 향후 연구 방향으로는 첫째, Knowledge Distillation과 모델 경량화 기법을 활용하여 Stable Diffusion Model의 빠른 추론 속도를 유지하면서도 성능 저하를 최소화하는 방안을 모색할 수 있다. 이는 모델 구조 최적화와 학습 전략 개선을 통해 실현 가능할 것으로 기대된다. 둘째, Few-shot Learning 기법을 도입하여 산업 단지나 농촌 지역 등 다양한 환경에서의 모델 적응성을 향상시키는 연구가 필요하다.

더 나아가, 본 연구의 방법론은 건물 변화 탐지를 넘어 다양한 분야로 확장될 수 있다. 예를 들어, 재난 상황에서의 피해 평가, 무단 채굴 감시, 산림 훼손 모니터링 등에도 적용될 수 있을 것이다. 특히 Dual Encoder의 특징 추출 능력과 DDPM의 정교한 생성 능력은 이러한 응용 분야에서도 큰 잠재력을 가진다.

마지막으로, 본 연구는 위성영상 분석에 있어 생성 모델의 활용 가능성을 보여주었다는 점에서 의의가 있다. 향후 더욱 발전된 생성 모델과의 결합을 통해 위성영상 분석의 새로운 지평을 열 수 있을 것으로 기대된다.

감사의 글

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원-지역지능화혁신인재양성사업(IITP-2025-RS-2022-00156287, 33%), 학석사연계 ICT핵심인재양성사업의 연구결과(RS-2022-00156385,33%), 그리고 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(RS-2024-00357917,34%).

참고문헌

- [1] D. Wang, J. Zhang, M. Xu, L. Liu, D. Wang, E. Gao, ... and L. Zhang, "MTP: Advancing Remote Sensing Foundation Model via Multi-Task Pretraining," arXiv:2403.13430, 2024. <https://doi.org/10.48550/arXiv.2403.13430>
- [2] S. Fang, K. Li, and Z. Li, "Changer: Feature Interaction is what You Need for Change Detection," arXiv:2209.08290, September 2022. <https://doi.org/10.48550/arXiv.2209.08290>
- [3] Z. Zheng, A. Ma, L. Zhang, and Y. Zhong, "Change is Everywhere: Single-Temporal Supervised Object Change Detection in Remote Sensing Imagery," in *Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, Canada, pp. 15173-15182, October 2021. <https://doi.org/10.1109/ICCV48922.2021.01491>
- [4] Y. Li, Q. Hou, Z. Zheng, M.-M. Cheng, J. Yang, and X. Li, "Large Selective Kernel Network for Remote Sensing Object Detection," arXiv:2303.09030, March 2023. <https://doi.org/10.48550/arXiv.2303.09030>
- [5] F. Rahman, B. Vasu, J. Van Cor, J. Kerekes, and A. Savakis, "Siamese Network with Multi-Level Features for Patch-Based Change Detection in Satellite Imagery," in *Proceedings of 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Anaheim: CA, pp. 958-962, November 2018. <https://doi.org/10.1109/GlobalSIP.2018.8646512>
- [6] K. Li, X. Cao, and D. Meng, "A New Learning Paradigm for Foundation Model-Based Remote Sensing Change Detection," arXiv:2312.01163v1, December 2023. <https://doi.org/10.48550/arXiv.2312.01163>
- [7] D. Wang, J. Zhang, B. Du, G.-S. Xia, and D. Tao, "An Empirical Study of Remote Sensing Pretraining," arXiv:2204.02825v4, May 2023. <https://doi.org/10.48550/arXiv.2204.02825>
- [8] 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 2015)*, Munich, Germany, pp. 234-241, October 2015. https://doi.org/10.1007/978-3-319-24574-4_28
- [9] V. Loos, R. Pardasani, and N. Awasthi, "Demystifying the Effect of Receptive Field Size in U-Net Models for Medical Image Segmentation," arXiv:2406.16701, June 2024. <https://doi.org/10.48550/arXiv.2406.16701>
- [10] R. Han, X. Fan, and J. Liu, "EUNet: Edge-UNet for Accurate Building Extraction and Edge Emphasis In

Gaofen-7 Images,” *Remote Sensing*, Vol. 16, No. 13, 2397, July 2024. <https://doi.org/10.3390/rs16132397>

- [11] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” in *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS’ 20)*, Vancouver, Canada, pp. 6840-6851, December 2020. <https://doi.org/10.48550/arXiv.2006.11239>
- [12] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in *Proceedings of the 15th European Conference on Computer Vision (ECCV 2018)*, Munich, Germany, pp. 3-19, September 2018. https://doi.org/10.1007/978-3-030-01234-2_1
- [13] J. Wolleb, R. Sandkühler, F. Bieder, P. Valmaggia, and P. C. Cattin, “Diffusion Models for Implicit Image Segmentation Ensembles,” arXiv:2112.03145, December 2021. <https://doi.org/10.48550/arXiv.2112.03145>
- [14] T. Amit, T. Shaharbany, E. Nachmani, and L. Wolf, “SegDiff: Image Segmentation with Diffusion Probabilistic Models,” arXiv:2112.00390v3, September 2022. <https://doi.org/10.48550/arXiv.2112.00390>
- [15] J. Wu, R. Fu, H. Fang, Y. Zhang, Y. Yang, H. Xiong, ... and Y. Xu, “MedSegDiff: Medical Image Segmentation with Diffusion Probabilistic Model,” arXiv:2211.00611v5, January 2023. <https://doi.org/10.48550/arXiv.2211.00611>
- [16] H. Chen and Z. Shi, “A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection,” *Remote Sensing*, Vol. 12, No. 10, 1662, May 2020. <https://doi.org/10.3390/rs12101662>
- [17] M. Berman, A. R. Triki, and M. B. Blaschko, “The Lovász-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks,” in *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City: UT, pp. 4413-4421, June 2018. <https://doi.org/10.1109/CVPR.2018.00464>
- [18] D. Müller, I. Soto-Rey, and F. Kramer, “Towards a Guideline for Evaluation Metrics in Medical Image Segmentation,” *BMC Research Notes*, Vol. 15, No. 1, 210, June 2022. <https://doi.org/10.1186/s13104-022-06096-y>
- [19] T. Lin, Z. Chen, Z. Yan, W. Yu, and F. Zheng, “Stable Diffusion Segmentation for Biomedical Images with Single-Step Reverse Process,” in *Proceedings of the 27th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI 2024)*, Marrakesh, Morocco, pp. 656-666, October 2024. https://doi.org/10.1007/978-3-031-72111-3_62



류현승(Hyunseung Ryu)

2022년 : 전남대학교 입학

2022년~현 재: 전남대학교 컴퓨터정보통신공학과 학사과정
※관심분야: 컴퓨터 비전, 딥러닝 등



전지현(Jihyeon Jeon)

2020년 : 전남대학교 입학

2020년~현 재: 전남대학교 컴퓨터정보통신공학과 학사과정
※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 지능전자컴퓨터공학부 교수

※관심분야: 딥러닝, 영상처리