

일반논문 (Regular Paper)

방송공학회논문지 제31권 제1호, 2026년 1월 (JBE Vol.31, No.1, January 2026)

<https://doi.org/10.5909/JBE.2026.31.1.123>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

명시적 깊이 사전 정보를 활용한 건물 변화 탐지

최혜정^{a)}, 박장수^{a)}, 오승준^{b)}, 심동규^{a)†}

Building Change Detection with Explicit Depth Prior Knowledge

Hye-jeong Choi^{a)}, Jangsoo Park^{a)}, Seoung-Jun Oh^{b)}, and Donggyu Sim^{a)†}

요약

본 논문은 원격탐사 변화 탐지에서 잡음·센서 왜곡에 강인한 성능을 위해 광학 영상과 대응 깊이 정보를 함께 사용하는 Depth-Guided Change Detection Network (DGCD-Net)을 제안한다. 두 입력에서 추출한 특징을 Adaptor 모듈로 융합해 변화 영역을 식별하는 핵심 표현을 구성하고, 이를 기반으로 변화 맵을 생성한다. 또한 Semantic Distortion Filter (SDF)로 깊이 추정 과정의 왜곡·잡음을 보정해 실제 기하 구조를 더 정확히 반영하며, 변화 단서에 집중한 안정적 특징 학습을 유도한다. WHU-CD에서 F1-score 96.23%로 FTA-Net 대비 약 1% 향상된 결과를 보여준다.

Abstract

This paper proposes DGCD-Net (Depth-Guided Change Detection Network, DGCD-Net) for remote-sensing change detection robust to noise and sensor distortions. Optical images and corresponding depth are encoded and fused via an adaptor module to form a discriminative representation and generate a change map. A Semantic Distortion Filter (SDF) corrects noisy depth-estimation artifacts to better reflect scene geometry. On WHU-CD, DGCD-Net achieves 96.23% F1, about 1% higher than FTA-Net.

Keyword : Remote Sensing, Change Detection, Explicit Prior, Depth, Guidance Information

a) 광운대학교 컴퓨터공학과(Department of Computer Engineering, Kwangwoon University)

b) 광운대학교 전자공학과(Department of Electronic Engineering, Kwangwoon University)

† Corresponding Author : 심동규(Donggyu Sim)
E-mail: dgsim@kw.ac.kr
Tel: +82-2-940-6470
ORCID: <https://orcid.org/0000-0002-2794-9932>

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음(IITP-2025-RS-2023-00258639)

※ 이 논문은 정부 (과학기술정보통신부)의 재원으로 정보통신기획평가원·학·석사연계 ICT 핵심인재양성 지원을 받아 수행된 연구임 (IITP-2025-RS-2022-00156215)

· Manuscript November 18, 2025; Revised December 24, 2025; Accepted December 24, 2025.

1. 서론

변화 탐지는 동일 지역의 다중 시기 영상 데이터를 비교하여 변화의 유형, 크기, 위치를 분석하는 기술로, 국토 모니터링, 자연재해 대응, 도시 확장 분석 등 다양한 분야에서 활발히 활용되고 있다^[1]. 그러나 다중 시점에서 획득된 원격 탐사 영상은 촬영 시간대, 그림자, 계절적 차이 등 외부 요인 때문에 이질적인 특성을 보이게 된다. 이러한 특성은 밝기나 색상 차이 및 기하학적 불일치를 유발하여, 변화 영역을 정확하게 탐지하는 데 큰 어려움을 일으킬 수 있다.

Copyright © 2026 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

이런 문제를 해결하기 위하여 최근에는 딥러닝 기술을 적용하여 원격 탐사 영상 분석의 정밀성과 효율성이 크게 향상했다. 이는 딥러닝 기술이 공간적·시간적 특징을 동시에 학습할 수 있어, 고해상도 및 광범위한 지역에서 수집된 대규모 데이터에 유용하게 적용될 수 있기 때문이다. 최근에는 U-Net^[2], ResNet^[3], Transformer^[4] 등과 같은 딥러닝 아키텍처를 적용하여 고해상도 원격탐사 데이터를 활용한 정량적 분석과 광범위한 지역의 변화를 파악할 수 있게 되었다.

변화 탐지 분야에는 다양한 딥러닝 모델이 활용됐다. 초기 연구는 영상 분류 모델을 기반으로 각 시점의 영상을 독립적으로 분류한 후, 두 결과를 비교하여 변화를 추론하는 접근이 시도되었다^[5]. 이후에는 객체 탐지 모델을 활용하여 건물이나 도로 등 주요 객체의 위치를 Bounding Box 형태로 추출한 뒤, 시점 간 객체의 존재 여부나 위치 변화를 분석하는 방법이 제안되었다^[6]. 그러나 변화 탐지의 핵심 목표는 픽셀 단위에서 변화 영역을 정밀하게 분할하는 데 있기 때문에, 최근 연구들은 주로 영상 분할 모델을 기반의 모델 구조를 중심으로 발전하고 있다. 특히 U-Net과 Convolutional Neural Network (CNN)과 같은 구조는 픽셀 단위 예측에 적합하여 널리 활용되고 있으며^[7], ResNet^[3], VGGNet^[8] 등의 백본 (Backbone)을 Siamese 구조와 결합함으로써 시점 간 특징 차이를 효과적으로 학습한다. 더 나아가 Transformer 및 주의 메커니즘 (Attention Mechanism)을 이용하여 더욱 정교한 변화 표현을 학습하는 방향으로 연구가 확장되고 있다^[9-12]. 하지만 기존 연구들은 높은 정확도 향상에도 불구하고, 수용영역 (Receptive Field)의 제한과 장거리 의존성 (Long-Range Dependency) 부족으로 인해 변화 경계가 불완전하거나 내부가 불연속적으로 검출되는 한계가 있다.

최근 변화 탐지 기술은 이러한 문제를 해결하기 위하여 보조 정보를 활용하는 연구들이 진행되고 있다. GCNet^[11]은 백본에서 추출되는 특징을 이용하여 암시적 (Implicitly) 변화 사전 정보 (Change Prior)를 활용한 멀티스케일 특징 융합을 통해 경계 불완전성과 내부 공백 현상을 효과적으로 완화하였다. 하지만, 암시적 변화 사전 정보 (Implicit Change Prior)는 백본으로 추출되는 특징에 큰 영향을 받으며

잘못 추출된 특징으로부터 전파되는 오류에 민감하다는 한계가 있다. 이를 개선하기 위하여 본 논문의 선행 연구에서는 명시적 변화 사전 정보 (Explicit Change Prior)를 활용하는 변화 탐지 기술을 개발하였다. SD-Net^[13]은 간단한 형태의 변화 탐지 구조의 입력으로 영상과 함께 깊이 정보를 활용하는 경우 F1-Score 관점에서 정확도 성능 향상이 가능함을 보였다. 하지만, SD-Net이 깊이 정보 추출에 활용한 DepthAnything^[14]은 일반적인 자연 영상만을 대상으로 학습된 모델이기 때문에, 원격탐사 영상에 적용할 경우 추출되는 깊이 정보에 많은 잡음이 포함된다. 이에 따라 기대했던 수준의 정확도 성능 향상을 이루지는 못하였다. PDE-Net^[15]은 CNN으로 구성되는 비선형 필터를 활용하여 바닥과 같이 평지에서 발생하는 잘못된 깊이 정보를 억제하였다. 하지만, PDE-Net은 의미 정보의 부족으로 고가도로와 같은 비관심 객체와 관련된 잡음을 억제하지 못하여 시각적 결과 대비 F1-Score 관점 정확도 성능은 저하되는 한계를 보였다.

이러한 한계를 극복하기 위하여, 본 논문에서는 의미적 일관성을 강화하여 깊이 정보의 왜곡과 잡음을 효과적으로 억제하고, 이를 통해 추출된 명시적 변화 사전 정보를 기반으로 변화 탐지 구조를 확장하였다. 깊이 정보의 품질 저하로 인한 오탐 (False Detection)과 누락 (Miss Detection)을 완화하기 위하여 의미 기반 왜곡 필터 (Semantic Distortion Filter, SDF)를 도입하였으며, 깊이 정보와 영상 특징을 융합하여 변화 탐지를 위한 단서를 더욱 명확히 반영하는 깊이 정보 기반 변화 탐지 네트워크 (Depth-Guided Change Detection Network, DGCD-Net)를 제안한다. 제안하는 DGCD-Net은 명시적 변화 사전 정보로써 깊이 정보를 의미적으로 활용하여 정교한 변화 탐지가 가능함을 보였다. WHU-CD^[16]를 이용한 실험에서 DGCD-Net은 F1-score 관점 96.23%로 기존 연구들 대비 높은 수준의 정확도 성능을 달성하였으며 주관적 성능 비교에서 변화 탐지 결과의 경계 불완전 문제와 내부 불연속 문제를 효과적으로 개선됨을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 소개하고, 3장에서는 제안하는 DGCD-Net의 구조와 동작 원리를 상세히 기술한다. 4장에서는 실험 및 결과 분석을

통해 제안 방법의 성능을 평가하며, 마지막으로 5장에서는 본 연구의 결론을 제시한다.

II. 관련 연구

서로 다른 시점에 촬영된 동일 지역의 원격탐사 영상은 밝기나 색상 차이, 기하학적 불일치로 인해 변화 영역을 정확히 구분하기 어렵다. 이러한 한계를 보완하기 위해 깊이 정보는 구조적 변화를 식별하는 중요한 단서로 활용될 수 있다. 이를 바탕으로 본 논문에서는 깊이 정보를 활용하는 변화 탐지 기술 개발을 위한 관련 연구를 소개한다.

1. 변화 탐지

원격탐사 영상의 발전으로 서로 다른 시점에 촬영된 영상을 활용한 변화 탐지의 응용 범위가 크게 확장되었다. 변화 탐지는 동일한 지역의 시계열 영상을 비교하여 건물, 도로, 산림 등 지표면의 토지 이용 및 토지 피복 변화를 분석하는 핵심 기술이다. 그러나 촬영 시점, 조명, 기상, 계절 등의 외부 요인으로 인해 영상 간 밝기나 색상 차이 및 기하학적 왜곡이 발생할 수 있으며, 이는 모델의 변화 식별 정확도가 저하할 수 있다. 특히 원격탐사 영상은 복잡한 배경과 다양한 지형적 요소를 포함하고 있어, 의미 있는 변화를 안정적으로 추출하는 데 여전히 어려움이 존재한다^[17].

이러한 문제를 해결하기 위해 다양한 딥러닝 기반 변화 탐지 기법이 제안됐다. 대표적으로 CNN 기반 변화 탐지 알고리즘^[18]은 변화 탐지를 네 단계로 수행한다. 그 과정은 (1) 시점 간 차이를 반영한 차 영상 생성, (2) 의미 분할을 통한 변화 후보 영역 추출, (3) 클러스터링을 통한 잠재적 변화 영역의 그룹화, (4) 분류기를 이용한 최종 변화 맵 산출과 같다. 이러한 구조는 Synthetic Aperture Radar (SAR) 영상^[19]과 같이 잡음이 많은 환경에서도 효과적으로 작동하며, 기존 전통적 방법보다 높은 신뢰도의 결과를 제공한다. 한편, ChangeMamba^[12]는 Mamba 구조를 변화 탐지에 도입한 최신 모델로, 시공간적 관계 모델링을 통해 정밀한 변화 정보를 효과적으로 추출한다. 이는 국소 정보와 전역 정

보를 동시에 활용할 수 있어 CNN의 제한적 수용영역 문제와 트랜스포머의 높은 계산 복잡도 문제를 모두 완화한다. 또한 이 구조는 이진 변화 탐지뿐 아니라 의미적 (Semantic) 변화 탐지 건물 손상 평가와 같은 다양한 응용에서도 우수한 성능을 보였다.

이와 같은 다양한 딥러닝 기반 접근법들이 변화 탐지 성능을 향상해 왔으나, 대부분의 기존 방법은 광학 영상 (Optical Imagery)과 같은 단일 모달리티 (Single Modality)만을 입력으로 사용한다. 이러한 접근은 조명 변화, 계절적 요인, 그림자, 배경 왜곡 등 외부 환경 요인에 민감하여, 실제 변화가 없는 영역에서도 오탐 (False Detection)이 발생하기 쉽다^[18]. 예를 들어, 도로, 적재물과 같은 건물과 비슷한 패턴을 건물의 변화로 잘못 인식하는 사례가 발생할 수 있다.

본 논문에서는 이러한 문제를 완화하기 위하여 기존의 단일 모달리티 기반 변화 탐지 방식을 확장하여 깊이 정보 (Depth Information)를 추가로 활용하는 다중 모달리티 (Multimodal) 변화 탐지 모델인 DGCD-Net을 제안한다. 원격 탐사 영상에서 깊이 정보는 지면과 대상 물체 간의 거리 차이 및 고도 분포 (Elevation Distribution)를 반영하므로, 광학 영상의 밝기나 색상 변화에 영향받지 않고 실제 구조적 변화를 명시적으로 구분할 수 있는 핵심 단서로 활용될 수 있다.

2. 깊이 정보 추정

깊이 정보는 조명, 그림자, 계절 변화, 촬영 시점 차이와 같은 외부 요인으로부터 비교적 영향을 덜 받기 때문에 왜곡이 적다. 이러한 특성으로 인해, 깊이 정보는 평면 정보만 담고 있는 2D 영상에 공간적·기하학적 단서를 제공하여 객체의 크기나 형태, 위치 관계를 보다 정확하게 추론할 수 있다. 따라서 복잡한 장면 구조를 이해하고 객체 간의 구성을 인식하는 데 중요한 역할을 하며, 의미론적 분할 성능 향상에도 기여한다.

깊이 정보는 일반적으로 다중 시점 기하 (Multi-View Geometry, MVG) 기반 영상 분석이나, Lidar와 RGB-D 센서 등과 같은 깊이 센서를 활용하여 획득할 수 있다^[20,21]. MVG 기반 접근은 여러 시점에서 촬영된 영상을 이용하여

기하학적 삼각 측량을 통해 깊이를 추정하지만, 영상 간 정합 정확도에 크게 의존하며 넓은 지역이나 상공 촬영 영상에는 시차 (parallax)가 작아 깊이 추정이 어렵다. 한편, 깊이 센서를 활용하는 방법은 장비의 높은 비용과 데이터 취득의 물리적 제약으로 인해 대규모 원격 탐사 환경에 적용하기 어렵다. 이러한 이유로 원격 탐사 영상에는 정확하고 일반화할 수 있는 깊이 정보를 얻는 것이 해결할 과제이다.

최근에는 딥러닝 기술의 발전으로 단일 영상만으로도 높은 품질의 깊이 마스크를 생성할 수 있는 단안 영상 기반 깊이 추정 (Monocular Depth Estimation, MDE) 연구가 활발히 진행되고 있다^[22,23,24]. 대표적으로 DepthAnything은 대규모 학습 데이터와 효율적인 네트워크 설계를 바탕으로 Zero-Shot 환경에서도 높은 정확도의 깊이 추정이 가능하다는 장점이 있다. 그러나 이는 주로 정면 시점 (front-view)에서 촬영된 장면에 최적화되어 있어, 상공 시점 (aerial-view)에서 촬영된 원격탐사 영상에서는 깊이 추정의 정확도가 저하되는 한계가 있다. 이러한 문제를 개선하기 위해, 본 연구에서는 DepthAnything으로부터 추출한 깊이 정보에 의미 정보를 활용한 왜곡 억제를 수행하는 의미 기반 왜곡 필터인 SDF를 제안한다. 제안하는 SDF는 광학 영상으로부터 활용 가능한 의미적 단서를 이용하여 깊이 맵의 잡음을 제거하고 구조적 왜곡을 보정함으로써, 건물의 형태와 구조를 더 정확하게 복원한다.

III. 제안하는 깊이 정보기반 변화 탐지 기술

DGCD-Net이 변화 탐지를 수행하는 과정은 크게 공간적 분석, 시간적 분석, 그리고 변화 맵 생성 단계로 구성된다. 공간적 분석 단계는 한 시점의 영상에서 변화 탐지 대상의 기하적, 구조적, 그리고 의미적 정보를 추출한다. DGCD-Net은 정확하고 정교한 변화 탐지를 위하여 공간적 분석 과정에 깊이 정보 기반 명시적 사전 정보 (Depth-Guided Explicit Prior Knowledge)를 활용한다. 이는 DepthAnything으로부터 추출된 깊이 정보에서 의미 정보를 활용하는 비선형 필터인 SDF를 이용하여 왜곡된 깊이 정보를 억제하며, 이를 통하여 추출되는 의미 정보 기반 가상 깊이 맵 (Semantic-Guided Pseudo Depth Map, SGPDM)을 사전 정보로 활용하여 더욱 정교한 공간적 표현을 추출한다. 시간적 분석 단계에는 절대 차이값을 이용하여 시간별로 추출된 공간적 표현 간 변화 단서를 추출하며 이는 최종 변화 맵 생성에 활용된다. 그림 1은 본 논문에서 제안하는 DGCD-Net의 구조이다. 그림의 우측은 제안하는 DGCD-Net의 전체적 과정이며 좌측은 DGCD-Net이 SDF를 통하여 추출되는 SGPDM을 활용하는 과정이다.

본 장에서는 제안하는 DGCD-Net의 구성요소를 단계별로 상세히 기술한다. 먼저 3.1의 선행 연구에서는 기존 방법들의 한계와 본 연구의 설계 배경을 간략히 정리한다. 이어서 3.2의 SDF에서는 의미 정보를 활용하여 깊이 왜곡을 억

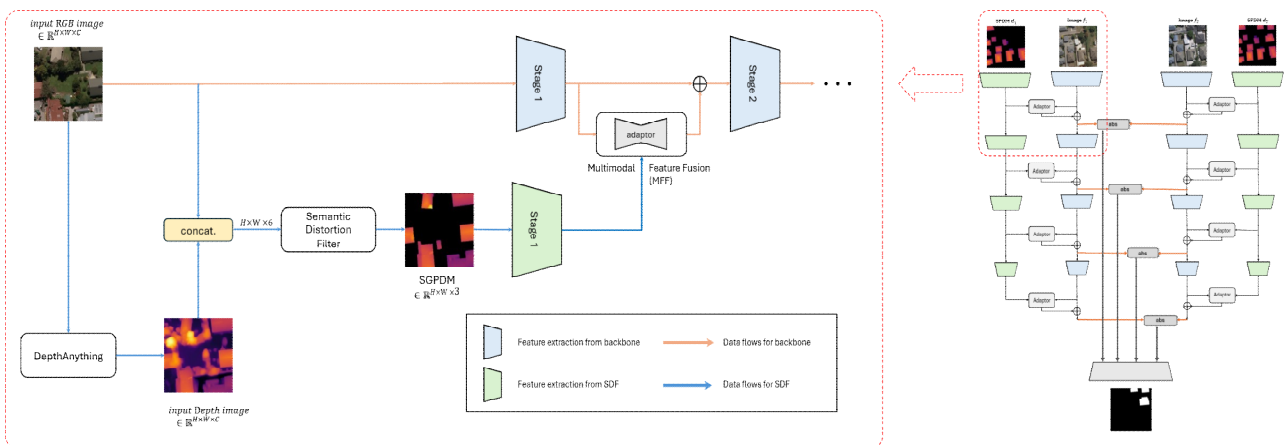


그림 1. 제안하는 DGCD-Net 구조

Fig. 1. Overview of the proposed DGCD-Net

제하고 정제된 SGPDM을 생성하는 과정을 설명한다. **Multimodal Feature Fusion** 절에서는 광학 영상과 깊이 정보를 통합하여 공간적·기하학적 특징을 효과적으로 결합하는 방법을 제시한다. 다음으로 변화 맵 생성 절에서는 시간적 차이를 기반으로 변화 단서를 추출하고 최종 변화 맵을 생성하는 과정을 다룬다. 마지막으로 구현 세부 사항 절에서는 네트워크 구조, 학습 설정, 라벨 생성과 같은 구현 세부 사항을 기술한다.

1. 깊이 정보를 명시적 사전 정보로 활용한 변화 탐지 방법

깊이 정보는 영상 내 객체의 기하적 구조와 거리 정보를 포함하고 있어, 조명 변화, 그림자, 계절적 차이와 같은 외부 환경 요인에 비교적 둔감하다. 이러한 특성으로 인해 깊이 정보는 복잡한 도시 장면이나 건물 밀집 지역에서 발생하는 비의도적 변화 (illumination-induced variation)를 완화하고, 실제 구조적 변화를 구분하는 데 효과적인 단서로 활용될 수 있다. 이러한 가능성에 기반하여, 선행 연구에서는 깊이 정보를 변화 탐지 모델에 직접 통합하는 다양한 시도가 이루어졌다.

첫 번째 연구에서는 Simple Depth-based Change Detec-

tion Network (SDCD-Net)^[13]를 제안하여 광학 영상과 깊이 영상을 동시에 입력으로 사용하는 구조를 설계하였다. 실험 결과, 깊이 정보를 추가로 활용한 모델이 기존의 단일 광학 영상 기반 모델 대비 F1-Score가 미세하게 향상된 결과를 보였으며, 이는 깊이 정보가 변화 탐지 성능 향상에 실질적인 기여 요인임을 실험적으로 입증한 결과로 해석된다. 이후 연구에서는 단순한 깊이 활용의 한계를 보완하기 위해, Pseudo Depth Estimation Network (PDE-Net)를 제안하였다. PDE-Net은 DepthAnything 모델을 통해 추출된 단안 기반 깊이 정보에 포함된 잡음을 완화하고, 건물 영역 중심의 Pseudo Depth Map을 생성하여 SDCD-Net^[13]의 입력으로 활용하였다. 이를 통해 왜곡이 억제되는 변화 정보를 확보함으로써 건물과 지표면을 명확히 구분하고, 세밀한 변화 탐지가 가능함을 보였다. 그러나 이 방법은 지표면에서 발생하는 왜곡만을 억제할 뿐, 특정 대상 (예: 건물, 고가도로 등)의 변화를 중점적으로 탐지해야 하는 응용 환경에서는 여전히 비관심 영역의 오탐 (False Detection)이 빈번히 발생하는 한계를 보였다.

기존 선행 연구에서 활용한 깊이 정보는 단안 기반 깊이 추정 결과 또는 지표면 중심의 왜곡 억제에 한정되어 있으며, 건물 변화 탐지와 같이 특정 응용을 고려한 상황에서 비관심 객체 (non-salient object)로부터 발생하는 의

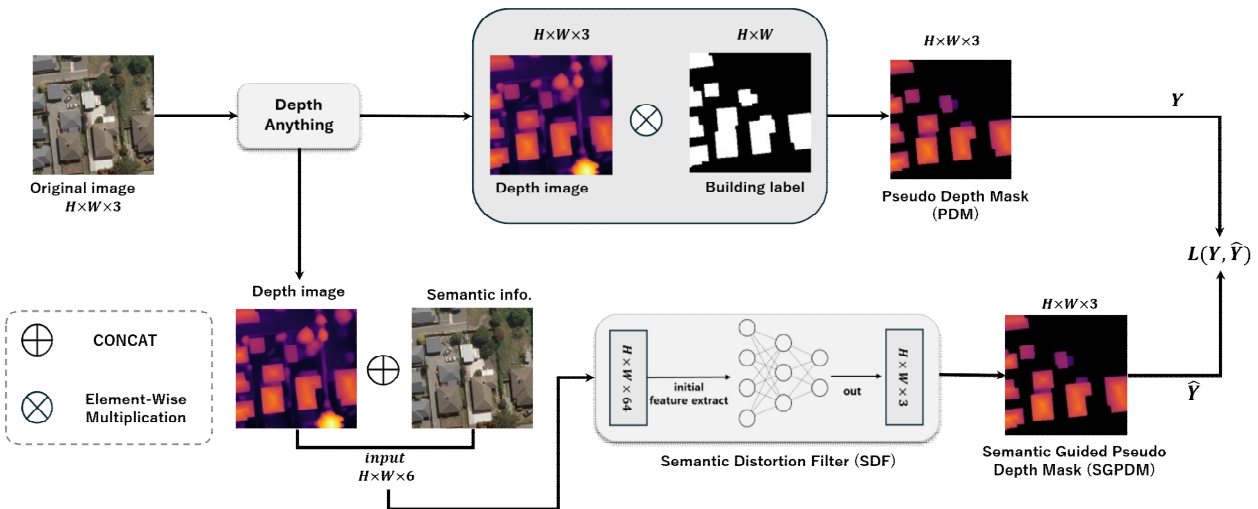


그림 2. Semantic Distortion Filter (SDF)의 학습 및 추론 과정
Fig. 2. Training and inference process of the Semantic Distortion Filter (SDF)

미적 왜곡 (Semantic Distortion)은 충분히 반영되지 않았다. 이러한 한계를 극복하기 위하여 본 연구는 의미 정보를 활용하여 깊이 왜곡을 억제하는 비선형 필터인 SDF를 제안한다. SDF는 Depth Anything으로부터 추출된 깊이 정보와 광학 영상으로부터 활용 가능한 의미적 단서 (Semantic Cues)를 융합하여 왜곡된 깊이 영상을 정제하며, 이를 통해 Semantic-Guided Pseudo Depth Map (SGPDM)을 생성한다. 생성된 SGPDM은 DGCD-Net의 명시적 사전 정보 (Explicit Prior)로 활용되어 공간적 분석 단계에서 신뢰성 있는 구조적 표현을 추출하는 데 기여한다.

2. Semantic Distortion Filter

본 논문에서 제안하는 SDF는 깊이 정보에서 건물과 같은 관심 객체 외의 영역으로부터 발생하는 의미적 왜곡을 효과적으로 억제하기 위해 설계되었다. 일반적으로 깊이 정보는 장면 내의 거리 정보를 제공하지만, 의미적 구분을 포함하지 않기 때문에 나무, 도로, 그림자와 같은 비관심 객체로부터의 깊이 변화가 탐지 과정에서 잡음으로 작용할 수 있다. 이러한 왜곡을 억제하기 위해서는 관심 영역과 비관심 영역을 구분할 수 있는 의미적 단서 (Semantic Cue)가 필요하다. 이를 위해서는 사전에 학습된 (Pretrained) 의미적 분할 (Semantic Segmentation) 모델을 활용할 수 있으나, 이는 추론 복잡도의 증가와 실시간 적용의 체약을 초래한다.

이러한 한계를 극복하기 위해 본 논문에서는 광학 영상을 활용하여 깊이 정보의 의미적 왜곡을 간접적으로 보정하는 SDF를 제안한다. 깊이 영상과 이에 대응하는 광학 영상은 조기 융합 (Early Fusion)을 통해 결합하며, 이 융합된 입력은 사전에 학습된 U-Net 기반의 Channel-Spatial Attention (CSA) 네트워크^[28]를 통하여 처리된다. CSA 네트워크는 입력으로부터 의미적 단서를 추출하고, 이를 통해 SGPDM을 생성한다. SGPDM은 위치별로 왜곡 발생 가능성을 추정하여 깊이 정보 내의 의미적 왜곡을 억제하는 적응형 게이트 마스크 (Adaptive Gating Mask)로 사용된다. 결과적으로 SDF는 깊이 정보 내 의미적 왜곡을 제거하면서도, 추가적인 의미적 분할 과정 없이 효율적으로 동작하

여 깊이 기반 변화 탐지의 강인성을 향상하게 한다. 그림 2는 SDF의 학습 및 추론 과정이다.

SDF는 깊이 영상 (D)과 광학 영상 (I)을 입력받아 SGPDM \hat{Y} 를 생성하도록 학습한다. SDF의 추론 과정은 다음과 같이 정의된다.

$$\hat{Y} = SDF(Concat([I, D]); \theta) \quad (1)$$

여기서 $Concat([I, D])$ 은 광학 영상과 깊이 영상의 조기 융합 (Early Fusion) 과정을 의미하며, SDF는 U-Net 기반의 CSA 네트워크로 구현된다. 네트워크의 출력은 의미적 왜곡을 억제한 깊이 지도인 \hat{Y} 이다. SDF의 학습 과정에서는 정답 깊이 지도 (label) y 를 다음과 같이 생성한다.

$$y = D \odot M \quad (2)$$

깊이 영상 D 에 건물 영역에 해당하는 의미 마스크 $M \in \{0, 1\}^{H \times W}$ 를 곱하여 비관심 영역의 깊이 값을 제거한 지도를 학습용 타겟으로 사용한다. 여기서 $M_{ij} = 1$ 은 건물(관심 영역), $M_{ij} = 0$ 은 비관심 영역을 의미한다. 이렇게 생성된 y 는 Semantic-Guided Ground Truth로서 SDF가 건물 영역 중심의 깊이 보정에 집중하도록 유도한다.

$$\hat{\theta} = \arg \min_{\theta} \mathcal{L}(y, \hat{Y}) \quad (3)$$

여기서 SDF의 파라미터 θ 는 예측 결과 \hat{Y} 와 정답 y 사이의 손실 함수를 최소화하도록 학습된다.

$$\mathcal{L}(y, \hat{Y}) = \frac{1}{HW} \sum_{i,j} \|y_{i,j} - \hat{Y}_{i,j}\|_1 \quad (4)$$

여기서 $\mathcal{L}(\cdot)$ 은 pixel-wise reconstruction loss로 L1 손실을 사용한다. 이 목적함수는 SDF가 비관심 영역의 왜곡을 억제하면서도, 관심 영역의 깊이 구조를 정밀하게 복원하도록 학습을 유도한다.

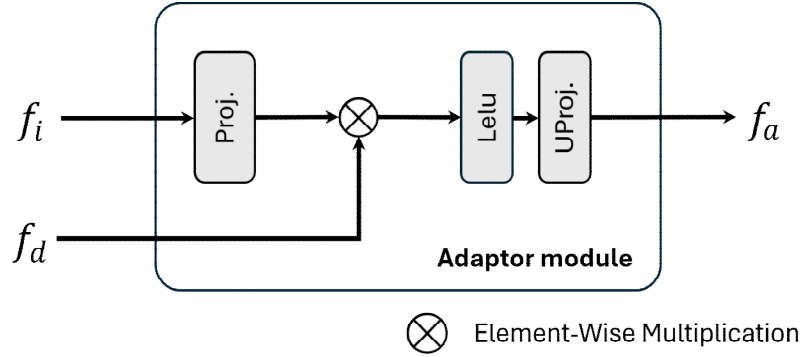


그림 3. Multimodal Feature Fusion (MFF)을 위한 Adapter 구조
Fig. 3. Adapter structure for Multimodal Feature Fusion (MFF)

3. Multimodal Feature Fusion (MFF)

깊이 정보는 광학 영상에서 얻기 어려운 장면의 구조적 단서 (Structural Cue)를 제공하므로 변화 탐지의 신뢰도를 향상할 수 있다. 그러나 광학 영상과 깊이 영상은 물리적 의미와 통계적 분포가 상이하기 때문에, 단순히 병합 (Concatenation)만으로는 두 모달리티 간의 상호 보완적 정보를 효과적으로 활용하기 어렵다. 특히 깊이 지도는 조도, 반사, 질감의 영향을 받지 않지만, 광학 영상은 이러한 요인에 민감하므로 적절한 융합 구조 (Fusion Architecture) 없이 결합할 때 오히려 특징 불일치 (Feature Inconsistency)를 유발할 수 있다. 이러한 문제를 해결하기 위해 본 논문에서는 Adapter 기반의 Multi-Modal Feature Fusion (MFF) 모듈을 제안한다. MFF는 SDF를 통해 정제된 깊이 정보 (SGPDM)를 광학 영상으로부터 추출된 특징 (Feature Representation)에 명시적으로 반영 (Explicitly Inject)함으로써 깊이 기반 구조 단서를 효과적으로 통합하는 다중 모달 융합 (Multimodal Fusion) 메커니즘을 제공한다. 이를 통해 DGCD-Net은 광학 영상의 시각적 변화(밝기, 그림자 등)에는 둔감하면서도 구조적 변화 (Structural Displacement)에 민감한 변화 표현을 학습할 수 있다. 그림 3은 Multimodal Feature Fusion을 위한 Adapter 구조이다.

광학 영상으로부터 추출된 특징 $f_i \in \mathbb{R}^{C_i \times H \times W}$ 와 깊이 정보로부터 얻어진 특징 $f_d \in \mathbb{R}^{C_d \times H \times W}$ 은 서로 다르게 통계적 분포와 표현 공간을 갖는다. 두 표현 공간 차이로 인한 특징 불일치를 보정하기 위하여 Adapter 모듈은 두

특징에 결합을 수행한다. 그 과정은 다음과 같이 정의된다.

$$f_a = UProj(\sigma(f_d \odot Proj(f_i))) \quad (5)$$

여기서 $Proj(\cdot)$ 은 1×1 합성곱 (Convolution) 연산을 이용한 투영 (Projection) 변환으로 광학 특징 f_i 를 깊이 특징 f_d 와 유사한 통계적 분포로 사상한다. \odot 은 element-wise multiplication으로 깊이 특징이 광학 특징의 반응을 조절하는 adaptive modulation 역할을 수행한다. $\sigma(\cdot)$ 은 비선형성을 위한 활성화 함수로 Leaky ReLU 함수를 사용한다. $UProj(\cdot)$ 은 또 다른 1×1 합성곱 연산으로 조정된 특징을 다시 광학 특징 공간으로 되돌리는 역 투영 (Inverse Projection) 연산이다. 이 과정을 단계적으로 표현하면 다음과 같다.

$$f_i \xrightarrow[Proj]{Transform} f_i' \xrightarrow[f_d]{modulated\ by} f_a' \xrightarrow[UProj]{Inverse\ Transform} f_a \quad (6)$$

즉, Adapter는 두 모달리티의 통계적 불일치를 최소화하기 위해 Transform - Inverse Transform 구조를 통해 깊이 특징을 광학 특징에 정렬 (Alignment)시킨다. 이렇게 변환된 f_a 는 MFF 내에서 백본의 다음 단계 (Stage) 및 시간적 분석의 입력으로 사용되어 신뢰성 있는 구조적 특징을 추출한다.

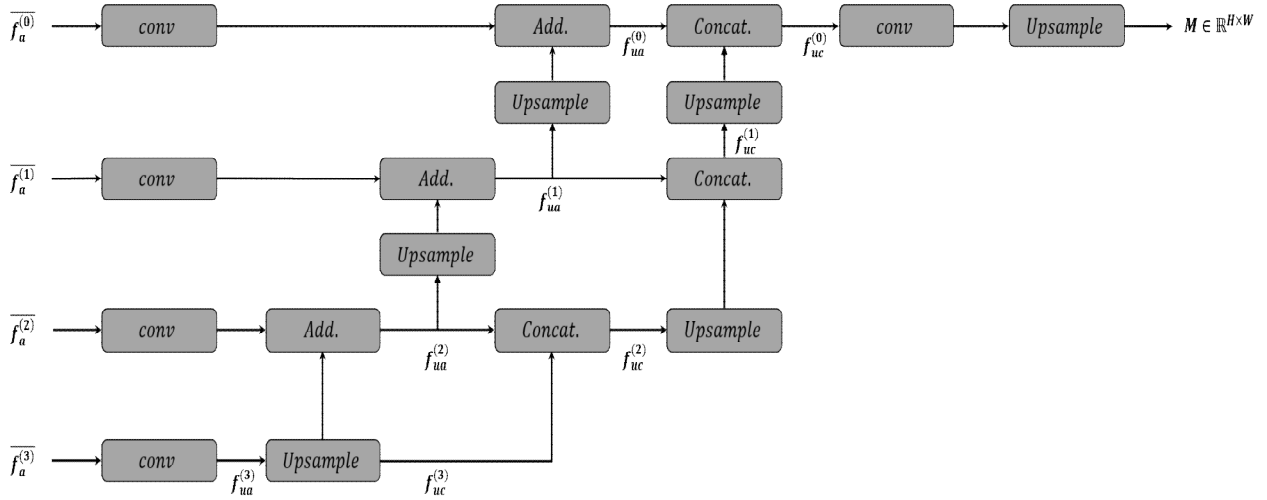


그림 4. 변화 맵 생성 네트워크 구조

Fig. 4. Network architecture for change map generation

4. 변화 맵 생성

변화 맵 생성 모듈은 두 시점의 입력 영상 (I_0, I_1) 과 이에 대응하는 깊이 정보 (D_0, D_1) 간의 융합을 통해 추출된 단계적 특징 $\{f_{a,0}^{(i)}\}_{i=0}^3$ 과 $\{f_{a,1}^{(i)}\}_{i=0}^3$ 을 입력으로 받아 단계적 업샘플링 (Upsampling), 덧셈 (addition), 채널 결합 (concatenation)을 반복하여 최종 변화 맵을 생성한다. 먼저 각 단계 i 에서 두 시점의 절대적 특징 차이를 계산하여 변화 후보 특징을 구한다.

$$\bar{f}_a^{(i)} = |f_{a,0}^{(i)} - f_{a,1}^{(i)}| \quad (7)$$

여기서 $f_{a,0}^{(i)}$ 는 시점 I_0 으로부터 추출된 i 번째 특징, $f_{a,1}^{(i)}$ 는 시점 I_1 으로부터 추출된 i 번째 특징이다. $\bar{f}_a^{(i)}$ 는 두 시점 간의 구조적 변화를 반영하는 기본 입력으로 사용된다. 가장 상위 레벨 특징 $\bar{f}_a^{(3)}$ 은 1×1 합성곱 연산과 ReLU 활성화를 거쳐 업샘플링 가능한 형태로 변환된 후, 하위 단계 특징과의 업샘플링 및 덧셈 연산으로 단계적 융합된다.

$$f_{ua}^{(i+1)} = Up\left(conv(f_{ua}^{(i)})\right) + conv(\bar{f}_a^{(i)}), \quad i = 0, 1, 2 \quad (8)$$

여기서 $f_{ua}^{(i)}$ 는 시점 i 의 업샘플링 단계에서 추출된 상위 특징 맵을 의미하며, $\bar{f}_a^{(i)}$ 는 동일 단계에서 하위 계층으로부터 전달되는 보조 특징 맵을 나타낸다.

$$f_{uc}^{(i+1)} = Concat(f_{uc}^{(i)} + f_{ua}^{(i+1)}), \quad (9)$$

각 단계에서 융합된 특징은 추가적인 업샘플링과 채널 결합을 통해 다중 해상도 정보를 보존한 통합 특징 맵을 형성한다. 이러한 계층적 융합 구조를 통해 고해상도 세부 정보와 저해상도 구조 정보가 상호 보완적으로 결합하며, 결과적으로 공간적 정합성과 구조적 일관성이 높은 변환 표현을 생성한다. 최종 통합 특징은 1×1 합성곱 연산을 통해 채널 차원이 클래스 수(변화/비 변화)로 축소되고 양선형 보간법 (bi-linear Interpolation) 기반 업샘플링을 통해 입력 영상과 동일한 해상도로 복원되어 최종 변화 맵 $M \in \mathbb{R}^{H \times W}$ 이 출력된다.

5. 구현 세부사항

본 논문에서 제안한 DGCD-Net은 PyTorch^[25] 프레임워크를 기반으로 구현하였다. 백본 네트워크는 ImageNet 데

이터넷으로 사전 학습된 VGG11^[8]을 사용하였으며 최적화 알고리즘은 AdamW^[26]를 적용하였다. 여기서 백본 네트워크는 미세조정을 수행하였다. 초기 학습률은 0.0005로 설정 후 Cosine Annealing Warm Restarts^[27] 방법을 사용하였으며 초기 주기 T_0은 {5,10,15}, 배수 인자 T_mult는 {2,4}로 설정하여 grid search를 수행하였다. 배치 크기 (batch size)는 8로 설정하여 총 50 epoch 동안 학습을 수행하였으며 파라미터는 weight decay는 L2 정규화를 기반으로 0.0025로 수행되었다. 데이터 증강 (Data Augmentation)은 Random Horizontal Flip, Random Rotation, Random Exchange (두 시점 영상의 위치 교환) 연산을 포함하여, 다양한 기하학적 변형에 대한 일반화 성능을 확보하였다. 모델 학습 과정에서 Early Stopping 기법을 적용하여 검증 세트 기준 F1-Score가 최솟값을 기록한 모델을 최종 결과로 선정하였다. 출력 변화 맵은 Sigmoid 함수를 거쳐 [0,1] 범위의 확률값으로 변환된다. 손실 함수는 픽셀 단위의 Binary Cross-Entropy (BCE)로 정의되며 변화(1)와 비 변화(0) 클래스 간의 예측 오차를 최소화하도록 모델을 학습하였다.

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (10)$$

여기서 y_i 와 \hat{y}_i 는 각각 정답 라벨과 예측 확률을 의미한다.

SDF의 구조는 SGSR^[28]의 CSA U-Net을 기반으로 하되 입력 및 출력 채널 수만을 조정하여 구성하였다. 최적화 알고리즘은 ADAM^[29]을 사용하였으며 초기 학습률은 0.0004로 설정하였으며 별도의 스케줄링 기법은 적용하지 않았다. 배치 크기는 8로 설정하여 총 100 epoch 동안 학습을 수행하였다. 데이터 증강은 Random Crop, Random Horizontal Flip, 그리고 Random Rotate를 적용하였다. 학습 완료된 SDF 모델은 고정된 상태로 DGCD-Net의 학습 및 추론에 활용된다. 모델 학습 과정에서 Early Stopping 기법을 적용하여 검증 세트 기준 L1 Loss가 가장 낮은 모델을 최종 결과로 선정하였다. 입력 해상도는 모두 256×256 으로 구성하였다.

IV. 실험 및 결과 분석

1. 실험 환경

제안하는 DGCD-Net은 Python 3.11.9와 PyTorch 2.8.0을 기반으로 구현되었으며, 단일 NVIDIA RTX 4060 GPU 환경에서 실험을 수행하였다. GPU 가속을 위해 CUDA 12.1과 CuDNN 8.9.7 라이브러리를 사용하였다. 본 연구에서 사용된 실험 환경은 다음과 같다. 모델 학습과 평가는 AMD Ryzen 7 5800X 프로세서(8코어, 3.80GHz)를 기반으로 하는 시스템에서 수행되었다. 메모리는 충분한 학습 성능을 확보하기 위해 32.0GB RAM을 사용하였으며, 모든 실험은 동일한 설정에서 진행하였다.

2. 데이터셋

WHU-CD 데이터셋은 건물 변화 탐지를 목적으로 구축된 항공 영상 기반 벤치마크 데이터셋이다. 이 데이터셋은 뉴질랜드 (New Zealand) Christchurch 지역의 동일 지역을 서로 다른 시점(2012년과 2016년)에 촬영한 고해상도 (0.075m/pixel) 항공 영상 쌍으로 구성되어 있다. 2011년 발생한 규모 6.3의 지진으로 인한 도시 재건 과정을 반영하고 있어 실제 건물의 신축·철거와 같은 구조적 변화를 포함한다. 영상의 공간 범위는 약 20.5km이며 2012년 영상에는 12,796개, 2016년 영상에는 16,077개의 건물 객체가 포함되어 있다. 두 시점의 영상은 30개의 지상 기준점 (Ground Control Point, GCP)을 이용하여 약 1.6 픽셀 수준의 정합 정확도로 기하 보정되었으며, 각 시점의 건물 벡터 지도와 라스터 라벨이 함께 제공된다. 데이터셋의 변화 전(2012)과 변화 후(2016) 영상은 완전히 중첩되어 제공되므로 변화 탐지 알고리즘의 학습 및 정량적 평가를 위한 대표적인 표준 데이터셋으로 널리 활용되고 있다.

3. 실험 결과

표 1의 정량적 비교 결과는 본 연구에서 제안한 DGCD-Net이 WHU-CD 데이터셋에서 기존 변화 탐지 모델들을 전반적으로 상회하는 성능을 달성했음을 명확히 보여준다.

표 1. WHU-CD 데이터셋을 활용한 기존 변화 탐지 방법들과 DGCD-Net의 변화 탐지 정량적

Table 1. Quantitative comparison between DGCD-Net and existing change detection methods on the WHU-CD dataset

Model	Backbone	Precision	Recall	F1-Score
SNUNet-CD ^[30]	NestedUNet	83.52	82.46	53.99
CGNet ^[11]	VGG16	94.47	90.79	92.59
ChangeCLIP ^[31]	ResNet50	95.63	94.02	94.82
MambaBCD-Base ^[12]	Transformer	95.90	92.29	94.06
FTA-Net ^[32]	MobileNetV2	96.02	94.43	95.21
CDMamba-P ^[33]	CDMamba	95.58	92.01	93.76
DGCD-Net	VGG11	98.07	94.46	96.23

기존 사전 정보 기반 모델인 CGNet과 ChangeCLIP은 각각 RGB 영상에서 암시적 패턴을 통해 구조적 단서를 학습하거나, Vision-Language 모델의 semantic prior를 활용함으로써 일정 수준의 성능을 보였으나, 실제 건물의 높이나 형태와 같은 물리적 구조 정보를 반영하지 못해 복잡한 도시 장면의 세밀한 변화 구분에 한계가 있었다. 한편, MambaBCD-Base와 CDMamba-P와 같은 Mamba 기반 모델들은 장거리 의존성을 효과적으로 학습하여 광역 문맥 정보를 포착할 수 있으나, 입력 정보가 RGB appearance에 한정되어 조명·그림자·배경 질감 변화로 인한 비의도적 활성화를 완전히 억제하지 못하는 한계를 보였다. 또한 주파수 도메인 기반의 FTA-Net은 RGB-only 모델 중 가장 높은 정확도를 기록했으나, 구조적 변화 단서 부재로 인해 변화 경계의 안정적 분리에 제한이 존재했다. 이에 비해 DGCD-Net은 RGB 영상과 SGPDM 기반의 깊이 정보를 명시적으로 결합하여 준물리적 구조 단서를 제공함으로써 외란 요인에 강인한 변화 표현을 학습할 수 있었으며, 그 결과 Precision 98.07%, Recall 94.46%, F1-score 96.23%를 기록하며 모든 비교 모델 대비 일관적으로 향상됨을 보였다. 이러한 수치적 우위는 깊이 기반 구조 정보와 멀티모달 융합 전략이 변화 탐지의 신뢰도 향상에 결정적으로 기여했음을 실험적으로 입증하는 결과라 할 수 있다.

4. Ablation Study

본 연구에서는 제안하는 DGCD-Net의 핵심 구성요소인

깊이 정보 활용 (Depth-Guided Explicit Prior)과 Adaptor 모듈이 변화 탐지 성능에 미치는 영향을 검증하기 위해 정량적 비교 실험 (Ablation Study)을 수행하였다. 깊이는 외부 환경 변화에 대한 영향이 없고 영상 왜곡이 적어, 건물 경계를 정밀하게 표현하는 구조적 단서로서 중요한 역할을 한다. 또한 Adaptor 모듈은 이러한 깊이 정보를 효과적으로 반영하여 의미 있는 특징을 선택적으로 강조함으로써 모델의 표현력을 향상한다. 표 2는 WHU-CD 데이터셋을 이용해 깊이 정보 및 Adaptor 모듈의 유무에 따른 성능 차이를 비교한 결과이다. 실험 결과, DGCD-Net은 가장 높은 성능을 보인 기존 모델 SDGD-Net 대비 Precision

표 2. Depth를 활용한 기존 변화 탐지 방법과 DGCD-Net의 변화 탐지 정량적 비교 결과

Table 2. Quantitative comparison between DGCD-Net and depth-based change detection methods

Model	Backbone	Precision	Recall	F1-Score
SDGD-Net ^[13]	VGG11	94.93	90.66	92.75
PDE-Net ^[15]	VGG11	93.03	90.14	91.56
DGCD-Net	VGG11	98.07	94.46	96.23

표 3. DGCD-Net의 입력으로 사용된 깊이 마스크 유형별 성능 비교 결과
Table 3. Performance comparison according to different depth mask types used as DGCD-Net inputs

Depth Input	Precision(%)	Recall(%)	F1-score(%)
DepthAnything ^[14]	90.77	93.44	92.09
PDM ^[15]	97.73	94.46	96.05
SGPDM	98.07	94.46	96.23

3.14, recall 3.80%, F1-score 3.48% 향상을 달성하였다. 이는 건물 중심 영역에 집중된 의미 있는 정보만을 포함하는 SGPDM을 활용함으로써, 불필요한 배경 잡음을 억제하고 건물 변화 영역을 정밀하게 탐지할 수 있었기 때문이다. 또한, 표 3의 결과에서 확인할 수 있듯이 동일한 Adaptor 모듈을 적용한 조건에서도 입력 깊이 데이터의 종류에 따라 성능 차이가 나타났다. DepthAnything, PDM, SGPDM을 각각 입력으로 사용한 SGPDM을 활용했을 때 가장 높은 성능을 보였다. 이는 SGPDM이 형태적 일관성을 유지하면서 배경 잡음을 억제하여 모델이 변화 영역에 집중하도록 유도했음을 의미한다.

네트워크 구조 설계 관점에서 DGCD-Net의 일반화 성능 측정을 위한 정량적 비교 실험으로는 백본 네트워크와 깊이 추정 모델을 변경해가며 실험을 수행하였다. DGCD-Net에 다양한 백본 네트워크를 적용하여 성능을 비교한 결과는 표 4와 같다. 실험 결과 DGCD-Net은 다양한 백본 네트워크에 대하여 전반적으로 높은 변화 탐지 성능을 보임을 알 수 있다. 다만 VGG 및 Resnet 대비 Transformer 모델은 약간 낮은 성능을 보이는데 이는 Transformer 모델이 갖는 귀납 편향 (Inductive Bias) 문제로 인하여 패치 경계면에서

불연속성이 발생하기 때문으로 해석된다. 모델의 복잡도 관점에서 VGG11은 상대적으로 단순한 구조를 가짐에도 정확도 성능에는 큰 차이가 없어 실제 응용에 효율적으로 활용 가능할 것으로 판단된다. DGCD-Net에 다양한 깊이 추정 모델과 SDF를 적용하여 성능을 비교한 결과는 표 5와 같다. 실험에는 대표적인 깊이 추정 모델인 Depth Anything과 Midas^[36]를 사용하였으며, 표의 ‘O’와 ‘X’는 각각 SDF 적용 유무를 나타낸다. 실험 결과, Depth Anything에 SDF를 결합한 경우가 F1-score 96.23%를 기록하며 전체 비교군 중 가장 우수한 성능을 달성하였다. 이와 유사하게, Midas 모델 기반 실험에서도 SDF 적용 시 미적용 대비 성능이 향상됨을 확인하였다. 이러한 결과는 SDF가 깊이 데이터 내의 노이즈나 비관심 객체를 효과적으로 필터링 및 억제함으로써, 모델의 탐지 성능 최적화에 기여함을 보여 준다.

정성적 비교 결과는 그림 5와 같다. 영상에서 노란색은 정확히 예측한 영역 (True Positive), 파란색은 오검출 (False Positive), 빨간색은 미검출 (False Negative)을 의미한다. 그림 5에서 확인할 수 있듯이 ChangeCLIP을 포함한 기존 모델들은 텍스처 변화나 조명 차이에 반응하여 실제 변화가 없는 도로·지면·건물에 인접한 적재물 영역에서 불필요한 탐지가 발생하거나, 건물 경계 부근에서 변화가 부분적으로 누락됨을 보였다. 특히 ChangeCLIP의 경우 텍스트-비전 기반 구조로 인해 건물 형태 변화보다는 지역적 시각 변화에 더 크게 반응하며, 실제 변화 영역을 일관되게 포착하지 못하는 사례가 나타난다. 반면 제안하는 DGCD-Net은 SGPDM을 통해 정제된 깊이 기반 구조 단서를 활용하여 이러한 오검출·미검출 문제를 효과적으로 억제하였다. DGCD-Net은 건물의 외곽선과 증축·철거와 같

표 4. 백본 아키텍처 변경에 따른 모델 성능 평가

Table 4. Performance comparison according to backbone architecture changes

Backbone	Precision(%)	Recall(%)	F1-score(%)
VGG11 ^[8]	98.07	94.46	96.23
VGG16 ^[8]	98.11	93.98	96
ResNet18 ^[34]	97.99	94.57	96.25
Transformer ^[35]	93.37	97.9	95.58

표 5. SDF 적용 및 깊이 데이터 활용에 따른 모델 성능 평가

Table 5. Quantitative comparison of performance with and without SDF

Depth Estimation Model	SDF	Precision(%)	Recall(%)	F1-score(%)
Depth Anything ^[14]	X	95.5	89.29	92.29
Depth Anything ^[14]	O	98.07	94.46	96.23
Midas ^[36]	X	94.26	85.99	89.93
Midas ^[36]	O	97.36	94.63	95.98

그림 5. 기존 변화 탐지의 방법과 DGCD-Net의 정성적 비교 결과 (노란색 영역: GT와 예측이 일치한 부분, 파란색 영역: False Positive, 빨간색 영역: False Negative)

Fig. 5. Qualitative comparison between existing change detection methods and DGCD-Net (Yellow: True Positive, Blue: False Positive, Red: False Negative)

		Visualization Results					
Original Image		GT	CGNet ^[11]	ChangeClip ^[31]	SDCD-Net ^[13]	PDE-Net ^[15]	DGCD-Net
Image I_1	Image I_2						
WHU-CD							

은 구조적 변화를 정확하게 탐지하며, 비변화 영역의 불필요한 탐지를 최소화한다. 그 결과 DGCD-Net이 생성한 변화 맵은 정답 (GT)과 가장 유사한 형태를 보였으며, 이는 깊이 정보의 명시적 사전 정보와 Adaptor 모듈의 결합이 변화 탐지 정확도를 실질적으로 향상시킴을 시각적으로 보여준다.

5. 실험 결과 분석

표 1은 WHU-CD 데이터셋을 기준으로, 깊이 정보를 사용하지 않는 기존 변화 탐지 모델들과 제안하는 DGCD-Net을 세 가지 주요 평가 지표로 비교한 결과이다. 이를 통해

기존 모델 중 가장 높은 성능을 보인 FTA-Net은 F1-score가 약 95.21% 수준에 머무르며 RGB 기반 단일 모달리티 접근의 성능적 한계를 확인할 수 있다. 반면, DGCD-Net은 F1-score 96.23%를 기록하여 모든 비교 모델을 상회하는 성능을 보여주었으며, 이는 깊이 정보를 사전 정보로 도입한 설계가 변화 탐지 성능 향상에 유효함을 나타낸다. 또한, 깊이 정보의 품질이 성능에 미치는 영향을 검증하기 위해 표 3에서 수행한 깊이 마스크 유형별 비교 결과를 통해, 영상에 잡음이 많이 포함되어 있는 DepthAnything 그대로 사용하여 실험한 결과 F1-score는 92.09%, 정제되지 않은 깊이 정보인 PDM을 사용한 실험 결과 F1-score는 96.05% 수준에 머물렀다. 반면, SDF를 적용하여 생성한 SGPDm을

입력으로 활용한 경우 F1-score가 96.23%로 크게 향상됨을 확인할 수 있었다. 이는 깊이 정보의 의미 기반 정제가 변화 탐지에 필요한 구조적 단서를 안정적으로 확보하는 데 중요한 역할을 함을 보여준다. 더 나아가 동일한 깊이 입력을 사용하는 조건에서도 Adaptor 모듈의 적용 여부에 따라 성능 차이가 발생하는 것을 표 2를 통해 확인하였다. Adaptor를 적용한 설정에서 정확도, 재현율, F1-score가 모두 상승하였으며, 이를 통해 깊이 기반 구조 단서가 RGB 특징과 정렬된 표현 공간에서 결합할 때 두 모달리티의 상호 보완적 정보가 더 효과적으로 통합됨을 알 수 있다. 즉, 깊이 정보를 단순히 입력으로 병합하는 것보다, SGPDM 기반 정제와 Adaptor 기반 표현 정렬이 함께 이루어질 때 변화 여부에 대한 특징 반응이 더욱 안정적으로 나타난다는 사실을 실험적으로 입증한다. 이와 같은 정량적 분석 결과는 그림 5의 정성적 비교와도 일관된다. ChangeCLIP을 포함한 기존 모델들은 텍스처 변화나 조명 차이에 반응하여 실제 변화가 없는 도로·지면·건물 인접 적재물 영역에서 불필요한 탐지가 발생하거나, 건물 경계 부근에서 변화가 부분적으로 누락되는 경향을 보인다. 반면 DGCD-Net은 SGPDM으로 정제된 깊이 기반 구조 단서를 활용하여 비변화 영역의 과도한 활성화를 억제하고 실제 구조적 변화를 명확하게 포착하는 결과를 보였다. 이를 통해 DGCD-Net이 생성한 변화 맵이 정답과 가장 유사한 형태를 나타내며, 깊이 정보와 Adaptor 모듈의 결합이 변화 탐지 정확도를 시각적으로도 향상함을 확인했다.

V. 결 론

본 논문에서는 원격탐사 영상에서 건물 중심의 변화를 정밀하게 탐지하기 위해 DGCD-Net을 제안하였다. 제안된 모델은 SGPDM과 Adaptor 모듈을 결합하여 복잡한 배경이나 다양한 장면에서도 안정적이고 정밀한 변화 탐지 성능을 달성하였다. WHU-CD 데이터셋에서 평가한 결과, DGCD-Net은 Precision 98.07%, Recall 94.46%, F1-Score 96.23%로 기존 방법 대비 우수한 성능을 보였으며, 특히 객체의 경계 변화 영역을 더욱 정확히 추출함으로써 깊이 정보의 유효성을 입증하였다. 이는 깊이 정보가 건물의 구

조적 특성과 존재 유무를 보존하여, RGB 영상만으로는 구분하기 어려운 변화를 인지할 수 있게 한 결과로 해석된다.

참 고 문 헌 (References)

- [1] J. Chen, D. Hou, C. He, Y. Liu, Y. Guo and B. Yang, "Change Detection With Cross-Domain Remote Sensing Images: A Systematic Review," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 11563-11582, 2024. doi: <https://doi.org/10.1109/JSTARS.2024.3416183>
- [2] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Cham: Springer international publishing, 2015.
- [3] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. doi: <https://doi.org/10.1109/CVPR.2016.90>
- [4] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in Proc. Int. Conf. Learn. Representations (ICLR), 2021. Available: <https://openreview.net/forum?id=YicbFdNTTy>
- [5] Wiratama, Wahyu & Lee, Jongseok & Park, Sang-Eun & Sim, Donggyu. (2018). Dual-Dense Convolution Network for Change Detection of High-Resolution Panchromatic Imagery. Applied Sciences. 8. 1785. doi: <https://doi.org/10.3390/app8101785>
- [6] Alidoost, Fatemeh, and Hossein Arefi. "A CNN-based approach for automatic building detection and recognition of roof types using a single aerial image." PFG - Journal of Photogrammetry, Remote Sensing and Geoinformation Science 86.5 (2018): 235-248.
- [7] Lee, Jongseok & Wiratama, Wahyu & Lee, Wooju & Marzuki, Ismail & Sim, Donggyu. (2023). Bilateral Attention U-Net with Dissimilarity Attention Gate for Change Detection on Remote Sensing Imageries. Applied Sciences. 13. 2485. doi: <https://doi.org/10.3390/app13042485>
- [8] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in Proc. Int. Conf. Learn. Representations (ICLR), 2015. Available: <http://arxiv.org/abs/1409.1556>
- [9] W. G. C. Bandara and V. M. Patel, "A Transformer-Based Siamese Network for Change Detection," IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022, pp. 207-210. doi: <https://doi.org/10.1109/IGARSS46834.2022.9883686>
- [10] H. Chen, Z. Qi and Z. Shi, "Remote Sensing Image Change Detection With Transformers," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-14, 2022, Art no. 5607514.

- doi: <https://doi.org/10.1109/TGRS.2021.3095166>
- [11] C. Han, C. Wu, H. Guo, M. Hu, J. Li and H. Chen, "Change Guiding Network: Incorporating Change Prior to Guide Change Detection in Remote Sensing Imagery," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 8395-8407, 2023.
doi: <https://doi.org/10.1109/JSTARS.2023.3310208>
- [12] H. Chen, J. Song, C. Han, J. Xia and N. Yokoya, "ChangeMamba: Remote Sensing Change Detection With Spatiotemporal State Space Model," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1-20, 2024, Art no. 4409720.
doi: <https://doi.org/10.1109/TGRS.2024.3417253>
- [13] Hye-jeong Choi, Jangsoo Park, Seoung-Jun Oh, & Donggyu Sim (2024-11-20). Change Detection Method using Depth Information in Remotely Sensed Images. *Proceedings of the Korean Society of Broadcast Engineers Conference*, Seoul.
- [14] L. Yang, B. Kang, Z. Huang, X. Xu, J. Feng, and H. Zhao, "Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024.
- [15] Hye-jeong Choi, Jangsoo Park, Seoung-Jun Oh, & Donggyu Sim (2025-06-22). Pseudo Depth Information Prediction Network for Improving Building Change Detection. *Proceedings of the Korean Society of Broadcast Engineers Conference*, Jeju.
- [16] S. Ji, S. Wei, and M. Lu, "Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set." *IEEE Transactions on Geoscience and Remote Sensing*, Vol.57, No.1, pp.574-586, 2018.
- [17] Lee, Wooju, Donggyu Sim, and Seoung-Jun Oh. 2021. "A CNN-Based High-Accuracy Registration for Remote Sensing Images" *Remote Sensing* 13, no. 8: 1482.
doi: <https://doi.org/10.3390/rs13081482>
- [18] Wang, Lukang, Min Zhang, Xu Gao, and Wenzhong Shi. 2024. "Advances and Challenges in Deep Learning-Based Change Detection for Remote Sensing Images: A Review through Various Learning Paradigms" *Remote Sensing* 16, no. 5: 804.
doi: <https://doi.org/10.3390/rs16050804>
- [19] Vinholi, J.G.; Palm, B.G.; Silva, D.; Machado, R.B.; Pettersson, M.I. Change Detection Based on Convolutional Neural Networks Using Stacks of Wavelength-Resolution Synthetic Aperture Radar Images. *IEEE Trans. Geosci. Remote Sens.* 2022, 60, 5236414.
- [20] CHOI, Hansol; LEE, Jongseok; SIM, Donggyu. Dense-Depth Map Estimation with LiDAR Depth Map and Optical Images based on Self-Organizing Map. *Journal of Broadcast Engineering*, 2021, 26.3: 283-295.
- [21] W. Yuan, X. Gu, Z. Dai, S. Zhu, and P. Tan, "Neural window fully-connected CRFs for monocular depth estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 3916 - 3925.
- [22] Guo H, Zhu H, Peng S, Lin H, Yan Y, Xie T, Wang W, Zhou X, Bao H., "Multi-view Reconstruction via SfM-guided Monocular Depth Estimation," 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2025, pp. 5272-5282.
doi: <https://doi.org/10.1109/CVPR52734.2025.00497>
- [23] Lopes, Alexandre, Roberto Souza, and Helio Pedrini. "A survey on RGB-D datasets." *Computer Vision and Image Understanding* 222 (2022): 103489.
- [24] Uchitha Rajapaksha, Ferdous Sohel, Hamid Laga, Dean Diepeveen, and Mohammed Bennamoun. 2024. Deep Learning-based Depth Estimation Methods from Monocular Image and Videos: A Comprehensive Survey. *ACM Comput. Surv.* 56, 12, Article 315 (December 2024), 51 pages.
doi: <https://doi.org/10.1145/3677327>
- [25] Imambi, Sagar, Kolla Bhanu Prakash, and G. R. Kanagachidambaresan. "PyTorch." *Programming with TensorFlow: solution for edge computing applications*. Cham: Springer International Publishing, 2021. 87-104.
- [26] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2019. Available: <https://openreview.net/forum?id=Bkg6RiCqY7>
- [27] I. Loshchilov and F. Hutter, "SGDR: Stochastic Gradient Descent with Warm Restarts," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2017. Available: <https://openreview.net/forum?id=Skq89Scxx>
- [28] D. Kim and M. Kim, "SGSR: A Saliency-Guided Image Super-Resolution Network," 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 980-984.
doi: 10.1109/ICIP49359.2023.10222146
- [29] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2015. Available: <http://arxiv.org/abs/1412.6980>
- [30] S. Fang, K. Li, J. Shao, and Z. Li, "SNUNet-CD: A densely connected Siamese network for change detection of VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1 - 5, 2022.
doi: <https://doi.org/10.1109/TGRS.2024.3367948>
- [31] Dong, Sijun & Wang, Libo & Du, Bo & Meng, Xiaoliang. (2024). ChangeCLIP: Remote sensing change detection with multimodal vision-language representation learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. 208. 53-69.
doi: <https://doi.org/10.1016/j.isprsjprs.2024.01.004>
- [32] Zhu, Taojun & Zhao, Zikai & Xia, Min & Huang, Junqing & Weng, Liguang & Hu, Kai & Lin, Haifeng & Zhao, Wenyu. (2025). FTA-Net: Frequency-Temporal-Aware Network for Remote Sensing Change Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. PP. 1-14.
doi: <https://doi.org/10.1109/JSTARS.2025.3525595>
- [33] H. Zhang, K. Chen, C. Liu, H. Chen, Z. Zou and Z. Shi, "CDMamba: Incorporating Local Clues Into Mamba for Remote Sensing Image Binary Change Detection," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1-16, 2025, Art no. 4405016.
doi: <https://doi.org/10.1109/TGRS.2025.3545012>
- [34] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.

[35] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[36] Ranftl, René, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. "Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer." IEEE transactions on pattern analysis and machine intelligence 44, no. 3 (2020): 1623-1637.

저 자 소 개



최 혜 정

- 2024년 2월 : 고려대학교 데이터사이언스학부 학사
- 2024년 9월 ~ 현재 : 광운대학교 컴퓨터공학과 석사과정
- ORCID : <https://orcid.org/0009-0008-6901-8299>
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전



박 장 수

- 2016년 8월 : 광운대학교 컴퓨터공학과 학사
- 2019년 2월 : 광운대학교 컴퓨터공학과 석사
- 2019년 8월 ~ 2023년 1월 : 아이브스 주식회사 선임연구원
- 2023년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 박사과정
- ORCID : <https://orcid.org/0009-0005-9312-6781>
- 주관심분야 : 영상신호처리, 고성능컴퓨팅



오 승 준

- 1980년 2월 : 서울대학교 전자공학과 학사
- 1980년 2월 : 서울대학교 전자공학과 학사
- 1988년 5월 : 미국 Syracuse University 전기/컴퓨터공학과 박사
- 1982년 3월 ~ 1992년 8월 : 한국전자통신연구원 멀티미디어연구실 실장
- 1986년 7월 ~ 1986년 8월 : NSF Supercomputer Center 초청 학생연구원
- 1987년 5월 ~ 1988년 5월 : Northeast Parallel Architecture Center 학생연구원
- 1992년 3월 ~ 1992년 8월 : 충남대학교 컴퓨터공학부 겸임교수
- 2002년 3월 ~ 2017년 12월 : SC29-Korea 전문위원회 대표위원
- 2001년 8월 ~ 2022년 3월 : MPEG 뉴미디어 포럼 부의장
- 2023년 3월 ~ 현재 : 광운대학교 산학협력단 연구교수
- ORCID : <https://orcid.org/0000-0002-5036-3761>
- 주관심분야 : 비디오 데이터 압축, 머신러닝/딥러닝 기반 컴퓨터비전



심 동 규

- 1993년 2월 : 서강대학교 전자공학과 공학학사
- 1995년 2월 : 서강대학교 전자공학과 공학석사
- 1999년 2월 : 서강대학교 전자공학과 공학박사
- 1999년 3월 ~ 2000년 8월 : 현대전자 선임연구원
- 2000년 9월 ~ 2002년 3월 : 바로비전 선임연구원
- 2002년 4월 ~ 2005년 2월 : University of Washington Senior research engineer
- 2005년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 교수
- ORCID : <https://orcid.org/0000-0002-2794-9932>
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전