ISSN 1229-1889(Print)
ISSN 2287-9005(Online)
J Korean Soc Qual Manag Vol. 53, No.3: 315-328, September 2025
https://dx.doi.org/10.7469/JKSQM.2025.53.3.315

딥러닝을 이용한 항공영상 기반 무허가 건축물 탐지 알고리즘 비교 연구

손현* · 김동호**†

* 숭실대학교 일반대학원 IT정책경영학과
** 숭실대학교 글로벌미디어학부

Deep Learning-Based Detection of Unauthorized Buildings Using Aerial Images: A Comparative Study

Son, Hyun* · Kim, Dong ho**†

* Department of IT Policy and Management, Graduate School of Soongsil University
** Global School of Media, Soongsil University

ABSTRACT

Purpose: This study aims to identify the most effective deep learning algorithm for detecting unauthorized buildings in aerial images by comparing the performance of various image recognition models. While previous research primarily utilized YOLO-based object detection systems, those studies did not analyze algorithmic differences in depth and thus lacked insight into selecting the optimal model for practical administrative use.

Methods: Using aerial images labeled with unauthorized buildings and their corresponding historical images, this study compares CNN-based AlexNet, ResNet (ResNet18, ResNet50, ResNet152), and Transformer-based Vision Transformer (ViT) models. Given the limited number of annotated images, Image Augmentation techniques were applied to expand the training dataset. Pre-trained models were fine-tuned for the task, and training was optimized using Early Stopping and a ReduceLROnPlateau scheduler to prevent overfitting.

Results: Among the evaluated models, ViT achieved the highest prediction accuracy of 98.66%, completing fine-tuning in only 9 epochs. All models demonstrated high reliability, with ROC curves yielding AUC scores of 1.0. ViT's superior performance is attributed to its global visual attention mechanism, which is particularly effective in detecting structural changes across entire aerial scenes.

Conclusion: This study provides a comparative analysis of deep learning models for detecting unauthorized construction, offering valuable insights into selecting practical and accurate algorithms for urban planning and building regulation. ViT, in particular, shows strong potential for real-world deployment due to its high precision and efficient training characteristics, contributing to automated building monitoring and improved public safety.

Key Words: Unauthorized Buildings, Aerial Imagery, Deep Learning, Vision Transformer, Image Recognition, Fine-Tuning

● Received 17 June 2025, 1st revised 19 June 2025, accepted 27 June 2025
† Corresponding Author(dkim@ssu.ac.kr)
ⓒ 2025, The Korean Society for Quality Management
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.

1. 서 론

현재 우리나라는 지진과 화재 등으로 인해 안전 문화에 관한 관심이 크게 증가하고 있으며, 자연재해로 인한 붕괴 및 화재 위험이 커지면서 무허가 증축으로 인해 건물의 수평 이동, 침하, 기울임 등이 발생하여 국민의 안전을 위협하는 무허가 건축물이 증가하고 있는 추세이다. <Figure 1>은 2016년 서울시 기준으로 무허가 건축 이행 강제금 부과 건수는 46,052건에 달하며, 이러한 건수는 2014년 이후 계속 증가하는 사실을 보여 준다(Seoul City, 2016). 무허가 건축물에 대한 단속은 민원 신고, 항공사진 촬영 후 육안으로 구분하여 현장을 방문하는 방식으로 이루어지지만, 항공사진은 육안으로 구분하기 어려운 데다 가용 인원이 제한되어 있어 현실적인 한계가 존재한다. 또한, 항공사진을 통한 구분 역시 사람의 자의적 판단이 개입될 수 있어 정확한 처리가 어려운 상황이다. 현재의 일반적인 항공사진 분석 프로그램은 정확도 면에서 만족스러운 효과를 내지 못하고 있어 참고용으로만 사용되고 있으며, 무허가 건축물 및 불법건축물은 지속적으로 증가하고 있다. 이러한 문제를 해결하기 위해 본 연구는 딥 러닝 기술을 항공사진 분석에 접목하여 과거의 항공사진을 반복 학습시키고, 딥 러닝을 통해 추출된 위치 정보를 화면에 표시함으로써 항공사진 비교 분석의 정확성을 높이고자 한다. 이를 통해 무허가 건축물의 위치를 효과적으로 파악하고, 건축물대장 정보를 출력하여 정보 습득이 용이해지며, 가용 인원의 제한을 극복할 수 있을 것으로 기대된다. 이 시스템은 담당 공무원의 무허가 건축물 및 불법건축물에 대한 의사 결정 지원 도구로서 역할을 할 것이며, 궁극적으로 무허가 건축물 감소와 국민 안전 증진에 크게 이바지할 것으로 기대된다.

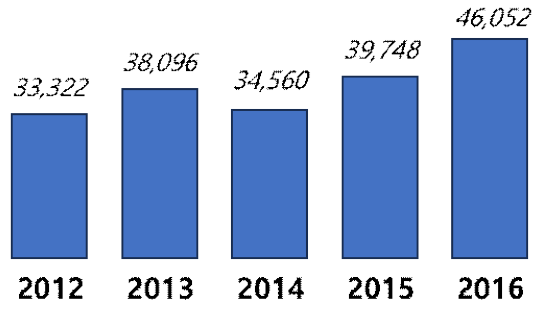


Figure 1. Number of cases of imposition of fines for unauthorized construction

기존 연구(Son and Hyun, 2022)는 YOLO 기반 특정 이미지 검출 및 이를 기반으로 한 무허가 건물 탐지 시스템 구축에 중점을 두었다. YOLO 기반으로 이미지 검출을 하는 실험에서는 알고리즘 종류에 대해서는 별로 관심을 갖지 않았고 이미 구축된 모델을 기반으로 학습 및 검출이 잘되는지 여부에만 중점을 둔 관계로 어떤 알고리즘이 무허가 건물 예측에 잘 맞는지 알 길이 없었다. 논문 게재 이후 이런 부분에 대한 궁금증을 갖게 됐고, 이후 이미지 인식이 가능한 다양한 딥러닝 알고리즘을 접하게 됐고 어떤 알고리즘이 무허가 건물을 과연 잘 예측하는지 확인해 보면 의미가 있을 것이라는 착안을 하게 됐다. 따라서 본 논문은 무허가 레이블링이 돼 있는 몇 장의 항공 사진과 해당 사진의 과거 항공 사진을 이용해 몇가지 이미지 인식 딥러닝 알고리즘을 이용해 학습해 보고 각 알고리즘의 성능 분석 결과를 제시 한다. 항공 사진 수가 적은 관계로 Image Augmentation 기법을 활용해 학습 및 검증용 이미지 수를 충분히 늘리고, 적절한 하드웨어의 준비가 안 될 경우 딥러닝 자체가 어려운 알고리즘을 제외하고 실험실에서 딥러닝 학습이 가능한 모델로 한정해 실험을 진행한다. 이 논문에서는 CNN 기반인 AlexNet, 잔차 학습 기반인 ResNet과 Transformer 기반인 ViT(Vision Transformer)를 이용해 예측 결과를 비교한다.

2. 문헌 연구

2.1 딥 러닝

딥 러닝은 여러 계층(layer)을 통해 초기 입력으로부터 점진적으로 더 복잡한 수준의 특성을 추출하는 머신러닝 알고리즘의 일종이다(Deng et al., 2014). 예를 들어, 이미지 처리에서 하위 계층은 가장자리와 같은 저수준 특징을 식별하며, 상위 계층은 숫자, 문자, 얼굴 등 고수준 객체를 인식할 수 있다. 딥 러닝에서 '딥(Deep)'이라는 용어는 처리 계층의 수를 의미하며, 이는 입력에서 출력까지 변환되는 경로인 크레딧 할당 경로(Credit Assignment Path, CAP)의 깊이로 정의된다. 피드 포워드 신경망의 경우, CAP의 깊이는 은닉 계층 수에 1을 더한 값과 같으며, 순환 신경망에서는 무제한으로 간주할 수 있다(Schmidhuber, 2015). CAP 깊이 2 이상은 딥 러닝으로 간주하며, 더 깊은 모델은 일반 모델에 비해 더 복잡하고 유용한 특성을 학습할 수 있는 능력을 제공한다.

딥 러닝 알고리즘에는 다양한 심층 신경망 구조가 존재하며, 그중 대표적인 구조는 심층 신경망(Deep Neural Network, DNN), 합성곱 신경망(Convolutional Neural Network, CNN), 순환 신경망(Recurrent Neural Network, RNN), 제한 볼츠만 머신(Restricted Boltzmann Machine, RBM), 심층 신뢰 신경망(Deep Belief Network, DBN), 심층 Q-네트워크(Deep Q-Networks, DQN) 등이 있다(Ciresan. D. et al., 2012; Krizhevsky et al., 2012).

심층 신경망(DNN)은 입력과 출력 계층 사이에 여러 은닉 계층을 포함하여 복잡한 분류 문제를 효과적으로 모델링할 수 있는 인공 신경망이다(Bengio et al., 2009). 은닉 계층을 통해 점진적으로 특징을 추출하며, 피드 포워드 신경망으로 설계되지만 순환 신경망(RNN) 구조에서도 활용된다(Sainath et al., 2013).

합성곱 신경망(CNN)은 시각적 이미지 처리에 특화된 다층 피드 포워드 신경망으로, 필터링 기법을 통해 이미지를 분류하고 분석한다. CNN은 계층적 패턴 인식을 통해 정규화된 다층 퍼셉트론의 구조적 복잡도를 줄이며, 데이터 과적합을 방지하기 위해 연결 구조를 단순화한다(Architectural Institute of Korea, 2020).

순환 신경망(RNN)은 내부 메모리를 통해 시간적 데이터 시퀀스를 처리할 수 있는 신경망 구조로, 음성 및 필기 인식과 같이 시간이 변하는 데이터를 분석하는 데 유용하다(Dupond et al., 2019; Abiodun et al., 2018; Tealab et al., 2018). RNN은 유한 및 무한 임펄스 응답 구조로 구분되며, 장단기 메모리(Long Short-Term Memory, LSTM)와 게이트 순환 유닛(Gate Circuit Unit) 등 게이트 메모리 구조를 통해 학습 성능을 높인다(Hyötyniemi et al., 1996; Sepp Hochreiter et al., 1997).

제한 볼츠만 머신(RBM)은 가시 유닛과 은폐 유닛으로 구성된 무방향 이분 그래프로, 은닉 유닛을 MCMC(Markov Chain Monte Carlo) 및 CD(Contrastive Divergence) 기법으로 추정한다. 이 구조는 간결하면서도 샘플링 기반 학습법을 통해 복잡한 데이터 모델링을 가능하게 한다(Sherrington et al., 1975).

심층 신뢰 신경망(DBN)은 비지도 학습을 통해 계층별 학습이 이루어지는 생성 모델로, 훈련 데이터가 적은 상황에서도 선행 학습을 통해 성능을 높일 수 있다. DBN은 통상적으로 제한 볼츠만 머신의 형태를 취하며, 각 계층이 제한 볼츠만 머신으로 구성된다(Hinton G, 2009).

심층 Q-네트워크(DQN)는 강화 학습의 Q 값을 딥 러닝을 통해 학습하는 구조로, CNN을 이용해 픽셀 데이터를 통해 게임을 학습할 수 있으며, 다양한 게임에 하나의 Agent로 대응할 수 있도록 설계되었다(Mnih. V et al., 2013; X. Xiong et al., 2020; L. Liang et al., 2020).

2.2 이미지 인식 딥러닝 알고리즘

이미지 분류를 위한 CNN 구조의 딥러닝 알고리즘으로 AlexNet, VGGNet, ResNet 등이 있다. CNN 기반 이미지 분류에서 획기적인 성능을 보인 AlexNet은 Relu 활성화 함수와 GPU 병렬 처리를 통해 대규모 이미지넷 데이터셋에서 성공을 거두었다(Seungzoon Lee et al., 2023). VGGNet은 단순한 3x3 필터 구조를 반복적으로 쌓아 깊은 신경망을 구성함으로써 이미지 분류의 정확도를 향상시켰다(Simonyan et al., 2014). ResNet은 잔차 학습(residual learning) 구조를 도입하여 네트워크가 깊어져도 성능이 감소하지 않도록 하였으며, 이미지 분류 및 탐지에서 광범위하게 사용되고 있다(Myung Hwan Na et al., 2020).

위 알고리즘의 경우 인식률은 향상시킬 수 있으나 신규 이미지를 이용해 딥러닝하려면 고비용이 발생하는 관계로 경량화 알고리즘이 고안됐고 해당 알고리즘으로 MobileNet, EfficientNet이 있다. MobileNet은 깊이별 분리 합성곱(depthwise separable convolution)을 도입하여 모바일 환경에서도 효율적으로 이미지 인식이 가능하도록 경량화된 모델을 제안하였다(Howard et al., 2017). EfficientNet은 모델의 깊이, 너비, 해상도를 동시에 최적화하는 복합 스케일링 방식을 통해 높은 정확도와 효율을 동시에 달성하였다(Tan et al., 2019).

CNN 구조가 아닌 Transformer 구조를 이용한 이미지 인식 알고리즘도 등장했고 ViT, Swim이 이에 해당한다. Vision Transformer(ViT)는 이미지를 패치 단위로 나눈 후 순차 데이터를 처리하는 Transformer 구조로 학습하여 CNN 없이도 높은 이미지 분류 성능을 달성하였다(Dosovitskiy et al., 2020). Swin Transformer는 윈도우 기반의 로컬 어텐션과 계층 구조를 통해 다양한 크기의 시각 정보를 효과적으로 처리할 수 있는 Transformer 모델이다(Liu et al., 2021).

위 알고리즘을 이용한 실제 응용 라이브러리로 YOLOv4, SAM이 있다. YOLOv4는 실시간 객체 탐지에서 속도와 정확도를 균형 있게 향상시키기 위해 다양한 최적화 기법을 통합하였다(Bochkovskiy et al., 2020). Segment Anything Model(SAM)은 사용자의 입력 없이도 이미지 내의 다양한 객체를 자동으로 분할할 수 있는 범용 세그멘테이션 모델을 제안하였다(Kirillov et al., 2023).

2.3 GIS 관련 연구

피해조사 자동화 시스템, 기상 시스템, 수도 조사 시스템, 국가 공간정보 종합시스템 등과 유사한 연구에는 이미지 분석, 스펙트럼 분석 기술, 측지측량 기반의 좌표계 및 좌표 보정, OGC 표준을 적용한 GIS 관련 연구 등이 포함된다. <Table 1>과 같이 최연오 외(2023)에서는 위성 영상에서 3차원 건물을 추출하기 위한 Frame Field Learning, PolyWorld, HiSup 알고리즘의 성능을 비교한 후 Frame Field Learning이 3차원 건물을 추출하기 위한 최적의 알고리즘으로 제시했다(Yeon Oh Choi et al., 2023).

김종우 외(2017)에서는 특수 카메라를 설치한 드론으로 구조물의 외관을 촬영하고 전처리 과정을 거친 후 80장의 이미지에 결함 여부를 표시 후 머신러닝을 해 균열이 0.3 mm인 경우 손상 검출률이 90.5%이고, 0.3mm 미만인 경우 손상 검출률이 75.0%임을 확인했다(Jong-Woo Kim et al., 2017).

김동은(2024)은 기존의 합성곱 신경망(CNN) 기반 모델과 달리 국내 항공 영상 기반 건물의 변화 탐지 학습 데이터를 구축하고 멀티스케일의 장거리 의존성을 효과적으로 포착할 수 있는 ChangeFormer를 활용해 국내에서 수집된 항공영상 데이터를 활용하여 학습시키고, 이를 통해 건물 변화 탐지의 정확도와 신뢰도를 향상시켰다(Dong-eun Kim, 2024).

Teo 외(2025)에서는 FT-UNetFormer라는 딥러닝 모델을 활용하여 다중 시점의 RGB 이미지, DSM, 기존 건물 폴리곤 데이터를 통합하여 건물 변화 탐지를 수행했다(Tee-Ann Teo et al., 2025).

Qinglie Yuan(2025)는 Transformer와 CNN을 결합한 하이브리드 네트워크를 통해 다양한 크기와 형태의 지붕을 정확하게 추출하는 방법을 제안했다. 제안된 모델은 WHU 및 Massub 데이터셋에서 높은 F1 점수를 기록했다(Qinglie Yuan, 2025).

본 연구는 딥러닝을 활용한 항공영상 기반 무허가 건축물 탐지를 목표로 하며, 기존 관련 연구들과 다음과 같은 차별점을 가진다. 첫째, 기존 연구들이 건물 변화 탐지(Teo et al., 2025; Kim, 2024) 또는 지붕 객체 추출(Yuan, 2025), 3차원 건물 모델링(Choi et al., 2023) 등 물리적 변화 또는 객체 검출에 집중한 데 비해, 본 논문은 '무허가 건축물'이라는 행정적·법적 개념을 기준으로 탐지하는 데 초점을 맞추었다. 즉, 단순히 건물의 변화 여부를 식별하는 것이 아니라, 항공사진에서 추출된 건물 위치 정보를 건축물대장 등 행정 DB와 연계하여 허가되지 않은 불일치 건축물의 존재 여부를 판별하고자 한다. 둘째, 기존 연구들이 RGB 영상, DSM, 다시점 항공사진 등의 정밀한 영상 데이터를 주로 활용한 데 반해(Teo et al., 2025; Yuan, 2025), 본 연구는 항공영상 분석 결과와 함께 건축물 행정 정보를 결합하는 방식을 통해 도시 행정 실무에 실질적으로 활용 가능한 자동화 탐지 시스템을 구현한다. 이러한 데이터 융합 방식은 공무원의 의사결정을 보조하고, 인력 부족이나 육안 탐지의 한계 등을 극복할 수 있는 현실적인 해법을 제시한다는 점에서 차별성을 지닌다. 셋째, 본 논문은 한국 실정에 맞는 항공영상과 무허가 건축물 사례를 기반으로 데이터셋을 직접 수집·라벨링하여 한국 도시 환경과 제도에 특화된 모델을 학습시켰다. 김동은(2024)의 연구가 국내 항공영상을 활용한 변화 탐지 모델을 제시하였으나, 변화 유무에 대한 일반적 분석에 그친 반면, 본 연구는 건축물대장과 실존 건축물 간의 불일치 사례를 활용하여 무허가 건축물 여부를 판단하는 실질적인 탐지 모델을 구현한다.

Table 2. GIS-related research

Subject	Detail
Choi et al.,(2023)	* After comparing the performance of Frame Field Learning, PolyWorld, and HiSup algorithms for extracting 3D buildings from satellite images, Frame Field Learning is presented as the optimal algorithm for extracting 3D buildings.
Kim et al.,(2017)	* The exterior of the structure was photographed with a drone equipped with a special camera, and after a preprocessing process, 80 images were marked for defects and machine learning was performed to confirm that the damage detection rate was 90.5% when the crack was 0.3 mm, and 75.0% when it was less than 0.3 mm.
Kim(2024)	* Trained ChangeFormer using aerial image data collected in Korea, thereby improving the accuracy and reliability of building change detection.
Teo et al.,(2025)	* Performed building change detection by integrating multi-view RGB images, DSM, and existing building polygon data using a deep learning model called FT-UNetFormer.
Yuan(2025)	* Proposed a method to accurately extract roofs of various sizes and shapes using a hybrid network combining Transformer and CNN.

3. 연구 방법

3.1 GIS 지도 무허가 건축물 비교 분석 환경

<Figure 2>는 GIS 기반 지도 비교 분석 환경을 나타낸다. 분석 환경은 과거 항공사진을 입력으로 받아 이미지 처리 및 지적도와 항공사진의 병합(Merging) 작업을 수행한 후, 딥 러닝 학습을 통해 모델을 구축한다. 이후 새로운 항공사진에서 해당 건물을 식별하고, 건물의 변화 상태를 분석하여 결과를 출력하는 방식으로 동작한다.

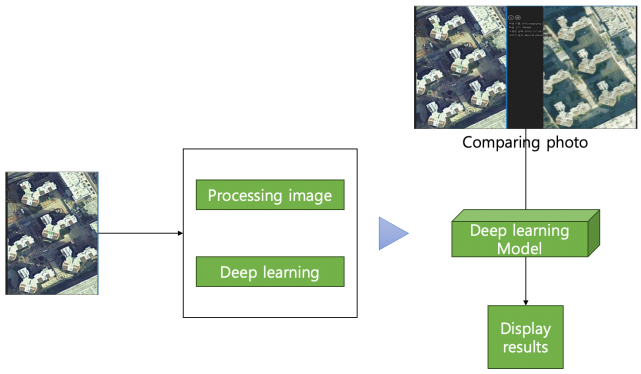


Figure 2. GIS Map Comparison Analysis Environment

3.2 GIS 지도 무허가 건축물 비교 분석 처리 절차

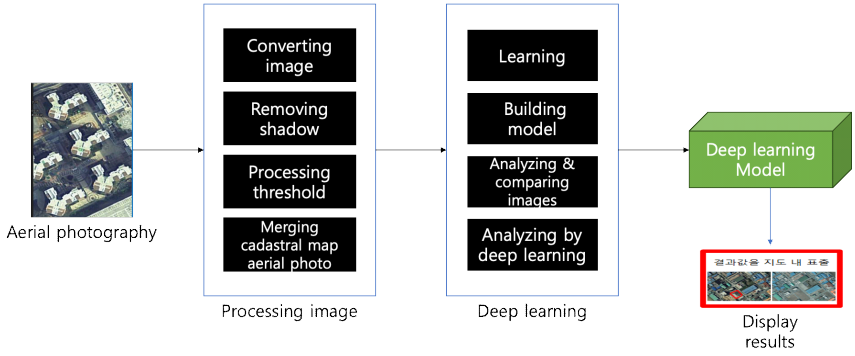


Figure 3. GIS Map Comparison Analysis Processing Procedure

<Figure 3>은 GIS 지도 비교 분석 환경의 처리 절차를 나타낸다. 본 분석 환경은 항공사진에 대한 이미지 처리 작업과 딥 러닝 작업으로 구성된다. 이미지 처리 작업은 Image Augmentation 기법을 이용해 학습할 이미지 수를 늘리고, 영상 이미지의 변환, Gray Scaling을 통한 그림자 제거, 이미지에 대한 임계치(Thresholding) 처리, 항공영상 이미지와 지적도 및 건축물대장 정보를 연계하여 병합(Merging)하는 단계로 이루어진다. 딥 러닝 작업은 몇몇

신경망을 이용한 신규 데이터 학습, 딥 러닝 모델 생성, 생성된 모델을 활용한 이미지 비교 분석, 분석 결과 도출 순으로 진행된다. 이미지 분석은 픽셀 단위로 비교하며, 단순 비교 결과는 딥 러닝 모델과 연계하여 최종 분석 결과를 지도 플랫폼으로 전송한다.

3.2.1 이미지 처리

<Table 2>는 GIS 지도 비교 분석의 처리 절차에서 개발이 요구되는 사항을 정리한 것이다. 이미지 처리에는 연계된 항공 영상 데이터를 이미지로 변환하는 알고리즘 개발, 시간, 날씨, 계절 등에 따른 RGB 차이로 인해 무허가 건축물로 잘못 인식되는 오류를 최소화하기 위한 Thresholding 알고리즘 개발, 건물 및 환경에 따른 그림자를 무허가 건축물로 오인하는 오류를 줄이기 위한 Gray Scale 알고리즘 개발, 그리고 항공사진과 지적도를 병합하여 건물 경계를 정형적으로 표기하는 Merging 기법의 개발이 포함된다.

항공 영상 이미지 변환 알고리즘은 연도별로 구분된 항공 영상에서 화면에 사용된 이미지 목록을 추출하고 해당 데이터를 파라미터로 취합한 뒤 Ajax를 통해 Java 컨트롤러로 전송하는 방식으로 설계된다. 이를 통해 각 이미지 목록을 활용하여 해당 타일의 위치를 계산하고, 해당 위치에 항공 영상 이미지를 배치한 후 최종적으로 완성된 항공 영상을 저장하는 프로세스를 수행하는 알고리즘을 개발하는 것이 목표이다.

Gray Scale 처리를 위해 OpenCV(Open Computer Vision) 라이브러리를 활용한다. 이미지 변환을 통해 완성된 RGB 항공 영상을 MAT class로 변환 및 기본 Gray Scale 알고리즘을 사용하여 변환한다. 그림자 처리를 위한 Thresholding 알고리즘은 Gray Scale이 완료된 후 그림자 처리를 위해 사용하며 컬러 이미지를 0/1로 변환하는 이진화 기법이다. OpenCV 라이브러리에서 제공하는 Threshold 메소드를 사용한다. Merging 작업은 완성된 이미지를 지도 화면에 표출하는 작업이며 OpenLayers의 ImageLayer를 사용한다.

Table 2. Development of GIS map comparison analysis processing procedure

Procedure	Detail
processing images	* Development of an algorithm to convert aerial image data into images * Development of a Gray Scale algorithm to minimize errors in recognizing shadows as unauthorized buildings due to differences in shadows that occur depending on buildings, weather, and time * Development of a Thresholding algorithm to minimize errors in recognizing unauthorized buildings due to differences in RGB caused by time, weather, and season in converted aerial imagery * Merging work to formally mark building boundaries using a filtering technique that combines aerial photographs and cadastral maps
deep learning	* Development of an algorithm to learn the image processing stage result storage location and the image processing stage result image * Development of an algorithm to create a model using learning data * Development of an analysis algorithm to utilize the created model
displaying the result	* Development of an algorithm that compares each processed image pixel by pixel, links the simple comparison results to a deep learning platform, and transmits the analysis results to a map platform.

3.2.2 딥 러닝 처리 단계

딥 러닝 플랫폼 구축을 위해서는 이미지 가공 단계의 결과 저장 위치와 가공된 이미지를 학습하기 위한 알고리즘 개발이 필요하다. 또한, 학습 데이터를 기반으로 모델을 생성하는 알고리즘 개발과, 생성된 모델을 활용한 분석 알고리즘 개발이 요구된다. 딥 러닝 알고리즘의 정밀도는 1에 가까울수록 학습의 정밀도가 높은 것으로 해석되며, 이는 모델의 정확성을 평가하는 중요한 지표로 작용한다.

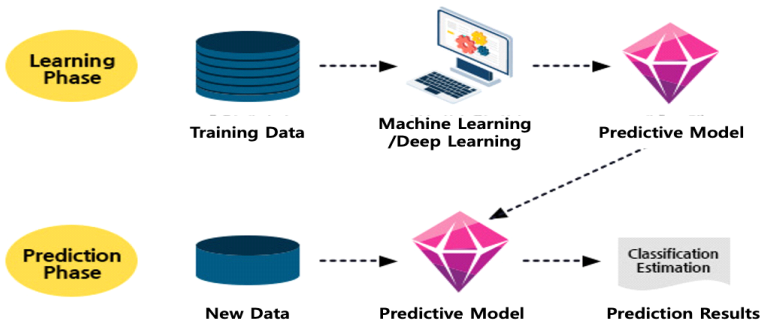


Figure 3. Deep learning process

<Figure 3>은 딥 러닝 프로세스이다. 학습 단계에서 훈련 데이터를 이용해 머신러닝/딥 러닝을 한 후 생성된 예측 모델을 이용하여 신규 데이터의 예측에 활용하여 분류가 정상인지를 확인한다.

본 연구에서는 PyTorch를 활용한 합성곱 신경망(Convolutional Neural Network, CNN) 알고리즘 기반의 AlexNet, 잔차 학습을 이용한 ResNet을 활용해 이미지 학습을 수행하고, Transformer 기반의 ViT(Vision Transformer)를 활용해 이미지 학습을 수행한다. Pytorch는 높은 수준의 딥 러닝 라이브러리로서, TensorFlow 등의 딥 러닝 엔진 위에서 작동하며 사용자 친화적인 인터페이스를 제공하여 모델 개발 및 학습에 용이하다. CNN 알고리즘은 이미지의 공간적, 계층적 패턴을 효과적으로 학습할 수 있는 구조로, 특히 이미지 분류와 객체 인식에서 뛰어난 성능을 보인다. Transformer 알고리즘은 입력 시퀀스 내의 모든 요소 간 관계를 self-attention 메커니즘을 통해 병렬적으로 계산하여, 장기 의존성과 문맥 정보를 효과적으로 학습할 수 있는 딥러닝 모델 구조이다(Vaswani et al., 2017). Vision Transformer(ViT)는 이미지를 고정 크기의 패치로 분할한 후 이를 시퀀스로 변환하여, Transformer의 self-attention 메커니즘을 통해 전역적인 시각 정보를 학습하는 이미지 인식 모델이다(Dosovitskiy et al., 2020).

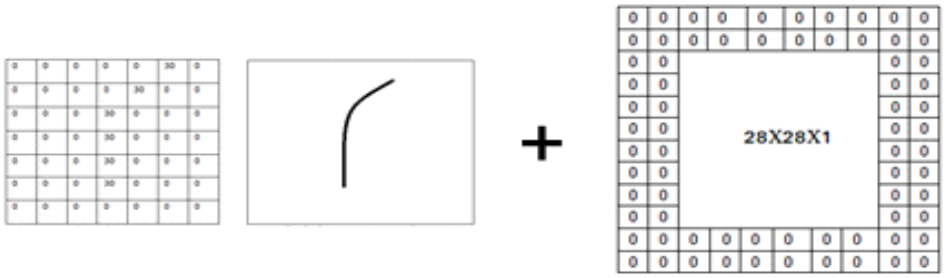


Figure 4. Aerial image comparison analysis process

<Figure 4>는 항공 이미지의 비교 분석 프로세스를 나타낸다. 본 시스템은 지도 비교 기능을 제공하며, 사용자 인터페이스에서 특정 주소를 검색하고 화면 분할 기능을 통해 해당 주소의 연도별 항공 영상을 시각화하여 건물 정보와 연도별 항공 영상을 육안으로 비교 분석할 수 있는 서비스 시스템이다. 프로세스는 필터 단계와 패딩 단계로 구성된다. 필터 단계는 그림자 및 RGB 처리 등 오류를 최소화하기 위한 전처리 과정을 포함하며, 패딩 단계는 일치하지 않는 부분을 탐지하고 저장하는 역할을 수행한다.

3.2.3 결괏값 표출

결과 출력을 위해서는 가공된 각 이미지를 픽셀 단위로 비교하고, 단순 비교 결과를 딥 러닝 플랫폼과 연계하여 분석된 결과를 지도 플랫폼으로 전송하는 알고리즘 개발이 필요하다. <Figure 5>는 최종적으로 외곽선만 추출된 결과 이미지의 예시를 나타낸다. 이미지 비교를 위해서는 먼저 Threshold 처리를 수행한 후 두 이미지를 비교하여 차이점만 남도록 처리해야 하며, 이후 Contours 검사를 통해 외곽선만을 추출하는 절차가 필요하다.

Figure 5. Example of an image with outline extraction

4. GIS 지도 무허가 건축물 비교 분석 모델 구현 및 실험

4.1 GIS 지도 무허가 건축물 비교 분석 모델 구현

4.1.1 항공 영상 이미지 처리 모듈

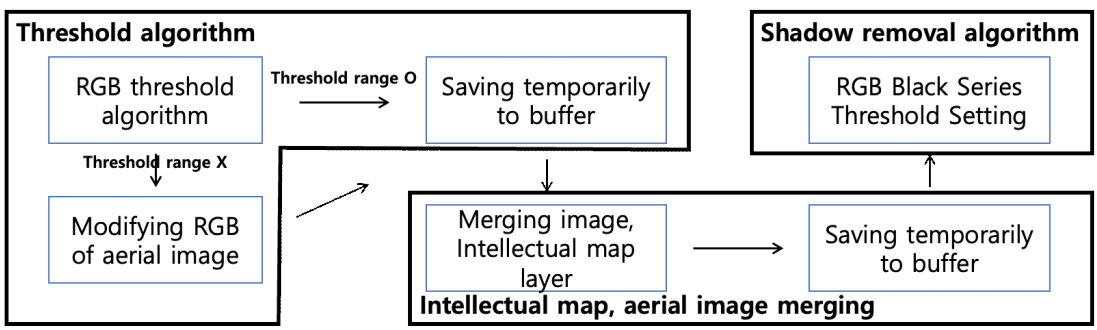


Figure 6. Aerial image conversion procedure

<Figure 6>은 항공 영상 이미지 변환 절차를 나타낸다. 이 절차에서는 RGB 임계치 기준을 설정하여, 항공사진이 임계치 범위 내에 있는 경우 Buffer에 임시 저장하고, 임계치를 초과하면 RGB 값을 조정한 후 Buffer에 저장한다. RGB 임계치 알고리즘을 통해 처리된 항공사진과 데이터 연계 알고리즘을 적용한 지적도 이미지를 중첩하여 Buffer에 임시 저장한 후, 그림자 제거 알고리즘에 전송된다. 시간과 건물에 의해 생성된 그림자를 무허가 건축물로 오인하는 오류를 방지하기 위해, 병합(Merging) 작업을 거친 이미지의 RGB 임계치를 블랙 계열로 설정하여 지적도 상 표시된 경계선 외부의 블랙 계열 부분에 명암 조절을 적용하고, 최종 결과를 Buffer에 저장한다.

항공 영상 변환 알고리즘은 이미지 목록을 추출하여 해당 파라미터와 함께 전송하는 부분과 이미지 목록을 기준으로 저장된 항공 영상 이미지로 구성된다. 이미지 목록을 기준으로 항공 영상 이미지 생성한다. 이미지 목록을 받아서 반복적으로 이미지 원본과 이미지의 좌표 정보를 받아서 이미지 크기를 계산한 후 항공 영상 이미지를 저장한다.

4.1.2 딥 러닝 처리 모델 학습

딥 러닝 처리 모듈은 딥 러닝 활용을 위한 훈련 데이터 생성, 알고리즘 연계, 예측 모델 구축의 단계로 구성된다. 예측 모델 구축을 위해, 먼저 훈련 데이터를 생성하며, 이를 위해 농지와 임야 지역 중 30개를 선정하고, 각 지역의 최근 3년간 항공 이미지를 수집한다. 딥 러닝 기법을 사용하기 위해, 각 항공사진의 크기를 조절하고 레이어를 반복적으로 쌓아 특징 추출(Feature Extraction) 단계에서 임계치 설정, 그림자 제거 알고리즘, 항공 이미지 병합(Merging) 작업을 연계하여 처리한다. 이 과정에서 각 항공 이미지에는 독립적인 연도 코드가 부여되며, 이를 활용하여 테이블을 생성하고, 알고리즘에 의해 추출된 이미지와 코드의 주솟값을 연도별로 정리하여 데이터베이스에 학습용 데이터를 기록한다.

국토교통부 국토정보맵의 공공정보인 항공사진과 지적도를 활용했으며 위와 같이 처리한 후 지적도에 없는 건물을 무허가 건축물로 수작업 라벨링하고, 농지 및 임야 지역 30개소에서 연도별 10개씩 총 300장의 데이터를 확보한 후 Image Augmentation 기법을 적용하여 50,000장의 학습 데이터를 구축한다. 학습은 M4 Pro GPU 환경(20코어)에서 수행하며, 각 모델의 출력 노드를 2로 설정하여 허가 여부를 판별하고, Learning rate 0.00001에서 시작해 ReduceLROnPlateau 스케줄러와 Early Stopping 기법을 적용하여 최적 성능을 도출한다.

4.2 GIS 지도 무허가 건축물 비교 분석 결과

4.2.1 딥 러닝 실행 결과

<Table 3>은 무허가 건물 이미지에 대한 Deep Learning 실행 결과이다. Deep Learning Model로 CNN 기반의 AlexNet, RestNet과 Transformer 기반은 ViT를 활용했다. ResNet은 계층 수에 따라 ResNet18, ResNet50, ResNet152를 활용했다. Deep Learning은 사전 학습 모델을 이용한 Fine Tuning방식으로 진행했고 과적합을 피하기 위해 더 이상의 성능 증가가 없을 때 학습을 종료하는 Early Stopping과 학습 결과가 거의 유사할 경우 Learning Rate를 줄이는 ReduceLROnPlateau 방식의 Scheduler를 활용했다. 이에 따라 AlexNet은 18번, ResNet18은 17번, ResNet50은 18번, ResNet152는 15번, ViT는 9번의 반복 학습이 진행됐다. 학습시간은 26분~4시간 정도였고, 단위 시간으로 보면 AlexNet의 학습시간이 가장 짧았고 ViT의 학습시간이 가장 길었다. AlexNet외에 Accuracy가 90% 이상이 나왔고, ResNet50 부터는 Accuracy가 96% 이상 나왔다. 가장 성능이 좋은 모델은 ViT이고 98.66%가 나왔으며 9번의 학습만으로 Fine Tuning이 완료됐다. 항공 사진의 경우 주변과의 관계 보다는 전체적인 그림에서 변경된 부분을 찾는 것이 용이하므로 전역적인 시각으로 이미지를 인식하는데 탁월한 ViT가 가장 높은 성능이 나온 것으로 추측할 수 있다.

Table 3. Deep Learning Result

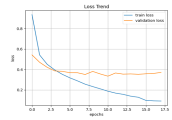
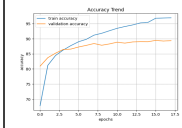
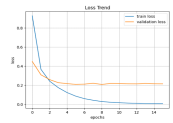
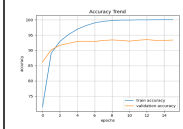
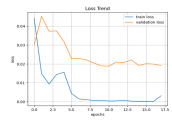
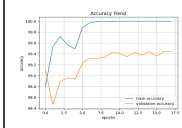
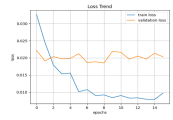
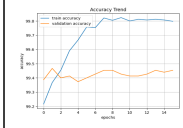
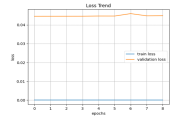
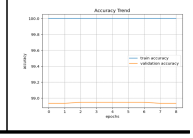
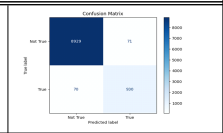
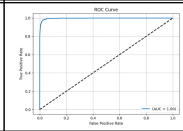
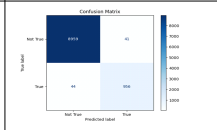
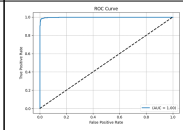
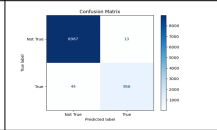
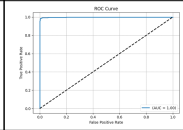
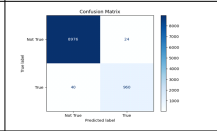
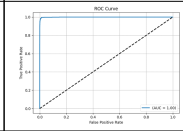
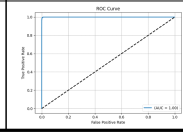
Deep Learning Model	Accuracy	Loss	Epoch	Elapsed Time	Loss Trend	Accuracy Trend
AlexNet	89.26%	0.3672	18	00:26:06		
ResNet18	92.99%	0.2305	17	00:49:01		
ResNet50	96.33%	0.1652	18	02:11:06		
ResNet152	96.20%	0.1322	15	04:05:15		
ViT	98.66%	0.0617	9	02:46:21		

Table 4. Deep Learning Detail Result

Deep Learning Model	Accuracy	Precision	Confusion Matrix		ROC
AlexNet	0.99	0.93	[[8929 71] [70 930]]		
ResNet18	0.99	0.95	[[8959 41] [44 956]]		
ResNet50	0.99	0.98	[[8987 13] [44 956]]		
ResNet152	0.99	0,98	[[8976 24] [40 960]]		
ViT	1.00	0.98	[[8985 15] [18 982]]		

<Table 4>는 각 모델에 대한 Accuracy Test 결과, Precision Score, 혼동 행렬과 ROC 커브 도표이다. 모든 모델의 Accuracy Test점수는 0.99~1.0으로 정확한 탐지가 가능한 것으로 나온다. CNN 기반은 0.99가 나오고 Transformer 기반은 1.0이 나오는 것을 알 수 있다. 학습의 정밀도를 확인하기 위한 Precision Score도 0.93~0.99로 정확한 탐지가 가능한 것으로 예상된다. ROC(Receiver Operating Characteristic curve)의 경우 사각형을 거의 채우는 형상으로 나오고 AUC(Area Under Curve)는 모두 1.0이 나오는 것을 확인할 수 있다.

5. 결 론

기존 연구(Son, Hyun, 2022)는 YOLO 기반의 특정 이미지 객체 검출 기법을 활용하여 무허가 건축물을 탐지하는 시스템을 구현하는 데 초점을 맞추었다. 이 연구는 YOLO 모델의 구조나 알고리즘 성능 차이에 대한 분석보다는, 사전에 구축된 YOLO 모델을 기반으로 탐지 기능이 작동하는지를 확인하는 실험에 집중하였다. 이에 따라 어떤 딥러닝 알고리즘이 무허가 건물 탐지에 가장 적합한지에 대한 분석은 부족한 실정이었다. 본 논문은 이러한 기존 연구의 한계를 인식하고, 무허가 건축물 탐지에 효과적인 딥러닝 알고리즘을 규명하고자 한다. 특히, 무허가 레이블이 부착된 일부 항공사진과 그 과거 영상들을 활용하여 다양한 이미지 인식 기반 딥러닝 모델들의 성능을 비교하고자 하며, 이를 통해 건축 행정 실무에 활용 가능한 고성능 탐지 모델을 제시하고자 한다.

이를 위해 본 연구에서는 CNN 기반의 AlexNet, 잔차 학습 구조를 갖는 ResNet(ResNet18, ResNet50, ResNet152), 그리고 Transformer 구조를 갖는 Vision Transformer(ViT)를 비교 대상으로 선정하였다. 학습 데이터는 국토교통부 국토정보맵에서 제공하는 지리정보 데이터에 무허가 여부가 라벨링된 데이터를 이용했으며, 이미지 갯수가 제한적인 현실을 고려하여 Image Augmentation 기법을 적용하였으며, 실험은 일반적인 실험실 환경에서 재현 가능한 수준의 하드웨어로 수행 가능한 모델들로 한정하였다. 학습 방식은 사전 학습된 모델을 활용한 파인튜닝(Fine-Tuning)을 기반으로 했으며, 과적합을 방지하기 위해 Early Stopping과 ReduceLROnPlateau 스케줄러를 함께 적용하였다. 실험 결과 ViT 모델이 가장 높은 정확도를 보이며(Accuracy: 98.66%), 단 9회의 반복 학습만으로도 안정적인 성능을 확보하였다. 또한 모든 모델이 ROC 커브 상 AUC 1.0을 기록함으로써 높은 신뢰도로 무허가 건물 탐지가 가능함을 입증하였다. 특히, ViT 모델은 이미지 전체를 전역적 시각에서 분석하는 특성 덕분에 건물 변화 탐지에 있어 우수한 성능을 보인 것으로 해석할 수 있다.

REFERENCES

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N., and Arshad, H. 2018. State-of-the-art in Artificial Neural Network Applications: A survey. Heliyon 4(11). DOI:10.1016/j.heliyon.2018.e00938.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, and Ross Girshick. 2023. Segment Anything. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV):4015-4026.

Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao 2004. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2020.

An image is worth 16x16 words: Transformers for image recognition at scale. arXiv:2010.11929.

Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861.

Architectural Institute of Korea. 2020. Online Dictionary of Architecure. Architectural Institute of Korea.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. 31st Conference on Neural Information Processing Systems.

Bengio, Y. 2009. Learning Deep Architectures for AI. Now Foundations and Trends. doi:10.1561/2200000006.

Choi, Yeon Oh, Park, Jinhyeok, Lee, Seung Woo, Shin, Sang Heon, and Kwak, Songyun. 2023. Comparative Analysis of Algorithms for Extracting Vector Polygons of Building from Satellite Imagery based on Deep Learning. Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography 41(5):321-330.

Deng, L. and Yu, D. 2014. Deep Learning: Methods and Applications. Deep Learning: Methods and Applications. New Foundations and Trends. doi: 10.1561/2000000039.

Dupond, S. 2019. A Thorough Review on the Current Advance of Neural network structures. Annual Reviews in Control 14:200-230.

Hinton, G. 2009. Deep Belief Networks. Scholarpedia 4(5):5947.

Hochreiter, S. and Schmidhuber, J. 1997. Long Short-term Memory. Neural Computation 9(8):1735-1780.

Hyötyniemi, H. 1993. Turing Machines are Recurrent Neural Networks. Proceedings of STeP'96-Genes, Nets and Symbols. Finnish Artificial Intelligence Society:13-24.

Karen Simonyan and Andrew Zisserman 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv:1409.1556.

Kim, Dong-eun. 2024. Fine-Tuning oftheChangeFormer Deep LearningModel for DetectingBuilding Changes in Korean Aerial Imagery, Masters' Thesis, Graduate School of AI·SW Science. Sogang University.

Kim, Jong-Woo, Jung, Young-Woo, Rhim, and Hong-Chul. 2017. Study on Structure Visual Inspection Technology using Drones and Image Analysis Techniques. Journal of the Korea Institute of Building Construction 17(6):545-557.

Krizhevsky, A., Sutskever, I., and Hinton, G. 2012. ImageNet Classification with Deep Convolutional Neural Networks. Communications of the ACM 60(6):84-90.

LeCun, B., Bengio, Y., Yann, and Hinton, G. 2015. Deep Learning. Nature 521:436-444.

Liang, L., Ye, H., Yu, G., Li, G. Y. 2020. Deep Learning based Wireless Resource Allocation with Application to Vehicular Networks. Proceedings of the IEEE 108(2):341-356.

Mingxing Tan and Quoc Le. 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. PMLR 97:6105-6114.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. 2015. Human-level Control through Deep Reinforcement Learning. Nature 518:529-533.

Myung Hwan Na, Wanhyun Cho, and SangKyoon Kim. 2020. A Construction of Web Application Platform for Detection and Identification of Various Diseases in Tomato Plants Using a Deep Learning Algorithm. J Korean Soc Qual Manag. 48(4):581-596..

Pham, D. T(ed.). 1991. Artificial Intelligence in Design. Springer-Verlag.

Qinglie Yuan. 2025. Building Rooftop Extraction from High Resolution Aerial Images Using Multiscale Global Perceptron with Spatial Context Refinement. Scientific Reports 15, Article number:6499.

Sainath, T. N., Mohamed, A., Kingsbury, B., and Ramabhadran, B. 2013. Deep Convolutional Neural Networks for LVCSR. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 8614-8618, doi: 10.1109/ICASSP.2013.6639347.

Schmidhuber, J. 2015. Deep Learning in Neural Networks: An Overview. Neural Networks 61:85-117.

Seoul City. 2016. Number of Cases of imposition of Compulsory Fines for Unauthorized Construction. Seoul City.

Seungzoon Lee, Jinsup Sim, and Jeongil Choi. 2023. A Case Study on Quality Improvement of Electric Vehicle Hairpin Winding Motor Using Deep Learning AI Solution. J Korean Soc Qual Manag. 51(2):283-296.

Sherrington, D. and Kirkpatrick, S. 1975. Solvable Model of a Spin-Glass, Physical Review Letters 35(26):1792. DOI:10.1103/PhysRevLett.35.1792.

Son, Hyun. 2022. Building a GIS Map Comparison System Using Deep Learning. Masters' Thesis, Graduate School of Information Science. Soongsil University.

Tealab, Ahme. 2018. Time Series Forecasting Using Artificial Neural Networks Methodologies: A Systematic Review. Future Computing and Informatics Journal 3(2):334-340.

Tee-Ann Teo and Pei-Cheng Chen. 2025. Building Change Detection in Aerial Imagery Using End-to-End. Deep Learning Semantic Segmentation Techniques: Buildings 15(5):695.

Xiong, X. Zheng, K., Lei, L., and Hou, L. 2020. Resource Allocation based on Deep Reinforcement Learning in IoT Edge Computing. IEEE Journal on Selected Areas in Communications 38(6):1133-1146.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. 2021. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV):10012-10022.

저자소개

손현 숭실대학교 정보과학대학원 ITCEO 학과 석사, 숭실대학교 IT정책경영학과 박사과정을 수료 하였으며, 현재 ㈜아이웨이 대표이사로 재직 중이다. 주요 연구 관심 분야는 Geographic Information System, Digital Twin, 3D 지도 데이터 연구 분야 등이다.

김동호 서울대학교 전기공학과(학사), KAIST 전기 및 전자공학과(석사), George Washington University 전산학과(박사), 숭실대학교 글로벌미디어학부 교수.