

CNN 모델과 Transformer 조합을 통한 토지피복 분류 정확도 개선방안 검토*

심우담¹·이정수^{1*}

Assessing Techniques for Advancing Land Cover Classification Accuracy through CNN and Transformer Model Integration*

Woo-Dam SIM¹·Jung-Soo LEE^{1*}

요 약

본 연구는 Transformer 모듈을 기반으로 다양한 구조의 모델을 구성하고, 토지피복 분류를 수행하여 Transformer 모듈의 활용방안 검토를 목적으로 하였다. 토지피복 분류를 위한 딥러닝 모델은 CNN 구조를 가진 Unet 모델을 베이스 모델로 선정하였으며, 모델의 인코더 및 디코더 부분을 Transformer 모듈과 조합하여 총 4가지 딥러닝 모델을 구축하였다. 딥러닝 모델의 학습과정에서 일반화 성능 평가를 위해 같은 학습조건으로 10회 반복하여 학습을 진행하였다. 딥러닝 모델의 분류 정확도 평가결과, 모델의 인코더 및 디코더 구조 모두 Transformer 모듈을 활용한 D모델이 전체 정확도 평균 약 89.4%, Kappa 평균 약 73.2%로 가장 높은 정확도를 보였다. 학습 소요시간 측면에서는 CNN 기반의 모델이 가장 효율적이었으나 Transformer 기반의 모델을 활용할 경우, 분류 정확도가 Kappa 기준 평균 0.5% 개선되었다. 차후, CNN 모델과 Transformer의 결합과정에서 하이퍼파라미터 조절과 이미지 패치사이즈 조절 등 다양한 변수들을 고려하여 모델을 고도화 할 필요가 있다고 판단된다. 토지피복 분류과정에서 모든 모델이 공통적으로 발생한 문제점은 소규모 객체들의 탐지가 어려운 점이었다. 이러한 오분류 현상의 개선을 위해서는 고해상도 입력자료의 활용방안 검토와 함께 지형 정보 및 질감 정보를 포함한 다차원적 데이터 통합이 필요할 것으로 판단된다.

주요어 : 원격탐사, 딥러닝, 트랜스포머, Unet, 토지피복

2024년 3월 8일 접수 Received on March 8, 2024 / 2024년 3월 14일 수정 Revised on March 14, 2024
/ 2024년 3월 15일 심사완료 Accepted on March 15, 2024

* 본 연구는 국립산림과학원 “산림자원 평가 및 모니터링을 위한 농림위성 융합 산출물 개발(과제번호:FM 0103-2021-04-2023)”의 지원으로 수행되었습니다.

1 강원대학교 산림경영학과, 박사 / Department of Forest Management, Division of Forest Sciences, College of Forest and Environmental Sciences, Kangwon National University, Ph.D

1 강원대학교 산림경영학과, 교수 / Department of Forest Management, Division of Forest Sciences, College of Forest and Environmental Sciences, Kangwon National University, Professor

※ Corresponding Author E-mail: jslee72@kangwon.ac.kr

ABSTRACT

This research aimed to construct models with various structures based on the Transformer module and to perform land cover classification, thereby examining the applicability of the Transformer module. For the classification of land cover, the Unet model, which has a CNN structure, was selected as the base model, and a total of four deep learning models were constructed by combining both the encoder and decoder parts with the Transformer module. During the training process of the deep learning models, the training was repeated 10 times under the same conditions to evaluate the generalization performance. The evaluation of the classification accuracy of the deep learning models showed that the Model D, which utilized the Transformer module in both the encoder and decoder structures, achieved the highest overall accuracy with an average of approximately 89.4% and a Kappa coefficient average of about 73.2%. In terms of training time, models based on CNN were the most efficient. However, the use of Transformer-based models resulted in an average improvement of 0.5% in classification accuracy based on the Kappa coefficient. It is considered necessary to refine the model by considering various variables such as adjusting hyperparameters and image patch sizes during the integration process with CNN models. A common issue identified in all models during the land cover classification process was the difficulty in detecting small-scale objects. To improve this misclassification phenomenon, it is deemed necessary to explore the use of high-resolution input data and integrate multidimensional data that includes terrain and texture information.

KEYWORDS: Remote Sensing, Deep Learning, Transformer, Unet, Land Cover

서 론

원격탐사 기술을 활용한 토지피복 분류는 지구환경 모니터링, 도시계획, 농업 관리 등 다양한 분야에서 중요한 역할을 수행해왔다. 과거부터 원격탐사를 활용한 영상분류 연구는 주로 최대우도법, 최단거리법과 같은 영상의 분광정보를 활용한 전통적인 기법과 SVM(Support Vector Machine), Random Forest, Decision Tree와 같은 기계학습 기법을 중심으로 수행하였으나 이러한 방법들은 복잡한 영상 패턴과 다양한 지표면 유형을 효과적으로 구분하는 데 한계점이 발생하였다(Qiu et al., 2024). 특히, 기계학습 기반의 모델들은 고정된 특성에 기반하여 학습되기 때문에, 복잡한 지형과 다양한 조건에서 발생하는 영상의 다양성과 복잡성을 완벽히 포착하기 어려우며, 새로운 데이터에 대한

일반화 능력이 부족한 한계점을 가지고 있다(Atkinson and Tatnall, 1997; Mountrakis et al., 2011). Foody and Mathur(2004)는 SVM 기법을 활용하여 위성 영상에서 도시 지역을 분류하는 연구에서, 다양화된 도시 환경의 복잡성 때문에 분류 정확도가 제한적이었다고 보고하였다. 또한, Pal and Mather(2003)의 연구에서는 Random Forest 알고리즘을 사용하여 다양한 유형의 토지피복을 분류하였으나, 영상의 미세한 변화를 감지하는 데 있어서 한계점이 있다고 보고하였다. 이러한 한계점에 따라 원격탐사 기반의 영상분류에서 보다 고도화된 분석기법의 필요성을 제기하게 되었으며, 이에 따라 딥러닝 기반의 영상분류 기법이 등장하였다. 특히, 합성곱 신경망(CNN; Convolutional Neural Network)을 활용한 연구가 활발히 진행되어 영상분류의 정확도가 크게 향상되었다(Handayanto, 2024). Zhang et al., (2016)은

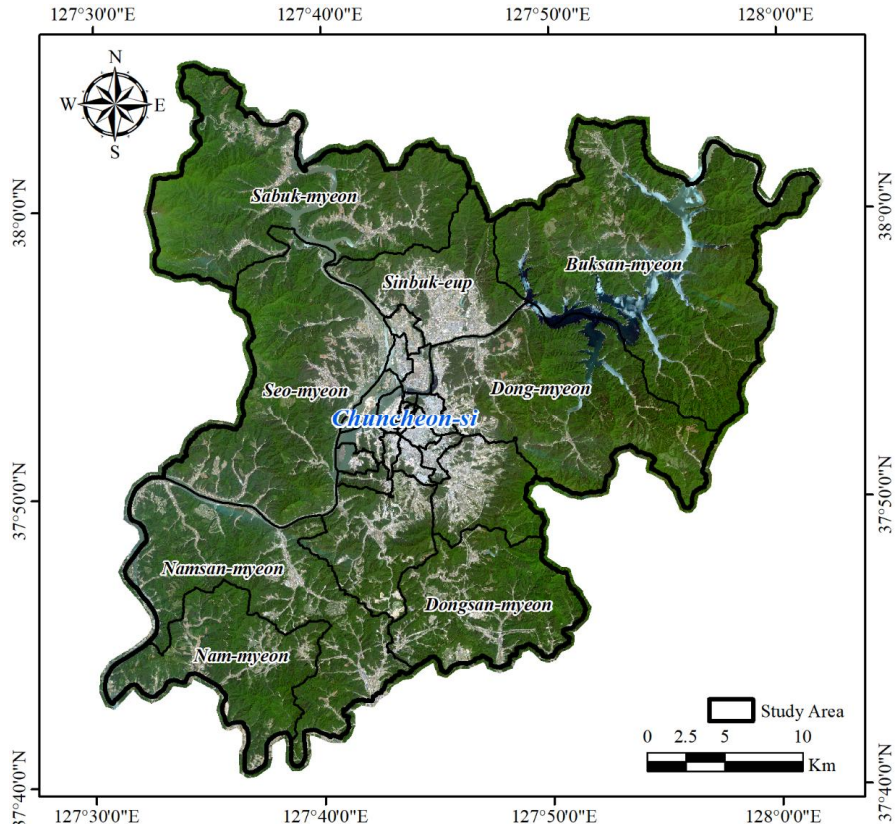


FIGURE 1. Study area

CNN 기반의 딥러닝 모델을 활용하여 고해상도 위성 영상을 기반으로 도시 지역 분류 연구를 수행하였으며, 딥러닝 기반의 접근방식이 SVM 과 같은 기계학습 방법에 비해 더 높은 분류 정확도를 달성했다고 보고하였다.

한편, CNN기반의 딥러닝 기법은 공간적 특성을 파악하는 과정에서 뛰어난 능력을 보여주었으나, CNN의 고정된 커널 크기와 국소적인 수용 영역은 대규모 데이터셋을 활용하거나 크기가 큰 영상에서 특징 정보를 추출할 때에는 일반화 성능의 한계를 보였다(Xu et al., 2024). 이에 따라, 영상 분류 작업에서 보다 복잡한 패턴과 상관 관계를 학습할 수 있는 Transformer 기반의 모델이 주목받기 시작했다(Zhao et al., 2024). Transformer 기법은 주목 메커니즘(Attention Mechanism)을 통해

영상의 다양한 부분을 효과적으로 파악하고, 전체 영상의 맥락을 이해하는 데 유리한 구조를 가지고 있다(Geng et al., 2024).

따라서, 본 연구는 위성영상을 활용한 토지피복 분류 분야에서 Transformer의 활용성을 검토하기 위해 CNN구조의 딥러닝 모델과 Transformer 모델을 결합하여 다양한 구조의 모델을 구성하였으며, 딥러닝 모델 기반 토지피복 분류를 수행하여 Transformer 모델을 통한 분류정확도 개선 여부를 검토하였다.

연구방법

1. 연구대상지

연구대상지는 강원도 춘천시이며, 지리적 위치는 경도 $127^{\circ} 30' 22.12'' \sim 128^{\circ} 01'$

48.39", 위도 37° 40' 28.28" ~ 38° 05' 06.17"에 위치하고 있다. 춘천시는 1읍 9면 15동으로 구성되어 있으며, 면적은 약 1,116km²로, 강원도에서 두 번째로 인구가 많이 분포하고 있다(춘천시, 2024). 2023년 기준 국토교통부 지적통계에 의하면, 춘천시의 토지이용현황은 임야가 약 75%로 가장 넓은 분포를 차지하고 있으며, 습지(하천·구거·유지)와 농경지(전·답·과수원)가 각각 약 9.2%, 약 8.9% 분포하고 있다(국토교통부, 2023). 특히, 춘천시는 소양댐 및 춘천댐 등 두 개의 댐과 호수 및 하천이 다수 분포하여 타도시보다 습지의 비율이 높으며, 분지형태의 지형적 특성으로 인해 연교차가 높은 지역이다(그림 1).

2. 연구방법

본 연구는 위성영상의 표면반사율(SR; Surface Reflectance) 정보를 활용하여 IPCC(Intergovernmental Panel on Climate Change) 토지이용 범주의 기준에 따라 딥러닝 모델의 학습 및 검증 데이터를 구축하였다. 또한, CNN(Convolutional Neural Network) 기반의 베이스 모델과 Transformer 모듈의 조합에 따라

4가지 딥러닝 모델을 구축하고 토지피복 분류를 수행하여 활용방안을 검토하였다(그림 2).

1) 딥러닝 모델의 데이터셋 구축

딥러닝 모델의 데이터셋은 학습자료와 검증자료, 테스트자료로 구성되어 있으며, 각각의 자료는 입력이미지와 참값이 기록된 라벨이미지로 구성된다. 입력 이미지는 2018년 5월 21일에 취득된 공간해상도 5m의 Rapideye 위성영상을 활용하였으며, 16bit로 기록된 5개의 표면반사율 정보(Red, Green, Blue, Red-Edge, NIR)를 포함하고 있다. 입력 이미지의 참값에 해당하는 라벨 이미지는 환경부 세분류 토지피복도를 IPCC의 6가지 토지이용범주(산림지, 농경지, 초지, 습지, 정주지, 기타토지)를 기준으로 재분류 하였다(그림 3). 데이터셋의 이미지 크기는 256×256 Pixel 크기로 설정하였으며, 입력 이미지 및 라벨 이미지를 해당 크기로 분할하여 타일형태의 이미지로 구축하였다. 분할한 타일 이미지는 랜덤샘플링 기법을 통해 각 연구대상지의 약 10% 면적에 해당하는 이미지를 선정하였으며, 7:3의 비율로 각각 학습자료, 검증자료로 활용하였다. 테스트 자료는 CNN기반 딥러닝 모델에서 이미지 외곽 부분의 분류정확도가 낮

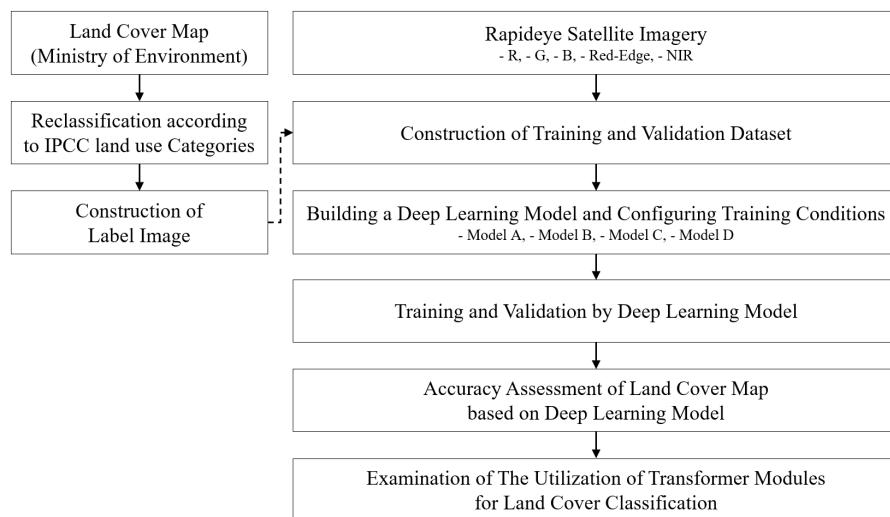


FIGURE 2. Research method for land-cover classification

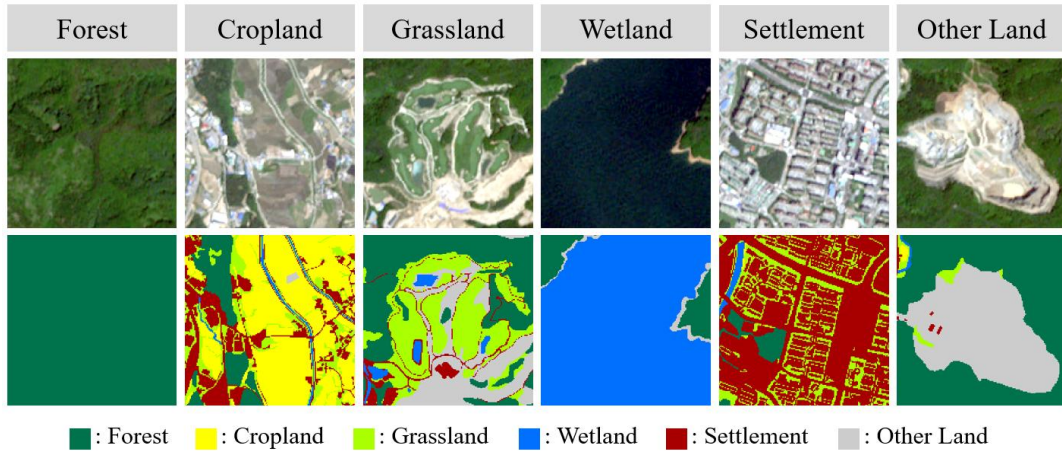


FIGURE 3. Training dataset by land cover categories

아지는 단점을 해결하기 위해 시군별 전체 지역 영상을 타일 형태의 이미지로 분할하는 과정에서 50%의 중첩률을 적용하였다. 학습자료는 총 41개의 이미지를 구축하였으며, 검증자료는 17개의 이미지, 테스트자료는 5,990개의 이미지를 구축하였다.

2) 토지피복 분류를 위한 딥러닝 모델 구축

토지피복 분류를 위해 CNN구조의 베이스모델과 Transformer 모듈의 결합을 통해 4종류의 모델을 구축하였다(그림 4). 베이스 모델로 선택한 Unet 모델은 2015년에 처음 소개된 이래로, Skip-Connection과 같은 효율적인 구조 덕분에 의료 영상 분석을 비롯한 여러 분야에서 광범위하게 적용되어 왔다(Ronneberger, et al., 2015). Unet 모델은 2개의 합성곱층으로 구성된 Unet Block과 다운샘플링을 4번 반복하는 단순한 구조로 구성되어 있으나 효율적인 구조로 인해 ResUnet, Unet++, Attention Unet, TransUnet 등 현재까지 많은 파생 모델들이 활용되고 있다(Oktay et al., 2018; Zhang et al., 2018; Zhou et al., 2018; Chen et al., 2021). 본 연구에서는 토지피복 분류의 정밀도 향상을 위해 4가지 모델을 설계하였다. Model A는 Unet과 동일한 구조를 가지며, 다른 모델과의 정확도 비교·검토를 위해 활용하였다.

Model B는 Unet 모델의 인코더 부분을 Transformer 모듈로 대체하여, 영상의 특징정보 추출과정에서 공간적 상관성을 더 잘 포착할 수 있도록 설계하였다. Model C는 Unet 모델의 인코더 부분을 Transformer 모듈로 대체하였으며, 특징 맵의 복원과 세부 정보의 정밀한 재구성에 초점을 맞췄다. Model D의 경우, Unet 모델의 인코더와 디코더 모두를 Transformer 모듈로 교체하여, 영상 내 공간적 관계성과 세부적인 특징정보 모두를 더욱 효과적으로 처리할 수 있도록 최적화하였다. 본 연구에서 활용한 Transformer 모듈은 Swin Transformer를 활용하였다. Swin Transformer는 Stage 1~4까지 구성된 계층적 구조를 통해 다양한 해상도에서 특징정보를 추출하도록 구성되어 있다. 또한, Shift Window 방식을 통해 각 계층별로 윈도우를 부분적으로 겹치게 배치함으로써, 국소적인 특징정보 뿐만 아니라 넓은 범위의 특징정보도 모델이 학습할 수 있도록 구성하였다. 이러한 유연한 구조적 특성에 따라 Swin Transformer는 이미지 분류, 객체 탐지, 객체 분할 등 다양한 분야에서 널리 활용되고 있다(Liu et al., 2021; Xie et al., 2021; Lin et al., 2022).

3) 딥러닝 모델의 학습조건 설정

딥러닝 모델의 학습 과정은 초기 단계에서 모

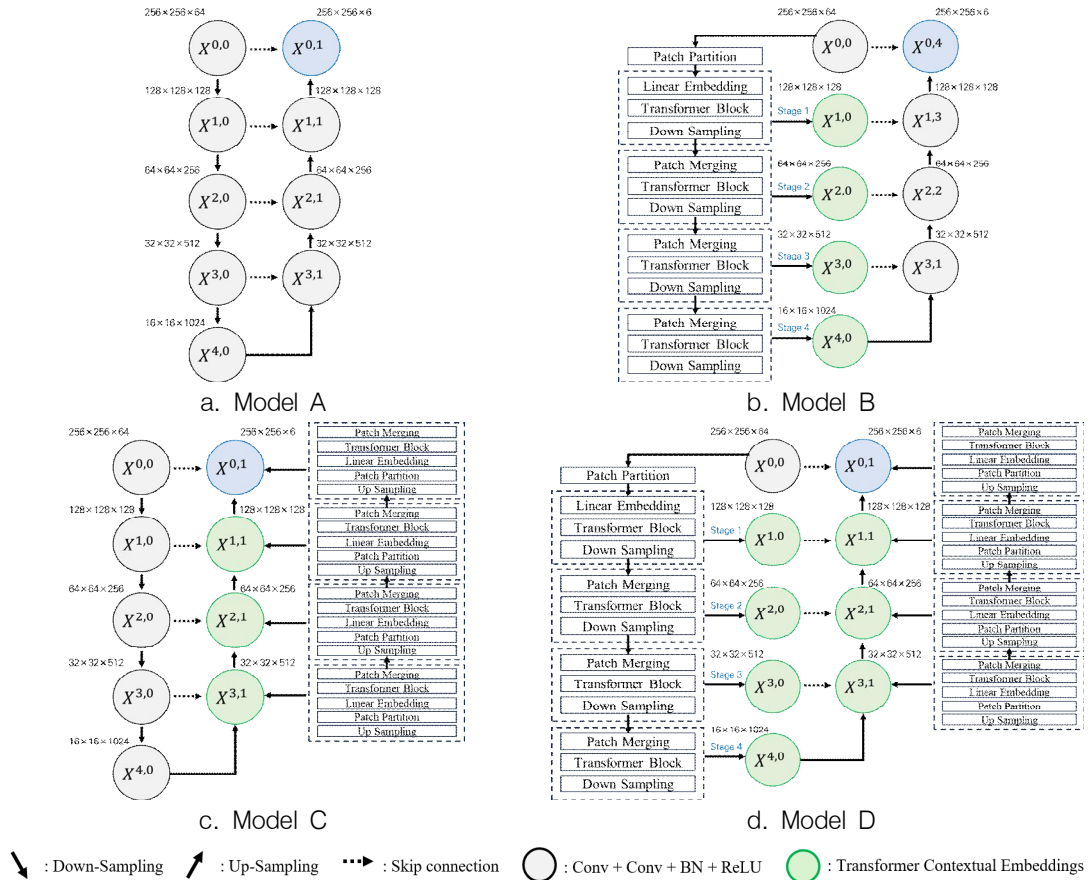


FIGURE 4. Architecture of the deep learning model for land cover classification

은 가중치를 작은 무작위 값으로 설정하면서 학습이 시작된다. 이후, 학습자료를 모델에 입력하여, 각 층(layer)을 통과하며 출력값을 계산하게 된다. 모델이 예측한 값과 실제 값 사이의 차이는 손실 함수를 통해 측정되며, 이 손실을 최소화하기 위해 각 가중치의 손실 기울기를 계산하여 가중치를 조정한다. 이 과정은 설정된 Epoch 수만큼 반복되며, 학습자료 및 검증자료와 모델 예측값과의 손실이 최소가 되도록 진행하게 된다 (Rumelhart et al., 1986; Kingma and Ba, 2014). 딥러닝 모델의 학습률은 Learning Rate Scheduler를 통해 학습률을 고정하거나 시간이 지남에 따라 점진적으로 조정(감소 또는 증가)하도록 설정할 수 있으며, Optimizer를 통해 학습

가중치를 업데이트하게 된다. 본 연구에서는 데이터세트의 클래스 불균형 문제와 과적합을 방지하기 위한 목적으로, MS-SSIM(Multi-Scale Structural Similarity), Focal Loss, Dice Loss를 포함하는 Hybrid Loss를 활용하였다. 또한, AdamW Optimizer와 OneCycleLR 기법을 활용하여 학습률을 유연하게 관리함으로써, 모델의 학습 효율성과 성능을 최적화하였다(Loshchilov and Hutter, 2018; Smith and Topin 2019). 딥러닝 모델의 학습횟수는 최대 1,000 Epoch까지 학습하도록 설정하였으며, 딥러닝 모델의 일반화 성능 평가를 위해 같은 학습 조건으로 10회 반복 학습하여 학습정확도 및 검증정확도를 비교·평가하였다.

4) 딥러닝 모델별 토지피복 분류 정합성 평가 학습이 완료된 딥러닝 모델은 테스트 데이터셋을 입력하여 토지피복도를 생성한 후, 이를 라벨 이미지와 비교함으로써 모델의 정확성을 평가하였다. 정확성 평가 과정에서는 라벨 이미지와 딥러닝 기반 토지피복도 간의 일치도를 확인하기 위해 혼동행렬(Confusion Matrix)을 작성하였으며, 혼동행렬을 기반으로, 전체 정확도(Overall Accuracy)와 Kappa 계수를 계산하여 모델의 분류 성능을 정량적으로 평가하였다(Rouhi et al., 2015; Huang and Rust, 2018). 또한, 범주별 분류정확도를 평가하기 위해 정밀도(Precision)와 재현율(Recall)을 산출하고, F1-Score를 활용하여 정확도를 평가하였다.

$$\text{Overall Accuracy (OA)} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

$$kappa = \frac{OA - p_e}{1 - p_e} \quad (2)$$

$$p_e = \left(\frac{TP + FN}{TP + FN + FP + TN} \times \frac{TP + FP}{TP + FN + FP + TN} \right) + \left(\frac{FP + TN}{TP + FN + FP + TN} \times \frac{FN + TN}{TP + FN + FP + TN} \right) \quad (3)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$F_1 - \text{Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

연구 결과

1. 딥러닝 학습자료의 범주별 분포 특성

학습자료의 범주별 면적 분포는 산림지가 64.5%로 가장 많은 면적을 차지하고 있으며,

습지, 초지 순으로 분포하고 있다. 특히, 기타토지의 경우, 전체 학습자료 면적의 2% 미만으로 분포하기 때문에 이는 학습 데이터 내에서 범주 간에 상당한 불균형이 있음을 나타낸다(표 1).

학습 자료의 범주별 표면 반사율(Surface Reflectance, SR)분포는 Red의 반사율은 기타토지가 가장 높았으며, 정주지, 농경지 순으로 높은 분포를 보였다. NIR의 경우, 산림지, 초지, 농경지 순으로 높은 분포를 보였다. 산림지와 습지는 다른 범주들과 확연히 구분되는 뚜렷한 분포 특성을 보였으나 농경지, 정주지, 기타 토지는 반사율이 서로 중첩되는 경향이 높아, 유사한 분포 특성을 가지는 것으로 분석되었다. 특히, 농경지 내에 포함된 시설재배지(예: 비닐하우스)는 정주지와 중첩률을 높이는 주요 요인으로 작용하였다. 초지는 농경지, 정주지, 산림지와 중첩률이 높으며 이는 초지가 다양한 환경 조건과 인접해 있다는 것을 의미한다.

2. 딥러닝 모델 별 학습 및 검증

딥러닝 모델의 학습 및 검증 정확도는 Epoch 1,000 기준, 같은 조건으로 10회 반복 학습한 결과를 측정하였다. 모든 모델에서 학습 정확도 약 93%, 검증정확도 약 88% 이상의 높은 정확도를 보였다(표 2). 딥러닝 모델 간 큰 차이는 발생하지 않았으나, 인코더에 Transformer 모델을 활용한 Model B는 타 모델 보다 학습 및 검증 정확도가 소폭 감소하였다. 한편, 학습 정확도는 전체 모델 평균 약 94.2%를 달성하였으나, 검증정확도는 평균 약 88.3%로 과적합 현상이 발생하였다. 과적합 현상은 모델이 학습데

TABLE 1. Spectral characteristics of training data for land cover classification

Land Cover Categories	Area Distribution	Surface Reflectance	
		Red	NIR
Forest	64.5%	249±109	866±235
Cropland	7.8%	794±456	1,434±430
Grassland	8.7%	563±315	1,172±328
Wetland	9.4%	220±207	379±343
Settlement	5.9%	1,051±557	1,422±511
Other Land	1.5%	1,166±622	1,524±604

TABLE 2. training and validation results of the deep learning model

Deep Learning Model	Training Results		Validation Results		Training Time (Minutes)
	Accuracy	Loss	Accuracy	Loss	
A	94.0±0.81	0.33±0.02	88.6±0.57	0.41±0.03	34.4±0.7
B	93.2±1.02	0.36±0.03	87.5±1.74	0.41±0.03	46.3±0.2
C	94.1±0.59	0.33±0.01	88.3±0.01	0.38±0.02	75.5±0.3
D	95.4±0.54	0.34±0.02	88.6±0.21	0.43±0.02	157.7±0.5

* The average values of models trained with 10 repetitions at 1,000 epochs.

이터에 지나치게 최적화되어 새로운 데이터에 대한 일반화 능력이 저하되는 현상이다. 본 연구에서는 과적합 현상을 방지하기 위해 Hybrid Loss를 활용하였으나, 딥러닝 모델의 학습자료가 특정 범주에 치중되어 있었기 때문에 학습과정에서 불균형한 데이터의 특징을 과하게 학습함으로써 검증정확도가 낮은 것으로 판단된다 (Das, 2024; Talukder et al., 2024). 이에 따라, 향후 데이터세트의 다양성과 균형을 고려하여 모델의 일반화 성능 향상이 필요할 것으로 판단된다.

딥러닝 모델 별 학습시간을 비교해 보면, Transformer 모듈을 활용한 Model B, C, D는 CNN 기반의 Model A보다 학습시간이 더 증가하였다. 특히, 인코더 및 디코더 모두 Transformer를 활용한 Model D는 Model A보다 약 4.5배 학습시간이 더 소요되었다. Dosovitskiy et al. (2020)은 Transformer 모델의 학습과정에서 일반적인 CNN 모델에 비해 더 많은 학습 시간과 계산비용이 소요된다고 언급하였으며, 이는 복잡한 패턴 인식과 처리 능력을 향상시키기 위해 모델의 구조적 복잡성이

증가하였기 때문으로 판단된다.

3. 딥러닝 모델 별 토지피복 분류 정확도 평가

딥러닝 모델기반 토지피복도의 면적 분포를 라벨 이미지와 비교해보면, CNN기반의 Model A는 산림지 및 기타토지가 과대추정되었으나, Transformer 기반의 모델들은 라벨 이미지에 비해 산림지, 초지, 기타토지가 과대추정되는 결과를 보였다(표 3). 또한, 딥러닝 모델의 토지피복 분류지도에서는 어떠한 범주로도 분류되지 않은 미분류 지역이 분포하였다. 딥러닝 모델에서 미분류가 발생한 원인은 모델의 학습과정에서 보지 못한 극단적인 특성을 가진 데이터를 입력 받았을 때 또는 범주에 속할 확률이 결정 임계값에 도달하지 못할 경우에 발생하게 된다. 본 연구에서 발생한 미분류 사례는 영상 내에서 빛 반사가 강하게 일어나 흰색 계열로 분포하는 지역에서 발생하였으며, 차후 다양한 분포패턴을 가진 데이터세트의 학습이 필요할 것으로 판단된다.

딥러닝 모델 별 토지피복 분류지도의 범주별 분포 특성을 비교·평가 하기 위해 Raster 형

TABLE 3. Estimated area by category in land cover maps for each deep learning model

Category	Label Image	Unit : ha(Model-Label)			
		Model A	Model B	Model C	Model D
Forest	84,452	86,767(+2,315)	86,884(+2,432)	87,068(+2,616)	86,534(+2,082)
Cropland	6,832	6,425(-406)	4,032(-2,799)	4,446(-2,385)	4,783(-2,048)
Grassland	6,656	6,269(-387)	7,720(+1,063)	7,870(+1,214)	7,781(+1,124)
Wetland	7,520	7,115(-405)	7,069(-452)	6,515(-1,005)	6,784(-736)
Settlement	5,262	3,852(-1,411)	4,279(-983)	4,112(-1,150)	4,816(-446)
Other land	847	1,066(+220)	1,582(+736)	1,479(+663)	789(-58)
Unclassified Area	-	73	3	77	82
Total	111,569	111,569	111,569	111,569	111,569

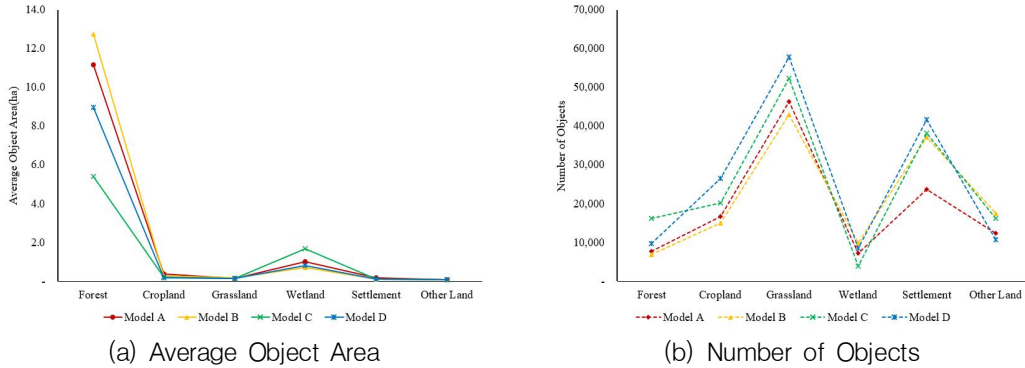


FIGURE 5. Object characteristics by land cover category

태의 토지피복 분류지도를 Vector 형태로 변환하여 객체별 평균 면적과 개수를 분석하였다. 딥러닝 모델 기반 토지피복도의 범주별 객체 분포를 비교해 보면, Transformer 기반의 모델들은 Model A에 비해 대부분 더 작은 크기의 객체로 분포하였다(그림 5). 이러한 현상은 Transformer가 특징정보 추출을 위해 입력 영상을 패치단위로 분할하고 상관성을 분석하기 때문으로 판단된다.

혼동행렬을 통한 딥러닝 모델별 토지피복도와 라벨자료의 정확도 비교 결과, OA(Overall Accuracy)는 전체 모델 평균 약 89.0%, Kappa는 평균 약 71.3%로 분포하였다(표 4). Model C의 경우, 타 모델보다 OA 및 Kappa가 다소 낮은 분포를 보였으나, Model A, B, D는 큰 차이가 발생하지 않았다. 토지피복 분류 정확도가 가장 높은 모델은 Kappa 기준 Model D가 평균적으로 가장 높은 정확도를 보였으며, 10회 반복 학습 과정에서 모델간의 편차가 가장 적어 높은 신뢰성을 보였다. Wang et al.

(2024)는 Transformer가 대규모 학습 데이터 세트에서 더 높은 정확도를 보인다고 언급하였으나, 이와 달리 본 연구에서는 41장의 소규모 학습자료를 활용하였음에도 Model D의 분류정확도가 가장 높게 분포하였다. Wang et al. (2022)와 Choi et al. (2023)의 연구에서도 CNN 구조의 딥러닝 모델에 비해 Transformer를 활용하였을 때 분류 정확도가 개선되었다. F1-Score를 활용한 범주별 정확도를 비교해 보면, 산림지와 습지는 모든 모델에서 90% 이상의 높은 정확도를 보였으나, 초지와 기타토지는 이전 연구와 같이 타 범주에 비해 상대적으로 낮은 정확도를 보였다(Stoian et al., 2019; Lee and Lee, 2020).

딥러닝 모델 별 토지피복 범주의 분류결과를 보면, 4종류의 모델 모두 임도와 같이 작은 객체는 탐지하지 못하는 모습을 보였으며, 벌채지와 같이 산림경영활동이 이루어진 지역은 초지로 오분류 하는 사례가 다수 발생하였다(그림 6-1). Kammerer et al. (2024)는 본 연구와

TABLE 4. Evaluation of land cover map accuracy by deep learning model

Deep Learning Model	Overall Accuracy	Kappa	F1-Score					
			Forest	Cropland	Grassland	Wetland	Settlement	Other land
A	89.4±0.5%	72.7±1.6%	96.9±0.4%	60.9±2.5%	41.3±2.8%	90.9±0.2%	60.7±0.9%	36.8±4.6%
B	89.3±0.2%	72.6±0.9%	97.1±0.3%	54.2±3.1%	44.6±1.8%	90.5±0.5%	61.7±0.4%	33.0±2.6%
C	87.9±1.2%	66.6±8.1%	96.0±1.7%	52.9±5.1%	39.8±9.6%	90.6±0.2%	46.7±17.7%	31.6±1.0%
D	89.4±0.2%	73.2±0.8%	97.1±0.2%	59.9±0.2%	45.9±0.9%	90.9±0.1%	60.6±2.2%	31.2±5.5%

* The average values of models trained with 10 repetitions at 1,000 epochs.

같이, 토지피복 분류과정에서 목초지/건초지와 관목지, 산림지 범주간의 오분류 사례가 다수 발생하였다고 언급하였다. 농경지 지역에서는 비닐하우스와 같은 시설재배지역을 정주지 또는 기타토지로 오분류 하는 사례가 많이 발생하였으나, 인코더 및 디코더 모두 Transformer를 활용한 Model D에서는 시설재배지를 기타토지로 오분류 하는 사례가 개선되었다(그림 6-2). 골프장과 같은 초지 지역에서도 앞선 사례와 마찬가지로 작은 도로나 모래와 흙으로 구성된 소규모 병커는 탐지하지 못하는 사례가 많았다. 또한, 잔디로 구성된 필드 지역을 CNN 모델의 경우 농경지로 오분류 하는 사례가 다수 발생하였으나, Transformer 기반의 모델에서는 이러한 오분류 사례가 개선되었다(그림 6-3). 습지의 경우, 모든 모델에서 높은 정확도를 보였으나, 하천에서 발생한 녹조현상이나 이끼류 자생 지역을 초지로 오분류 하는 사례가 가장 많이 발생하였다. 이러한 현상은 학습자료 내에서 초지 범주와 가장 가까운 객체들을 분석하였을 때 습지와 인접한 객체가 전체 초지의 약 57%를 차지하기 때문에 주변과의 관계까지 학습하는 딥러닝 모델의 특성상 오분류가 가장 많이 발생한 것으로 판단된다. 정주지의 경우 건축물과 같은 인공구조물은 대부분의 모델에서 정확도가 높았으나, 건설중인 토지, 토석채취지, 운동장과 같이 나지의 형태로 분포하는 지역을 기타토지가 아닌 정주지로 오분류 하는 사례가 가장 많이 발생하였다(그림 6-5). 이러한 현상은 범주 간 유사한 분광값의 분포 때문으로 판단되며, 기타토지 및 정주지의 분류정확도 개선을 위해서는 분광정보 이외에도 GLCM(Gray Level Co-occurrence Matrix)과 표고, 경사와 같은 지형정보 등 다양한 공간정보의 활용이 필요할 것으로 판단된다.

결론 및 고찰

본 연구는 Transformer의 활용가능성을 검토하기 위하여 CNN구조의 베이스모델과의 결합을 통해 4종류의 모델을 구성하고, 토지피복

분류를 수행하여 Transformer 모듈의 활용방안을 검토하였다. 학습 소요시간 측면에서는 CNN 기반의 모델이 가장 효율적이었으나 Transformer 기반의 모델을 활용할 경우, 분류 정확도가 Kappa 기준 평균 0.5% 개선되었다. 분류 정확도가 가장 높은 모델은 베이스 모델인 Unet 모델에서 인코더 및 디코더 부분을 모두 Transformer로 대체한 Model D의 정확도가 가장 높았으나, 분류 정확도 향상이 크지 않았기 때문에 CNN 모델과의 결합과정에서 하이퍼 파라미터 조절과 이미지 패치사이즈 조절 등 다양한 변수들을 고려하여 모델을 고도화 할 필요가 있다고 판단된다. 또한, 모델의 학습과정에서 과적합 현상을 방지하기 위해 Hybrid Loss를 활용하였으나, 학습정확도에 비해 검증정확도가 낮은 수치를 보였다. 이러한 현상은 딥러닝 모델의 학습자료가 특정 범주에 치중되어 있었기 때문에 학습과정에서 불균형한 데이터의 특징을 과하게 학습했기 때문으로 판단되며, 차후 모델의 일반화를 위해 데이터 증강기법의 적용, 정규화 기법 연구 등 데이터셋 구축을 위한 다양한 연구방안이 필요할 것으로 판단된다. 토지 피복 분류과정에서 모든 모델이 공통적으로 발생한 문제점은 작은 도로, 병커와 같은 소규모 객체들의 탐지가 어려운 점이었다. 이러한 오분류 현상의 개선을 위해서는 지형 정보 및 질감 정보를 포함한 다차원적 데이터 통합이 필요할 것으로 판단된다. **KAGIS**

REFERENCES

- Atkinson, P.M. and A.R.L. Tatnall. 1997. Introduction neural networks in remote sensing. *International Journal of remote sensing* 18(4):699-709.
- Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Lu, L., Yuille, A.L. and Y. Zhou. 2021. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv: 2102.04306*.

- Choi, S.H., Park, K.B. and J.Y. Lee. 2023. SwinResNet: Volumetric Medical Image Segmentation by Fusing Swin Transformer and ResNet. *Korean Journal of Computational Design and Engineering* 28(3): 282–293 (최성호, 박경범, 이재열. 2023. SwinResNet: Swin Transformer와 ResNet 융합을 통한 Volumetric 의료 영상 분할. *한국 CDE 학회 논문집* 28(3):282–293).
- Choung, Y.J. 2014. Analysis of Land Uses in the Nakdong River Floodplain Using RapidEye Imagery and LiDAR DEM. *Journal of the Korean Association of Geographic Information Studies* 17(4):189–199 (정운재. 2021. RapidEye 영상과 LiDAR DEM을 이용한 낙동강 범람 원 내 토지 이용 현황 분석. *한국지리정보학회지* 17(4):189–199).
- Das, S. 2024. A new technique for classification method with imbalanced training data. *International Journal of Information Technology*:1–9.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehahani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and N. Houlsby. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Foody, G.M. and A. Mathur. 2004. A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on geoscience and remote sensing* 42(6):1335–1343.
- Geng, S., Zhai, S. and C. Li. 2024. Swin transformer based transfer learning model for predicting porous media permeability from 2D images. *Computers and Geotechnics*. 168(106177).
- Handayanto, R.T. 2024. Land Cover Segmentation of Multispectral Images Using U-Net and DeeplabV3+ Architecture. *Jurnal Ilmu Komputer dan Informasi* 17(1):89–96.
- Huang, M.H. and R.T. Rust. 2018. Artificial intelligence in service. *Journal of service research* 21(2):155–172.
- Kammerer, M., Iverson, A.L., Li, K. and S.C. Goslee. 2024. Not just crop or forest: an integrated land cover map for agricultural and natural areas. *Scientific Data* 11(1):137.
- Kingma, D.P. and J. Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kwon, S.K., Kim, E.H., Lim, J.B. and A.R. Yang. 2021. The Analysis of Changes in Forest Status and Deforestation of North Korea's DMZ Using RapidEye Satellite Imagery and Google Earth. *Journal of the Korean Association of Geographic Information Studies* 24(4):113–126 (권수경, 김은희, 임중빈, 양아람. 2021. RapidEye 위성영상과 구글 어스를 활용한 북한 DMZ의 산림현황 및 산림황폐지 변화 분석. *한국지리정보학회지* 24(4):113–126).
- Lee, S.H. and M.J. Lee. 2020. A study on deep learning optimization by land cover classification item using satellite imagery. *Korean Journal of Remote Sensing* 36(6_2):1591–1604 (이성혁, 이명진. 2020. 위성영상을 활용한 토지피복 분류 항목별 딥러닝 최적화 연구. *대한원격탐사학회지* 36(6_2):1591–1604).
- Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G.

- and D. Zhang. 2022. Ds-transunet: Dual swin transformer u-net for medical image segmentation. *IEEE Transactions on Instrumentation and Measurement* 71: 1–15.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. and B. Guo. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*. pp.10012–10022.
- Loshchilov, I. and F. Hutter. 2018. Fixing weight decay regularization in adam. <https://openreview.net/forum?id=rk6qdGgCZ>. (Accessed Feb 15, 2024)
- Mountrakis, G., Im, J. and C. Ogole. 2011. Support vector machines in remote sensing: A review. *ISPRS journal of photogrammetry and remote sensing* 66(3):247–259.
- Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K. and D. Rueckert. 2018. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.
- Pal, M. and P.M. Mather. 2003. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment* 86(4):554–565.
- Qiu, C., Zhang, X., Tong, X., Guan, N., Yi, X., Yang, K., Zhu, J. and A. Yu. 2024. Few-shot remote sensing image scene classification: Recent advances, new baselines, and future trends. *ISPRS Journal of Photogrammetry and Remote Sensing* 209:368–382.
- Ronneberger, O., Fischer, P. and T. Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention MICCAI International Conference*. pp.234–241.
- Rouhi, R., Jafari, M., Kasaei, S. and P. Keshavarzian. 2015. Benign and malignant breast tumors classification based on region growing and CNN segmentation. *Expert Systems with Applications* 42(3):990–1002.
- Rumelhart, D.E., Hinton, G.E. and R.J. Williams. 1986. Learning representations by back-propagating errors. *nature* 323(6088):533–536.
- Smith, L.N. and N. Topin. 2019. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications Vol. 11006*, pp.369–386.
- Stoian, A., Poulain, V., Inglada, J., Poughon, V. and D. Derksen. 2019. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. *Remote Sensing* 11(17):1986.
- Talukder, M.A., Sharmin, S., Uddin, M.A., Islam, M.M. and S. Aryal. 2024. MLSTL-WSN: Machine Learning-based Intrusion Detection using SMOTETomek in WSNs. *arXiv preprint arXiv:2402.13277*.
- Wang, H., Xing, C., Yin, J. and J. Yang. 2022. Land cover classification for polarimetric SAR images based on vision transformer. *Remote Sensing*

- 14(18):4656.
- Wang, H., Lin, J., Li, Y., Dong, X., Tong, X. and S. Lu. 2024. Self-Supervised Pre-Training Transformer for Seismic Data Denoising. *IEEE Transactions on Geoscience and Remote Sensing* 62: 5907525.
- Won, M.S., Kim, Y.S. and K.H. Kim. 2014. Estimation on Greenhouse Gases(GHG)s Emission of Large Forest Fire Area in 2013. *Journal of the Korean Association of Geographic Information Studies* 17(3):54-67 (원명수, 김유승, 김경하. 2014. RapidEye 영상을 활용한 대형산불피해지의 온실가스 배출량 추정. *한국지리정보학회지* 17(3):54-67).
- Xie, Z., Lin, Y., Yao, Z., Zhang, Z., Dai, Q., Cao, Y. and H. Hu. 2021. Self-supervised learning with swin transformers. *arXiv preprint arXiv:2105.04553*.
- Xu, R., Dong, X.M, Li, W., Peng, J., Sun, W. and Y. Xu. 2024. DBCTNet: Double Branch Convolution-Transformer Network for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* Vol. 62.
- Zhang, L., Zhang, L. and B. Du. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and remote sensing magazine* 4(2): 22-40.
- Zhang, P., Ke, Y., Zhang, Z., Wang, M., Li, P. and S. Zhang. 2018. Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery. *Sensors* 18(11):3717.
- Zhao, Y., Bao, W., Xu, X. and Y. Zhou. 2024. Hyperspectral image classification based on local feature decoupling and hybrid attention SpectralFormer network. *International Journal of Remote Sensing* 45(5):1727-1754.
- Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N. and J. Liang. 2018. Unet++: A nested u-net architecture for medical image segmentation. *Proceedings of the In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop*. pp.3-11.
- 국토교통부. 2023. 지적통계. <https://stat.molit.go.kr/portal/cate/statMetaView.do?hRsId=24>. (Accessed February 2, 2024)
- 춘천시. 2024. 춘천 소개. <https://www.chuncheon.go.kr/cityhall/about-chuncheon/introduction/general/>. (Accessed February 2, 2024). 