

트랜스포머 기반 산림지와 도심지 의미론적 분할 및 탄소 저장량 예측 기법

우 승 우¹ · 장 은 서² · 곽 찬 희^{3*}

¹강남대학교 인공지능융합공학부 인공지능전공 학사과정

²강남대학교 소프트웨어응용학과 학사과정

³강남대학교 인공지능융합공학부 데이터사이언스전공 조교수

Transformer-based Semantic Segmentation and Carbon Storage Prediction Model for Application in Forest and Urban Areas

Seung-Woo Woo¹ · Eun-Seo Jang² · Chanhee Kwak^{3*}

¹Bachelor's Course, Department of Artificial Intelligence, Kangnam University, Yongin 16979, Korea

²Bachelor's Course, Department of Software Applications, Kangnam University, Yongin 16979, Korea

³Assistant Professor, Department of Data Science, Kangnam University, Yongin 16979, Korea

[요 약]

산림의 탄소 저장량 측정은 주로 현장 조사와 원격 탐사의 방법을 통해 이뤄지나, 이런 접근법은 측정의 정확도가 떨어지는 문제를 안고 있다. 따라서 본 연구에서는 항공 관측 이미지를 활용하여 산림을 구분하고 이에 상응하는 탄소 저장량을 측정하는 딥러닝 모델을 제안한다. 기존의 이미지 기반 객체 탐지 모델들은 매개 변수의 수가 많고 계산 비용이 높아 실시간 서비스나 자원 제약이 많은 경우 활용이 어렵다는 단점이 있다. 이에 본 연구가 제안하는 모델은 트랜스포머 구조를 기반한 SegFormer를 활용해 경량화와 분류 정확도의 두 가지 목표를 동시에 달성하고자 하였다. 실험 결과 1픽셀당 25cm 해상도의 항공 이미지를 활용했을 때 기존 알고리즘 대비 우수한 성능을 확인하였다. 본 연구의 결과를 활용하여 추후 탄소 저장량 예측 및 모니터링에 효과적으로 적용될 수 있을 것으로 기대한다.

[Abstract]

Carbon storage in forests is mainly measured by field surveys and remote sensing methods; however, these approaches afford reduced measurement accuracy. Therefore, this study proposes a deep learning model that utilizes aerial imagery to classify forests and measure the corresponding carbon storage. Existing image-based object detection models have a large number of parameters and high computational costs; therefore, it is difficult to apply them in real-time services or in resource-constrained situations. To address this, the proposed model leverages the SegFormer architecture based on transformers to simultaneously achieve both lightweight design and classification accuracy. The proposed model experimentally demonstrated superior performance compared to existing algorithms on aerial images with a resolution of 25 cm per pixel. The findings of this study are expected to be effectively applied in future carbon storage prediction and monitoring efforts.

색인어 : 컴퓨터 비전, 딥러닝, 항공 영상, 탄소 저장량, 의미론적 분할

Keyword : Computer Vision, Deep Learning, Aerial Image, Carbon Storage, Semantic Segmentation

<http://dx.doi.org/10.9728/dcs.2024.25.11.3495>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 30 August 2024; **Revised** 02 October 2024

Accepted 18 October 2024

***Corresponding Author; Chanhee Kwak**

Tel: +82-31-280-3718

E-mail: chk@kangnam.ac.kr

I. 서론

기후 변화는 인류가 직면한 시급한 문제로, 심각한 환경적, 사회적, 경제적 영향을 일으키고 있다. 세계기상기구(World Meteorological Organization)는 온실가스 농도, 해수 온도, 해양 산성화, 해수면 상승을 기후 변화의 핵심 지표로 보고하고 있는데, 이러한 지표들은 최근 몇 년간 정상 범위를 벗어난 채 유지되고 있다[1]. 현 상황이 지속될 경우 지구온난화의 가속화로 인한 심각한 환경적, 사회적, 경제적 영향을 피할 수 없기에, 지구온난화의 진행을 저지하고 기후 변화를 완화하기 위한 범세계적인 노력과 정책들이 지속적으로 개발 및 추진되고 있다.

대기 중 온실가스 감축을 위해 국가들은 다양한 접근을 시도하고 있는데, 그 중 산림의 역할은 매우 중요하다. 기후변화에 관한 정부간 협의체(IPCC)에서 발간한 「국가 온실가스 인벤토리 작성을 위한 2006 IPCC 가이드라인[2]에서는 토지 이용 범주를 6개의 분류로 구분하여 각 특정 토지이용에 대한 온실가스 배출량 및 흡수량을 산정하고 보고할 것을 요구하고 있다. 우리나라에서도 2020년에 발표한 ‘2050 탄소 중립 추진전략’을 통해 탄소 흡수원과 블루카본의 중요성을 인식하고 강조하였으며, 다양한 방법으로 산림의 탄소 저장량을 높이고 관리할 것을 명시했다.

탄소 중립 달성을 위해 각국은 블루카본 확대, 산림자원 조성 확대 등의 다양한 방법을 사용하고 있다. 특히 산림 자원을 활용하는 방법은 현재까지 개발된 어떤 인위적 이산화탄소 제거 기술보다 환경적이며, 비용 대비 이산화탄소 제거 효과가 높은 방법이기 때문에, 이를 정확하게 측정하고 관리하는 것이 중요하다. 탄소 저장량 예측 연구는 기술 발전을 통해 점점 더 정교해지고 있지만, 여전히 정확도가 떨어지는 문제와 높은 비용이 발생하는 한계점을 안고 있다. 최신 기술들은 다양한 방법을 통해 탄소 저장량을 추정하려고 시도하고 있으나, 현장 조사나 원격 탐사 기법에서 발생하는 오류는 예측의 신뢰성을 저하시킬 수 있다. 또한, 고해상도 데이터와 기계학습 기법을 활용하는 과정에서 소요되는 비용이 많아 실용적인 적용에 제약이 있다. 이러한 한계점들은 탄소 중립 목표 달성을 위한 연구의 효율성을 저해하며, 보다 효과적인 예측 모델 개발이 필요하다.

본 연구는 서울 수도권 지역을 대상으로 AIHub에서 제공하는 식생 탄소 저장량 데이터셋[3]에 SegFormer[4]를 접목해, 산림 및 도시 지역의 탄소 저장량을 예측하고 분석하는 새로운 모델을 제안한다. 본 연구에서 활용한 데이터셋은 고해상도의 항공 영상 및 위성 영상을 포함하며, RGB, 식생의 높이를 나타내는 임분고(SH) 정보와 Ground Truth, Carbon Map으로 구성되었는데, 이 데이터를 바탕으로 식생 유형을 정확하게 분류하고 각 지역의 탄소 저장량을 예측할 수 있는 모델을 개발했다. 본 연구에서는 SegFormer[4]를 바탕으로 커널 크기와 레이어 수를 조정하고 탄소 저장량 예측을 위한 헤드를 추가해 우수한 Segmentation 성능을 바탕으로 탄소

저장량 예측의 정확성을 높여 기존 머신러닝 기반 모델과 비교했을 때 높은 정확도와 효율성을 기대할 수 있다.

II. 관련 연구

2-1 Transformer Based Semantic Segmentation Model

비전 트랜스포머(Vision Transformer)는[5] 자연어 처리 분야에서 사용된 트랜스포머(Transformer)[6] 구조의 인코더 부분을 활용한 모델이다. 트랜스포머 구조 특성상 전역적 특징 추출능력이 뛰어나고, MHSA(Multi-Head Self Attention)를 통해 학습 속도와 표현력을 높이며, 입력 크기에 대한 영향 없이 변형이 용이하다. 또한 뛰어난 사전 학습 모델의 특징 추출 능력을 바탕으로 다양한 디코더 변형을 통해 분류(Classification), 분할(Segmentation) 및 검출(Detection) 등 다양한 분야에서 활발하게 적용되고 있다.

SegFormer[4]는 Transformer 기반의 Segmentation 모델로, CNN(Convolutional Neural Networks)와 Transformer의 장점을 결합해 기존 Transformer 구조보다 연산 효율성과 높은 성능을 보인다. 해당 모델은 Hierarchical Transformer Encoder를 사용하여 다중 해상도의 피쳐 맵을 생성하고, Multi-level Feature Aggregation Decoder를 통해 이를 결합해 최종 출력을 구성한다. 기존의 Transformer 모델에 비해 연산 복잡도와 메모리 사용량을 크게 줄이면서 높은 성능을 보여주며, 다양한 응용 분야에서 유용하게 활용될 수 있다. 해당 모델은 대규모 데이터셋에서 우수한 결과를 보였다.

2-2 Aerial Image Processing

항공 영상 처리는 지형 정보, 도시 구조, 자연 환경, 등을 식별하고 분류하는 중요한 기술로, 도시 계획, 환경 모니터링, 재난 대응 등 다양한 분야에서 핵심적인 역할을 한다. 비의넌트 정도를 항공 사진 분석으로 탐지해내거나[7], 이미지와 프롬프트 임베딩을 바탕으로 의미론적 분할을 수행하는 모델[8] 또한 우수한 성능을 보였다. 기존의 CNN 기반 모델들은 지역적인 특징 추출에는 강점이 있으나, 전역적인 문맥 이해에는 한계가 있었다. 이를 극복하기 위해 U-Net 구조의 개선 모델과 트랜스포머 기반 모델들이 도입되어, 전역적인 상호작용을 통한 더욱 정확한 분할이 가능해졌으며, 이는 항공 영상 의미론적 분할에도 유의미한 영향을 미쳤다.

Khan and Jung[9]은 U-Net에 Self-Attention과 Separable Convolutions를 결합하여 항공 영상의 의미론적 분할 성능을 향상시켰다. 이 모델은 Self-Attention 메커니즘을 통해 전역적인 문맥 정보를 포착하고, Separable Convolutions로 연산 효율성을 높여 고해상도 영상에서의 분할 정확도를 개선하였다. 또한, Anilkumar et al.[10]은

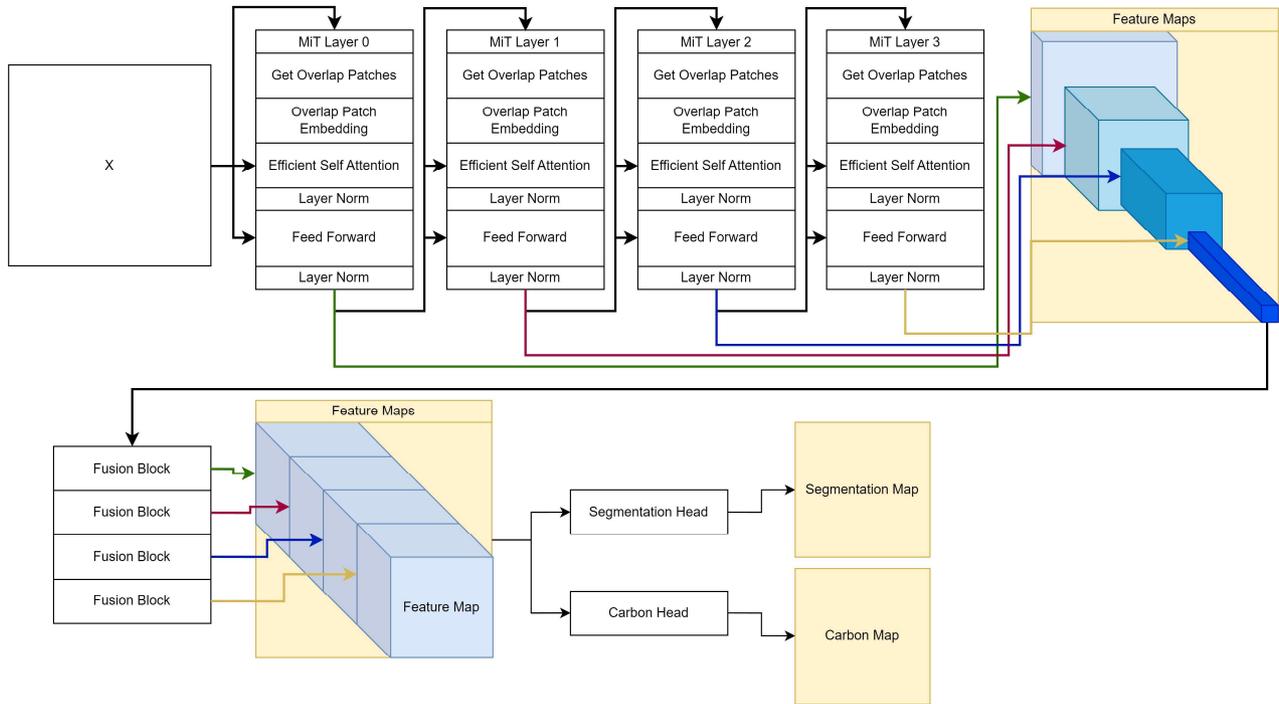


그림 1. 모델 구조
Fig. 1. Model architecture

Improved Golden Eagle Optimization Algorithm을 활용하여 Adaptive DeepLabv3+ 모델을 제안하였다. 이 방법은 최적화 알고리즘을 통해 모델의 파라미터를 효율적으로 조정하여 항공 영상의 의미론적 분할에서 높은 성능을 보였다.

한편, AerialFormer는[11] CNN과 Transformer의 장점을 결합한 모델로, CNN의 국소적 상관성 포착과 Transformer의 전역 상관성 포착 능력을 모두 활용하여 다양한 스케일의 항공 영상 분할에 효과적이다. 이 모델은 다중 해상도 CNN Stem, Swin Transformer Encoder, 다중 확장 CNN Decoder로 구성되어 있으며, iSAID, LoveDA, Potsdam과 같은 데이터셋에서 뛰어난 성능을 입증하였다. 이런 모델들의 한계점으로는 파라미터가 매우 많고, 계산비용이 매우 크기 때문에 실시간 응용이나 자원 제약이 있는 환경에서의 적용이 어려울 수 있다. 따라서 효율성을 증대를 위해 SegFormer를 기반으로 한 모델을 개발하였다.

2-3 Carbon Storage Prediction

현재 국내에서는 산림 자원이 흡수하는 탄소 저장량을 정밀하게 예측하여 탄소 중립을 실행하고 있으며, 기술 발전에 따라 다양한 방법들이 적용되고 있다. Yang and Kim[12]은 불암산의 생태복원지와 식생보전지 수목의 흉고직경을 측정하여 탄소 저장량 및 흡수량을 산정하여 비교, 분석하였다. Kim et al.[13]은 현장조사와 지상라이다 측정을 통해 수목의 탄소 저장량 및 이산화탄소 흡수량을 평가하였다.

최근에는 기술이 발전하며 현장조사 기법과 더불어 원격 탐사 기법을 사용하기도 한다. 원격 조사를 통해 특정 지역의 탄소량을 예측하기 위해서는 특정 지역의 산림과 비산림지를 분류한 뒤, 이를 수식에 대입하여 탄소 저장량을 예측한다. Lee et al.[14]은 다시기 Landsat TM 영상과 기계학습 알고리즘인 SVM(Support Vector Machine), SVR(Support Vector Regression), 강화 학습(Reinforcement Learning)을 이용하여 산림지를 분류하고 산림 탄소 저장량 변화를 분석하였다. Kim et al.[15]은 높은 정확도를 위해 단순한 회귀 분석에서 벗어나 기계학습 기법을 활용한 고성능 모델 제작을 시도하였다. 고해상도 위성 자료 및 공간정보를 이용하여 랜덤포레스트 기법을 활용한 이산화탄소 흡수량 평가 모델을 제작하였다.

이처럼 탄소 중립과 탄소 저장량에 관한 연구는 기술이 진화함에 따라 다양한 형태와 모델을 활용하며 점차 발전하고 있다. 그뿐만 아니라 컴퓨터 연산 능력의 향상과 고해상도 원격 탐사 자료의 증가는 더욱 광범위한 지역에 대한 상세한 연구를 가능하게 하였다[16].

III. 연구방법

3-1 Method

본 연구에서는 SegFormer[4]의 높은 효율성과 성능을

바탕으로 식생을 분류하고, 픽셀별 탄소 저장량을 예측하는 모델을 제작하였다. SegFormer[4]는 MiT(Mix Transformer Encoders) 인코더를 사용한다. 각 패치가 겹치도록 Embedding 하는 Overlap Patch Embedding 방식을 사용해 입력되는 이미지의 지역적인 정보를 잃지 않게 한다. Embedding 된 정보에서 특징을 추출하기 위해 효율적으로 개선된 트랜스포머 레이어를 통해 특징을 추출한다.

이후 트랜스포머 레이어에서는 Reduction Ratio를 커널 사이즈로 사용한 K, V 계산 바탕으로 Q, K, V의 효율적 연산을 위해 설계되었으며, 입력된 특징을 편향이 없는 컨볼루션 레이어에 통과시켜 Q와 K, V를 출력한다. 그다음 Q, K, V를 재정렬한 뒤 Attention 연산을 수행하고, 마지막으로 다시 재정렬해 최종 출력을 구성한다. 또한, 계층적 구조를 바탕으로 4개의 인코더 트랜스포머 블록에서 High Resolution Coarse Feature와 Low Resolution Fine-Grained Feature를 추출하며 너비와 높이의 크기가 작은 특징맵부터 Upsampling을 통해 보간한다. 이후 이를 하나의 피쳐맵으로 결합시킨 뒤 Segmentation Head를 통과해 결과를 출력한다.

근본적으로 SegFormer[4]는 RGB 이미지에 대한 Semantic Segmentation을 수행하는 모델이므로, 픽셀별 탄소 저장량을 정확히 예측하기 위해 변경된 입력 채널과 세부적인 특징 추출을 위해 적절한 조정이 필요하다. 본 연구에서 제시하는 모델은 SegFormer-B0 모델을 기반으로 연구의 특성에 맞게 다음과 같은 구조 변경을 적용하였다.

먼저 입력 채널의 구조를 수정하였다. 해당 모델에서 사용하는 입력 샘플은 이미지와 임분고 정보가 합쳐진 4채널의 데이터이다. SegFormer[4]는 계층적 구조로 인해 임베딩 과정에서 큰 커널 사이즈로 인해 입력되는 이미지가 입력 너비/4, 입력 높이/4로 감소하고, 레이어의 깊어짐에 따라 추출되는 특징맵의 스트라이드 값에 의해 Feature Map의 크기가 절반씩 감소한다. 기존 방식의 경우 출력 사이즈가 크게 감소해 저장량 예측능력이 감소할 것이며, 실제 탄소 저장량 값과 예측값 사이의 오차가 커질 것이다. 정확한 식생 분류와 탄소량 예측을 위해 초기 임베딩 레이어의 Kernel의 크기를 7에서 4로, 기존의 Overlap Patch Embedding 방식을 따르기 위해 Stride를 4에서 2로 수정했으며, 출력 크기를 맞추주기 위해 padding을 3에서 1로 감소시켰다.

기존에 큰 커널을 바탕으로 정보를 대략적으로 뽑아낸 것에 비해, 본 연구의 모델은 초기부터 좀 더 세부적인 정보를 추출함으로써 계산량은 증가하지만, 모델의 특징 추출 능력이 강화된다. 이에 더해 SegFormer-B0은 경량화를 위해 (2, 2, 2)로 구성되어있는데, 이는 모델의 특징 추출능력이 상대적으로 낮아진다. Segmentation의 정확도에 더해 각 픽셀별 탄소량을 정확하게 예측하기 위해 세부적인 특징을 기존의 인코더 구조보다 특징 추출을 상세하게 수행해야 한다. 따라서 앞서 커널 크기와 Stride를 변경한 것과 더불어 모델의 해석 능력 강화를 위해 각 레이어의 계층 구조를 (2, 2, 6, 3)으로 변경해 주었다. 이로써 Fine-Grained Feature에 대한 해

석 능력을 키워 모델의 성능 향상을 기대할 수 있다.

SegFormer[4]는 Segmentation Task만 수행하므로 탄소량 예측을 위해 MLP(Multi-Layer Perceptron) Head 뒤에 Segmentation Head와 병렬적으로 새로운 Regression Head를 추가하였다. 구조도는 아래와 같다.

기존 SegFormer[4]의 인코더-디코더 구조를 따르며, 인코더 블록의 Layer를 수정해 Fine Grained Feature 추출능력을 증가시켰다. SegFormer와 동일하게 MLP Head뒤 Fusion Block 통과한 후 Segmentation Head와 Regression Head에서 최종 출력인 Segmentation Map과 Regression Map이 나온다. 이 둘을 각각 식생 분류 예측과 탄소 저장량 예측으로 사용된다.

Carbon Head가 추가되면서 탄소 저장량 예측은 Regression 과정이므로 Loss로 MSE Loss(Mean Squared Error Loss)를 사용하며, Segmentation Loss는 Cross Entropy Loss를 사용한다. 변경된 출력구조는 2가지 출력을 예측하는 Multi Task이므로 적절한 로스를 관리하기 위해 Lambda 1과 Lambda 2를 사용한 Loss 업데이트 방식을 사용하였다. 식은 아래와 같다.

$$Loss = CELoss \times \lambda_1 + MSELoss \times \lambda_2 \quad (1)$$

두 Lambda는 0에서 1 사이의 값을 가진다. 본 연구에서는 Segmentation 성능을 바탕으로 Regression의 성능이 보장된다고 전제하였으며, 이에 따라 Lambda 1은 1로 설정해 Segmentation Loss를 모두 업데이트될 수 있게 하였다. Regression Loss는 픽셀별 MSE Loss 값을 측정된 뒤 그 결과를 모두 더한 결과를 사용하게 된다. 따라서 학습 초기에는 업데이트된 Segmentation Loss가 무시될 정도로 큰 값을 가지게 되어 모델이 올바른 방향으로 학습하지 못하고 Local Minimum에 빠질 위험이 크다. 따라서 Lambda 2는 0.0005로 설정해 상대적으로 작은 Step으로 업데이트되게끔 구성하였다. 본 연구에서 최종적으로 사용된 모델은 그림 1로 정리할 수 있다.

3-2 Dataset

본 연구에서는 AIHub의 식생 탄소 포집량 데이터셋[3]을 활용하였다. 식생 탄소 포집량 데이터셋은 수도권 산림지와 도심지의 탄소 저장량을 예측하기 위해 제작된 데이터셋이다. 크게 산림지와 도심지로 구분되며, 세부적으로 1픽셀당 (10cm, 10cm) 해상도의 항공영상, (25cm, 25cm)의 항공영상, (10cm, 10cm)의 NIR영상 그리고 (10m, 10m)의 위성영상을 포함하고 있다. 본 연구에서는 수도권을 대상으로 산림지와 도심지의 탄소 저장량을 알아보기 위하여 1픽셀당 (25cm, 25cm)의 해상도를 가진 산림지, 도심지의 항공 이미지를 활용하였다. 데이터셋은 국토지리정보원에서 제공되었으며, 2022년도 수도권의 공개가능 지역의 촬영본은 25%의 중복률을 가지고 (512, 512) 픽셀로 나눈 데이터이다.

한 샘플은 RGB채널의 이미지와 임분고(SH) 정보를 가지고 있으며, 라벨 데이터는 식생 분류를 위한 GIS 정보와 식생의 라벨 정보, 픽셀당 탄소 저장량을 나타낸 탄소 저장량 지도를 포함한다. 임분고 정보는 해당 픽셀의 식생의 높이를 파악하기 위한 데이터로 해당 데이터셋에서는 한 픽셀당 식생 높이의 평균을 의미한다.

표 1. 식생 정의
Table 1. Vegetation classes

Area	Class Number	Class Name
Forest	0	Unidentifiable
	110	Pine
	120	Larch
	130	Other Conifers
	140	Broadleaf
	190	Non-forested
City	0	Unidentifiable
	110	Pine
	130	Other Conifers
	140	Broadleaf
	210	Street trees
	220	Grassland

해당 데이터는 지구 식생의 높이를 파악하기 위해 2018년 ISS(International Space Station)에 장착된 Spaceborne LiDAR 데이터를 활용했으며, 탐지되지 않은 지역에서는 GIS 프로그램을 사용해 식생, 인공물, 등이 포함된 표면 자료인 수치표면모델(DSM)에서 식생, 인공물이 포함되지 않은 지표 표면 자료인 수치표고모델(DTM)을 뺀 값이 사용되었다. 탄소 저장량을 예측하기 위한 산림 지역의 식생 분류는 표 1과 같다. 산림지의 탄소 저장량은 바이오매스 탄소 저장량 산정식을 사용하였으며, 식은 아래와 같다.

$$\text{산림탄소저장량}(tC) = \text{줄기재적}(m^3) \times D \times BEF \times (1 + R) \times CF \quad (2)$$

식 (2)에서의 줄기 재적은 각 샘플의 픽셀당 임분고 데이터가 사용되었으며, D는 목재기본밀도를 말하며, BEF는 바이오매스 확장계수이며, 계산시 가지와 잎을 포함하도록 하는 상수이다. R은 뿌리-지상부 비율이며, D와 R은 국가 고유 배출, 흡수 계수로 승인된 임상별 고유 계수가 적용되었다.

도심지의 경우 산림지의 바이오매스 탄소 저장량 산정식을 적용하되, 수목의 생장이 저해되는 도심지의 특성을 반영하기 위해 도심 수목 전환계수 0.8이 적용된 것을 탄소 저장량으로 사용하였다. 도심지의 초지와 관목은 연간 바이오매스 탄소 축적 증가량 산정식을 활용해 탄소 저장량을 측정했으며, 식은 아래와 같다.

$$\Delta C_G = A \times G_c \quad (3)$$

ΔC_G 는 연간 바이오매스 탄소 축적 증가량이며, A는 카테고리별 해당 지목의 면적 변화량이다.

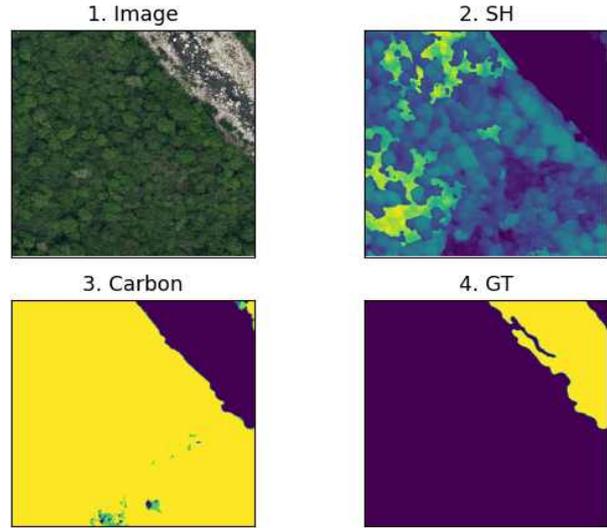


그림 2. 항공 데이터셋 예시(25 cm, 25 cm)
Fig. 2. Aerial dataset sample(25 cm, 25 cm)

마지막으로 G_c 는 면적당 연간 바이오매스 탄소 축적 증가량이다. 즉 지목의 면적 변화에 따른 바이오매스 탄소 축적 증가량이 도심 초지와 관목의 탄소량이 된다.

해당 데이터셋 다음과 같은 전처리과정 후 활용하였다. 먼저 GIS 정보로 존재하는 식생 분류를 GeoPandas를 활용해 샘플과 같은 크기의 Ground Truth로 제작하였다. 산림지와 도심지의 식생 분류가 달라 분류가 같은 소나무, 기타 침엽수, 활엽수는 그대로 활용했으며 산림 지역의 낙엽송과 도심 지역의 가로수를 활엽수로 재분류 해주었다. 도심 지역의 초지, 관목은 산림 지역에서 따로 분류되지 않았고, 산림 지역의 샘플을 확인한 결과 비산림지(非산림지)로 구분했기 때문에 비산림지로 재조정 해주었다. 따라서 최종적인 클래스는 식별불가, 침엽수, 활엽수, 비산림지 4개다.

그림 2는 본 연구에서 사용한 데이터셋의 예시이다. 1번은 입력된 항공 이미지이며, 2번은 한 픽셀당 수목의 평균 높이인 임분고 정보이다. 3번은 해당 이미지의 탄소량을 나타낸 정보이며, 마지막 4번은 식생의 Ground Truth이다. 항공 영상은 위성영상과 비교하면 낮은 시간 해상도를 가지지만 좁은 범위의 지역을 지면 가까이에서 촬영하기 때문에 공간 해상도가 높아 정밀한 예측이 가능하다. 따라서 상세한 산림지 분할과 탄소 저장량 예측에 유리한 모델 제작에 목적을 두고자 1픽셀당 25cm의 해상도를 가지는 항공 데이터를 활용하였다.

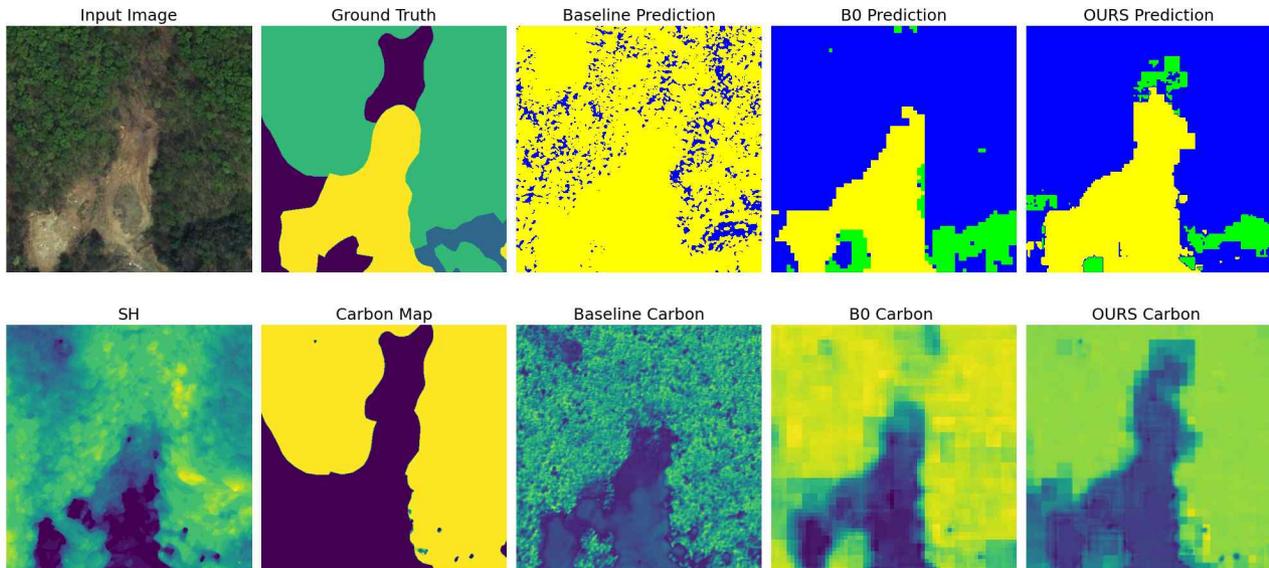


그림 3. 예측 결과
 Fig. 3. Prediction results

IV. 연구 결과 및 분석

성능 검증은 2가지 지표를 사용하였다. 식생 분류 성능은 mIoU (mean Intersection over Union)로, mIoU는 Segmentation 성능을 측정하기 위한 지표로 예측된 분할 영역과 실제 정답 (ground truth) 영역 간의 겹침 정도를 정량화한 것이며, 수식은 아래와 같다.

$$mIoU = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FP_i + FN_i} \quad (4)$$

식 (4)에서 n은 클래스의 개수이며, TP는 각각의 클래스의 True Positive, FP는 False Positive, FN는 False Negative를 나타낸다. 이 지표는 각 클래스에 대한 IoU를 계산한 후 이들의 평균을 구함으로써 산출된다. mIoU는 0에서 1까지의 값을 가지며 1에 가까울수록 더 정확한 분류 성능을 의미한다.

픽셀당 탄소저장량 예측 정확도는 회귀 모델에서 자주 사용되는 RMSE(Root Mean Square Error)로 평가하였다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

식 (5)는 RMSE (Root Mean Squared Error)로, 예측값과 실제값 사이의 오차를 제곱하여 평균을 구한 후 제곱근을 취해 원래 데이터 단위로 표현한 값이다. RMSE 값이 낮을수록 예측값이 실제값에 더 가까워 모델의 예측 오차가 작음을 의미한다. 즉, RMSE가 낮다는 것은 모델이 주어진 데이터를 더 정확하게 예측하고 있음을 뜻한다. 특히 탄소 저장량을 예

측하는 모델에서 RMSE가 낮다는 것은 모델이 실제 탄소 저장량에 더 근접한 값을 예측하고 있음을 보여준다. 이처럼 RMSE는 모델의 예측 정확도를 평가하는 중요한 지표로 사용된다.

표 2는 데이터셋과 함께 제공된 Unet 기반의 탄소저장량 예측모델, SegFormer-B0 모델에 인코더 수정 없이 Carbon Head를 추가하고 Lambda를 통한 Loss 계산으로 학습을 진행한 모델, 본 연구의 모델을 비교한 결과이다. 두 모델 모두 학습에 충분한 Epoch을 반복해 학습했으며, 검증 시 Loss가 가장 낮을 때의 결과를 사용하였다. 본 연구에서 제안한 모델은 SegFormer의 MiT(Mix Transformer Encoders)의 구조를 따라 컨볼루션을 통한 Attention 연산을 바탕으로 경량화되었기 때문에 Baseline 모델 대비 50%의 파라미터를 가진다. 또한, FLOPs는 본 연구의 모델이 Unet의 약 10%의 크기를 가지며, 이는 실제 추론시간에 큰 영향을 미치는 차이이다.

mIoU는 제안한 모델이 0.5145로 Baseline의 0.3602보다 우수한 것으로 나타났으며, RMSE 결과 또한 0.1291로 Baseline의 0.2089보다 뛰어났다. 이를 통해 더 적은 파라미터와 FLOPs를 가져 추론시간도 더 빠르며 높은 mIoU와 낮은 RMSE 점수를 통해 본 연구 모델의 우수성을 확인할 수 있다. SegFormer-B0 모델과 본 연구모델을 비교할 때 인코더의 첫 레이어의 커널 사이즈가 크지만 Stride는 4이며, 인코더의 3번째와 4번째 층의 레이어 개수도 더 적기 때문에 전체적으로 더 적은 파라미터와 FLOPs를 가진다. 하지만 성능 비교했을 때, mIoU에서는 본 연구 제안 모델이 약 10% 높은 점수를 보이며, RMSE에서도 약 10% 낮은 점수를 기록해 본 연구 모델의 우수성을 알 수 있다.

본 연구에서는 식생 분류와 탄소 저장량 예측을 위한 새로운 모델을 제안하며, 기존의 Unet 모델 및 Segformer-B0

모델과 비교하였다. 그림 3은 이러한 비교 결과를 시각적으로 제시한다. 상단 행은 입력 이미지, Ground Truth, 그리고 세 모델의 식생 분류 예측 결과이며, 하단 행은 해당 지역의 수고(SH), 실제 탄소 저장량(Carbon Map), 그리고 각 모델의 탄소 저장량 예측 결과를 나타낸다. Ground Truth에서 녹색은 침엽수, 파란색은 활엽수, 노란색은 비산림지, 보라색은 식별 불가 영역을 의미한다. 탄소 저장량 맵에서는 밝은 색상일수록 높은 저장량을 나타낸다.

표 2. Unet, SegFormer-B0, 연구 모델의 결과
Table 2. Result of Unet, SegFormer-B0 and our model

	Baseline(Unet)	SegFormer-B0	OURS
Params(M)	28.952	7.978	13.335
FLOPs(G)	193.968	4.357	20.792
mIoU	0.3602	0.4629	0.5145
RMSE	0.2089	0.1432	0.1291

Unet 모델의 Segmentation 예측 결과를 살펴보면, 식생 유형을 제대로 구분하지 못하고 있다. 대부분의 영역을 비산림지로 오분류하고 있으며, 일부 지역만을 활엽수로, 산발적으로 분류하고 있다. 이는 Unet 모델이 복잡한 식생 패턴을 인식하는 데 한계가 있으며, 학습 방식에 차이가 있음을 보여준다. Segformer-B0 모델의 예측 결과는 Unet보다 개선되었으나, 여전히 한계를 보인다. 중앙의 비산림 지역은 대체로 잘 식별했으나, 우측 하단의 침엽수 군락을 과대 예측하는 등 정밀도가 부족하다. 본 연구에서 제안한 모델의 예측 결과는 가장 우수한 성능을 보인다. 전체적으로 분포된 활엽수와 우측 하단의 침엽수 군락을 정확히 분류했으며, 중앙의 비산림 지역도 명확하게 식별하였다. 이는 제안된 모델이 복잡한 식생 패턴을 효과적으로 학습하고 분류할 수 있음을 나타낸다. 탄소 저장량 예측 결과를 보면, Unet 모델은 픽셀 단위로 세밀한 예측을 보이나, 실제 탄소 맵과의 차이가 크다. 특히 상단의 식별 불가 영역의 예측 결과와 실제 결과의 차이가 매우 크다. Segformer-B0 모델(B0 Carbon)은 개선된 결과를 보이지만, 여전히 정확도가 떨어진다. 본 연구의 모델(OURS Carbon)은 중앙의 비산림지 탄소량을 낮게 예측하고, 주변 산림 지역의 높은 탄소 저장량을 잘 포착하여 가장 정확한 예측을 보여준다. 특히 Segmentation 결과에 따라 탄소량 예측이 달라지는 것을 확인할 수 있다. 이는 제안된 모델이 식생 유형과 탄소 저장량 간의 관계를 효과적으로 학습했음을 시사한다. 제안된 모델의 예측 결과에서 나타나는 사각형 패턴은 인코더 레이어의 낮은 채널 수로 인한 것이다. 반면, Unet 모델의 결과가 더 세밀해 보이는 것은 더 큰 파라미터 수와 입출력 이미지 크기 때문이다. Segformer-B0 모델의 출력 크기가 작은 것은 더 큰 커널 크기와 Stride 때문이다. 결론적으로, 본 연구에서 제안한 모델은 식생 분류와 탄소 저장량 예측 모두에서 가장 우수한 성능을 보이며, 특히 다양한 식생 유형이 혼재된 복잡한 지역에서도 높은 정확도를 유지

하고 있다.

이러한 점수 향상은 두 가지 측면에서 발생했는데, 첫 번째는 컨볼루션 기반 모델 구조보다 트랜스포머 기반 모델의 우수성이다. 트랜스포머 기반 모델은 Segmentation을 수행할 때, 컨볼루션 기반 모델과 달리 이미지 내 다양한 위치 간의 상관관계를 동시에 고려할 수 있어, 주변 지형과 관련된 정보를 더 효과적으로 포착할 수 있다. 따라서, 지형적 특성과 객체 간의 관계에 기반한 정밀하고 일관성 있는 분할이 가능해진다. 이는 탄소 저장량 예측에서도 효과적인데, 어텐션 연산을 통해 장거리 의존성을 학습함으로써, 단순한 픽셀 단위 예측을 넘어 넓은 지역에 걸친 공간적 상호작용과 패턴을 파악할 수 있기 때문이다. 트랜스포머 모델은 이미지 내의 모든 픽셀 간 관계를 고려하여, 개별적인 위치의 탄소 저장량뿐만 아니라 주변 지역의 지형적 특성까지 종합적으로 반영해 탄소 저장량 예측을 수행한다.

두 번째는 연구 모델의 첫 번째 레이어인 Overlap Patch Embedding 과정이다. 첫 번째 레이어는 저수준 특징 추출뿐만 아니라, 전체적인 학습 진행의 안정성과 효율성의 기초가 되기 때문에 계층적 구조의 딥러닝 모델에서 중요하다. 기존 MiT 구조의 첫 번째 레이어는 넓은 수용 영역을 확보하기 위해 커널 크기를 7로 설정했으며, 이로 인해 높은 연산량을 요구하게 되어 스트라이드를 4로 설정하였다. 일반적인 Segmentation 작업에서는 이러한 큰 커널 크기로 넓은 영역을 한 번에 포착하는 것이 효과적일 수 있다. 그러나 본 연구는 항공 영상 Segmentation 및 픽셀별 탄소 저장량 예측을 수행하기 때문에, 적절하지 않다.

항공 영상은 매우 높은 공간 해상도를 가지며, 본 연구에서 사용한 데이터셋 또한 1픽셀당 25cm의 고해상도를 가지기 때문에 작은 객체와 복잡한 지형 패턴을 포착하는 것이 중요하다. 기존 MiT의 첫 번째 레이어의 Patch Embedding 레이어는 커널 크기 7, 스트라이드 4로 설정되었다. 이는 넓은 영역의 정보를 포착할 수 있으나, 공간 해상도가 높은 영상에서는 세부 정보가 손실될 위험이 크다. 이에 따라, 본 연구에서는 세부 정보 손실을 최소화하고, 기존의 Patch Embedding 방식의 장점을 유지하기 위해 커널 크기 4, 스트라이드 2로 설정하였다.

이로써, 모델은 정밀한 지역적 정보를 바탕으로 기존 구조보다 세부적인 특징을 더 효과적으로 추출할 수 있게 된다. 레이어가 깊어지면서, 초기에 정밀하게 추출된 저수준 특징들이 상위 레이어에서 고수준 특징으로 통합되며, 이를 통해 더욱 정확한 특징 추출이 가능해진다. 또한 출력 특징맵의 해상도를 높여 원본 이미지의 세부 정보를 충실히 반영하게 하였다. 결과적으로, 이러한 개선은 모델이 작은 객체와 다양한 지형 패턴을 더욱 정확하게 학습할 수 있게 하여 Segmentation 성능을 향상에 기여한다. 이러한 결과는 제안된 모델이 실제 환경에서의 탄소 저장량 예측 및 모니터링에 효과적으로 활용될 수 있음을 시사한다.

V. 결 론

기존의 탄소 저장량 측정의 비용 및 정확도 문제를 해결하기 위해, 본 연구는 산림지 및 도심지 의미론적 분할 및 탄소 저장량을 예측하기 위해 항공 이미지와 트랜스포머 구조를 기반으로 한 SegFormer를 활용하였다. 임분고 정보를 함께 입력으로 활용하였으며, 항공 영상에서의 분류 정확도를 높이기 위해 기존 모델의 임베딩 커널 사이즈를 줄였으며, 강건한 학습을 위해 패치를 교환하는 방식을 사용하여 트랜스포머 구조가 항공 영상을 기반으로 한 산림지 및 도심지 의미론적 분할과 저장량 예측에 효과적임을 알 수 있다. 또한, 산림지 및 도심지를 분할 후, 저장량을 예측하는 연쇄적 방식이 아닌 하나의 특징 추출부의 출력을 바탕으로 두 가지 Task를 독립적으로 진행하는 방식을 사용하여 우수한 활용성을 가진다. 연구 과정에서 고해상도에서 저해상도 위성영상으로 전이학습을 진행하였지만, 사용 가능한 위성영상 이미지가 매우 적어 활용하지 못한 한계가 있다. 높은 정확도를 가진 모델인 만큼 추후 전이학습, 헤드 구조 변경, 디코더 레이어 추가 등을 통해 위성 영상에도 해당 모델을 적용한다면 범용성이 높고 높은 시간-공간 해상도를 가진 모델을 제작할 수 있을 것이다.

감사의 글

본 연구는 2022년 강남대학교 교내연구비 지원에 의해 수행되었음.

참고문헌

[1] Environmental Preservation Association. Climate Change in Korea [Internet]. Available: <https://www.gihoo.or.kr/menu.es?mid=a30101030000>.

[2] H. S. Eggleston, L. Buendia, K. Miwa, T. Ngara, and K. Tanabe, 2006 IPCC Guidelines for National Greenhouse Gas Inventories, Institute for Global Environmental Strategies, Hayama, Japan, July 2006.

[3] Neighbor System. Vegetation Carbon Capture Identification Data [Internet]. Available: <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=data&dataSetSn=71449>.

[4] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," in *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS '21)*, Online, pp. 12077-12090, December 2021. <https://doi.org/10.48550/arXiv.2105.15203>

[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, ... and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv:2010.11929v1, October 2020. <https://doi.org/10.48550/arXiv.2010.11929>

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... and Illia Polosukhin, "Attention Is All You Need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*, Long Beach: CA, pp. 6000-6010, December 2017. <https://doi.org/10.48550/arXiv.1706.03762>

[7] H. Myung, S. Kim, K. Choi, D. Kim, G. Lee, H. Ahn, ... and B. Kim, "Diagnosis of the Rice Lodging for the UAV Image Using Vision Transformer," *Smart Media Journal*, Vol. 12, No. 9, pp. 28-37, October 2023. <https://doi.org/10.30693/SMJ.2023.12.9.28>

[8] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, ... and R. Girshick, "Segment Anything," in *Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, pp. 3992-4003, October 2023. <https://doi.org/10.1109/ICCV51070.2023.00371>

[9] B. A. Khan and J.-W. Jung, "Semantic Segmentation of Aerial Imagery Using U-Net with Self-Attention and Separable Convolutions," *Applied Sciences*, Vol. 14, No. 9, 3712, April 2024. <https://doi.org/10.3390/app14093712>

[10] P. Anilkumar, P. Venugopal, P. K. R. Maddikunta, T. R. Gadekallu, A. Al-Rasheed, M. Abbas, and B. O. Soufiene, "An Adaptive DeepLabv3+ for Semantic Segmentation of Aerial Images Using Improved Golden Eagle Optimization Algorithm," *IEEE Access*, Vol. 11, pp. 106688-106705, 2023. <https://doi.org/10.1109/ACCESS.2023.3318867>

[11] T. Hanyu, K. Yamazaki, M. Tran, R. A. McCann, H. Liao, C. Rainwater, ... and N. Le, "AerialFormer: Multi-Resolution Transformer for Aerial Image Segmentation," *Remote Sensing*, Vol. 16, No. 16, 2930, August 2024. <https://doi.org/10.3390/rs16162930>

[12] K. C. Yang and J. S. Kim, "Carbon Storage and Absorption of Trees in the Ecological Restoration Area and Vegetation Conservation Area of Bulamsan Urban Nature Park," *Ecology and Resilient Infrastructure*, Vol. 7, No. 4, pp. 284-293, December 2020. <https://doi.org/10.17820/ERI.2020.7.4.284>

[13] H. K. Kim, Y. S. Hong, Y. K. Lim, I. S. Yun, K. S. Do,

C. H. Jung, ... and C.-B. Kim, "Estimation of Carbon Stock and Annual CO2 Uptake of Four Species at the Sejong National Arboretum -Pinus densiflora, Metasequoia glyptostroboides, Aesculus turbinata, Chionanthus retusus-," *Journal of Environmental Impact Assessment*, Vol. 32, No. 1, pp. 41-48, February 2023. <https://doi.org/10.14249/eia.2023.32.1.41>

[14] J.-H. Lee, J.-H. Im, K.-M. Kim, and J. Heo, "Change Analysis of Aboveground Forest Carbon Stocks according to the Land Cover Change Using Multi-Temporal Landsat TM Images and Machine Learning Algorithms," *Journal of the Korean Association of Geographic Information Studies*, Vol. 18, No. 4, pp. 81-99, December 2015. <https://doi.org/10.11108/kagis.2015.18.4.081>

[15] M. Kim, J. U. Park, J. Park, J. Park, and C.-U. Hyun, "Predicting the Effects of Rooftop Greening and Evaluating CO2 Sequestration in Urban Heat Island Areas Using Satellite Imagery and Machine Learning," *Korean Journal of Remote Sensing*, Vol. 39, No. 5-1, pp. 481-493, October 2023. <https://doi.org/10.7780/kjrs.2023.39.5.1.2>

[16] J. Shin, S. Jeong, and D. Chang, "Estimation of Forest Carbon Stock in South Korea Using Machine Learning with High-Resolution Remote Sensing Data," *Atmosphere*, Vol. 33, No. 1, pp. 61-72, February 2023. <https://doi.org/10.14191/Atmos.2023.33.1.061>

우승우 (Seung-Woo Woo)



2022년~현재 : 강남대학교 인공지능융합공학부 인공지능전공 학사과정

※ 관심분야 : 딥러닝(Dep Learning), 컴퓨터 비전(Computer Vision) 등

장은서 (Eun-Seo Jang)

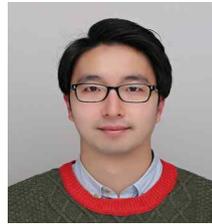


2022년 : 송의여자대학교 문헌정보과 (문헌정보관리전문학사)

2022년~현재 : 강남대학교 소프트웨어융합학과 학사과정

※ 관심분야 : 딥러닝(Dep Learning), 빅데이터 분석(Bigdata Analysis)

곽찬희 (Chanhee Kwak)



2020년 : KAIST 경영공학 (경영공학박사-IT경영)

2020년~현재 : 강남대학교 인공지능융합공학부 조교수

※ 관심분야 : 정보보호(Personal Information), 머신러닝(Machine Learning)