

Segmentation 기반 전동킥보드 주차/비주차 구역 분류 기술의 개발*

조 용 현*, 최 진 영**

요 약

본 논문은 공유형 전동킥보드 시스템 운영 시, 관리 상 발생할 수 있는 주차 문제를 해결하기 위해 반납 인증사진으로 주차, 비주차 구역을 판단하는 AI모델을 제시한다. 본 연구에서는 주차/비주차 구역 배경 관련 객체를 판별하기 위해 ADE20K에 Pre-trained된 Segformer_b0 모델과 점자블록, 전동킥보드에 Fine_tuning한 Segformer_b0 모델을 통해 주차/비주차에 관련된 객체의 Segmentation map을 추출하고, Swin 모델을 통해 주차/비주차 구역을 이진 분류하는 방법을 제시하였다. 최종적으로 총 1,689장을 직접 라벨링한 후 진행한 Fine_tuning SegFormer 모델은 mAP가 81.26% 수준으로 전동킥보드와 점자블록을 인식하였으며, 총 2,817장을 훈련한 Classification 모델은 92.11%의 정확도와 91.50%의 F1-Score로 주차구역과 비주차 구역을 분류하는 것이 가능하였다.

Development of segmentation-based electric scooter parking/non-parking zone classification technology

Yong-Hyeon Jo*, Jin Young Choi**

ABSTRACT

This paper proposes an AI model that determines parking and non-parking zones based on return authentication photos to address parking issues that may arise in shared electric scooter systems. In this study, we used a pre-trained Segformer_b0 model on ADE20K and fine-tuned it on tactile blocks and electric scooters to extract segmentation maps of objects related to parking and non-parking areas. We also presented a method to perform binary classification of parking and non-parking zones using the Swin model. Finally, after labeling a total of 1,689 images and fine-tuning the SegFormer model, it achieved an mAP of 81.26%, recognizing electric scooters and tactile blocks. The classification model, trained on a total of 2,817 images, achieved an accuracy of 92.11% and an F1-Score of 91.50% for classifying parking and non-parking areas.

Key words : Personal Mobility, Image Segmentation, Parking/Non-parking zone classification

접수일(2023년 11월 21일), 게재확정일(2023년 12월 26일)

* 부산대학교 정보컴퓨터공학부 석사과정(주저자)

** 동신대학교 기계공학과 교수(교신저자)

★ 본 논문은 2023년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과입니다.(2021RIS-002)

1. 서 론

최근에는 전동 킥보드, 전동 스쿠터 또는 전동 자전거와 같은 개인용 전동기(PM)을 이용하는 사용자가 증가하고 있으며, 특히 도심지에서는 건강을 위해서는 물론, 대중 교통 수단을 대체하는 교통 수단으로서 많은 사람이 이용하고 있다. 이러한 개인용 전동기(PM)는 일반적인 이동수단과 달리 화석연료를 사용하지 않기 때문에 환경오염이 발생되지 않고, 근거리 통학 및 출퇴근 수단으로써 경비절감과 출퇴근 시간 절약 또는 근거리 이동에서 그 활용성이 급격하게 늘어나고 있다.

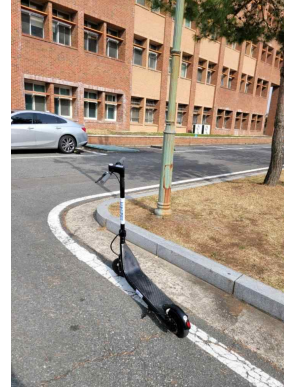
2018년 이후 국내 공유형 전동킥보드 업체가 성장함에 따라 전동킥보드가 새로운 교통수단으로써 각광받고 있으며, 공유 시설 업체가 늘어나며 전동킥보드 대여수와 이용자가 증가하였다. 하지만 이후부터 현재까지 공유킥보드 주차, 무단 방치에 관한 문제가 커지고 있다[1]. 공유킥보드는 아파트 단지, 초등학교와 중학교와 같은 기초 교육 시설만 아니라면 어디에서든지 반납할 수 있어 지역 곳곳에 퍼져 주차되어 있으며, 이러한 공유 킥보드들은 이용자가 보도 위에 무분별하게 주차, 인도나 도로 위에 쓰러져 주차함에 따라 차량 운행, 보행의 방해요소로 작용하여 민원의 불만이 커지고 있다[2].

해결방안으로 서울특별시 정차, 주차 위반차량 견인 등에 관한 조례 일부 개정조례안에서 서울시는 사고 발생 가능성이 높은 구역에 방치된 개인형 이동장치에 5만원의 견인료를 부과하며, 50만원 한도 내 30분당 700원의 보관료를 해당 업체에 매기는 법을 시행하고 있다[3]. 또한 개인형 이동장치 거치 시설이나 주차구역을 마련하는 등 다양한 방안이 제안되었다.

한계는 해당 규모에 비해 현실적인 시행 기준과 관리 측면이 부실하다고 평가되고 있으며, 제도적인 측면과 함께 기술적으로 주차 질서 확립할 수 있는 기술적 연구가 필요하다.

본 논문에서는 대학 캠퍼스 내에 사용되는 자체 공유형 킥보드를 관리하기 위해 주차, 비주차구역을 분류하는 AI모델을 제안한다. 먼저 전동킥보드의 주차·비주차구역의 특징을 추출하기 위해 도로교통법 및 관련 지침, 지자체별 가이드 라인, 대학 캠퍼스 내에

구조, 배경을 참고하여 주차·비주차구역에 관련된 객체를 선별한다[4].



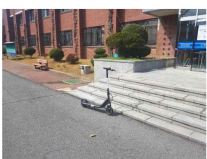


(그림 1) 대학 캠퍼스 내의 공유형 전동킥보드 모델

도로교통법 제32조 및 제33조에 현행 주·정차 제한구역, 대통령 직속 4차산업혁명위원회의 주·정차 제한구역, 서울시 공유 PM 업체 이용질서 확립 및 활성화 MOU에 따른 주·정차 제한구역 가이드라인에 따라 주차 관련 특징을 갖는 객체는 벽, 빌딩, 벤치, 가로등, 전봇대, 도로, 도보, 계단, 점자블록, 전동킥보드로 총 10개를 선별하였고 주차·비주차구역을 분류하는 8가지 기준을 두었다.

주차구역이라 판단하는 기준은 (1) 벽, 빌딩 옆, (2) 도보 옆, (3) 벤치 옆, (4) 가로등, 전봇대 옆이고, 비주차구역이라 판단하는 기준은 (1) 도로 안, (2) 도보 중앙, (3) 계단 진입로 근처, (4) 점자블록 위로 예시는 표 1과 같다.

AI 시스템의 주차·비주차구역을 판단하는 과정은 먼저 ADE20K 데이터셋에 Pre_trained Segmentation 모델을 활용하여 전동킥보드 주차·비주차구역과 관련된 벽, 빌딩, 도보, 벤치 등의 객체 특징을 추출하며, ADE20K classes에 포함되어 있지 않는 전동킥보드와 점자블록 객체는 실제 사진을 수집한 후 라벨링 작업을 진행하여 Segmentation 모델에 Fine_tuning을 진행한다. 따라서 최종 두 Segmentation 모델로 선별된 객체의 Segmentation map을 추출하고 Image Classification 모델로 주차구역과 비주차구역을 분류할 수 있도록 한다.

<표 1> 주차·비주차구역 판단 예시

주차구역	비주차구역
	
(1) 벽, 빌딩 옆	(1) 도로 안
	
(2) 도보 옆	(2) 도보 중앙
	
(3) 벤치 옆	(3) 계단 진입로 근처
	
(4) 가로등, 전봇대 옆	(4) 점자블록 위

본 논문은 Segmentation, Image Classification을 활용한 전동키토보드 주차 상태 및 주차, 비주차를 분류하는 AI 모델을 제안하며, 제안된 모델은 공유키토보드 반납 시 찍어야 하는 인증사진에 적용함으로 대학 캠퍼스 내에 전동키토보드 주차 문제를 해결할 수 있을 것으로 기대한다.

2. 관련 연구

2.1 Transformer

NLP 영역에서는 sequence transduction 모델에 encoder, decoder를 포함하는 복잡한 순환 또는 CNN(Convolutional Neural Network)이 기초가 되었다. 하지만 2017년 Vaswani 등의 연구에서 단지

attention mechanisms만 사용되는 단순한 network 구조의 Transformer가 제안되었으며[5], 기존 LSTM, GPU 모델의 Sequence 길이가 길어질수록 Long-term dependency가 취약해지는 단점을 보완하였다.

Vaswani 등의 연구[5]에서는 기존 순차적으로 데이터를 다루는 RNN, LSTM과는 다르게 가중치로부터 embedding 된 단어 벡터들에 Query, Key, Value 값을 구하고 Query, Key에 score 함수를 적용하여 Attention score를 구한 후 식 (1)과 같이 Attention Value를 구한다. (Scaled Dot-Product Attention)

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (1)$$

여기서 Attention 값은 단어 벡터들의 전 위치에 score 함수를 적용하기 때문에 global한 정보를 공유하여 Long-term dependency의 문제점을 해결하였다. 그리고 NLP에 중요한 Position 정보 또한 식 (2)와 같이 단어 벡터에 PE(Position Embedding)를 연산하여 sequence 정보도 주었으며 성능을 개선하기 위해 Attention 함수의 구조에 num_head를 설정하여 병렬화 과정도 진행하였다.

$$\begin{aligned} PE_{(pos, 2i)} &= \sin(pos/10000^{2i/d_{model}}) \\ PE_{(pos, 2i+1)} &= \cos(pos/10000^{2i/d_{model}}) \end{aligned} \quad (2)$$

이러한 Attention 함수만을 사용한 Transformer 구조는 기존 CNN, RNN을 벗어난 단순하고 신선한 아이디어였으며, 이후 Vision 분야에도 구조로 다양한 연구가 진행되었다. 대표적으로 Visual Transformer(ViT)[6]는 Transformer 구조를 활용하여 컴퓨터 비전 Image Classification 영역에서 좋은 성능을 달성한 예다.

현재는 컴퓨터 비전의 Classification, Segmentation, Object Detection 등의 다양한 분야에서도 Transformers 구조가 제안되고 있으며, 최근에는 Transformer 구조의 [7,8] 모델 등이 제안되며 SOTA(State of the Art)를 달성하였다.

2.2 Swin Transformer[8]

Swin Transformer는 Transformer 구조로 컴퓨터 비전의 여러 분야의 backbone을 제공하기 위해 제안되었다. 대표적인 아이디어는 이미지를 local window(patch)로 나누고 local window 안에서만 self-attention(W-MSA)을 수행한 후 shift하여 local window 간의 self-attention(SW-MSA)을 수행하는 방법이다.

먼저 W-MSA는 식 (3)과 같이 기존 ViT의 MSA와 비교하였다.

$$\begin{aligned} \Omega(MSA) &= 4hwC^2 + 2(hw)^2C \\ \Omega(W-MSA) &= 4hwC^2 + 2M^2hwC \end{aligned} \quad (3)$$

W-MSA 식의 M(Local Window Size)값은 모델 구조에 따라 이미 고정된 실수이기 때문에 기존 MSA의 quadratic한 연산을 linear하게 만들었으며, 이로 인해 연산량을 크게 줄일 수 있게 되었다.

다음 SW-MSA는 local window를 M/2(window size//2) 만큼 shift해서 local window간 정보를 공유할 수 있도록 self-attention을 수행하였다. 이 과정에서 불필요한 연산을 줄이기 위해 Cyclic Shift와 Attention Mask를 진행하여 W-MSA와 동일한 window 개수로 줄일 수 있었다.

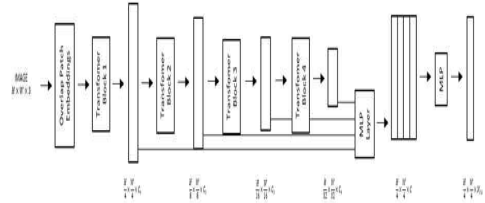
Swin Transformer 모델은 Classification, Object Detection, Segmentation 등 다양한 컴퓨터 비전 분야의 backbone으로 활용되었으며, 특히 Classification에서 Swin-L은 ImageNet dataset에 87.3% Accuracy와 197M Parameter로 우수함을 보였다.

2.3 SegFormer[7]

Semantic Segmentation은 이미지 분류에서 확장된 영역으로 픽셀별로 객체를 분류하기 때문에 이미지 내의 구체적인 객체를 인지할 수 있다는 장점이 있으며 대표적으로 자율 주행 자동차, 뇌 종양 분야에 활용되고 있다.

Segmentation 영역에서 2021년에 발표한 SegFormer 모델[7]은 encoder에 Transformer 구조를 활용하여 제안되었다. 기존 Transformer 구조가 기 초된 방법이 매우 효율적이지 않고, 실시간 어플리

케이션으로 배포하기 어렵다는 문제점을 해결하고 자 단순한 두 모듈의 구조로 (그림 2)의 파이프라인과 같이 제안되었다.



(그림 2) SegFormer 파이프라인

두 모듈은 Hierarchical Transformer encoder와 Lightweight ALL-MLP decoder이다.

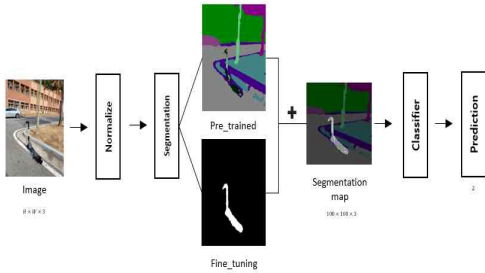
먼저 Hierarchical Transformer encoder는 기존 PE(Position Embedding)에 resolution이 고정되어 다른 resolution의 데이터 경우 보간법을 이용하여 성능이 떨어진다는 문제를 해결하고자, Mix-FFN이 제안되었으며, patch merging 과정에서 4x4 patch size로 computation cost가 커지는 문제점을 해결하기 위해 Reduction ratio와 linear 함수를 활용해 feature 차원을 줄여주기도 하였다.

Lightweight ALL-MLP decoder는 encoder에서 추출된 feature level 값들을 연결하여 local attention과 global attention의 관계도 고려하였으며, Lightweight의 MLP layer로 통일하여 단순한 구조로 만들었다.

SegFormer는 이전에 제안된 Segmentation 모델의 복잡한 구조를 피했으며, 가벼운 구조로 높은 효율과 성능을 보였다. 대표적으로 ADE20K dataset에 SegFormer_B5는 51.8%의 높은 mIoU를 달성함과 동시에 183.3(G)의 Flops의 적은 연산량을 보였다. 이에 다양한 downstreaming에 SOTA(State of the Art)를 달성했을 뿐 아니라 zero-shot robustness에서도 강건함을 보였다.

3. AI 모델 제안

본 논문에서 제안하는 AI 모델의 구성도는 (그림 3)과 같다.



(그림 3) 제안된 AI 모델의 블럭도

첫째, $H \times W \times 3$ 의 해상도를 가지는 이미지는 먼저 전처리 과정을 거치며 값의 범위가 평균 0, 분산 1의 값으로 변환된다. 둘째, ADE_20K의 Pre_trained 모델의 Segmentation map과 전동킥보드, 점자블록에 Fine_tuning 모델의 Segmentation map을 병렬 과정으로 계산한다. 셋째, Pre_trained 모델과 Fine_tuning 모델에서 분류된 두 Segmentation map을 제한된 크기의 맵($100 \times 100 \times 1$)으로 추출한다. 넷째, Classifier 모듈로 주차구역과 비주차구역을 분류한다.

3.1 Pre-trained SegFormer

Semantic Segmentation 데이터셋은 주차에 관련된 객체('wall', 'building', 'road', 'sidewalk', 'sign board', 'stairs', 'bench', 'streetlight', 'stairway', 'traffic')를 포함하고 있는 ADE20K 데이터셋을 활용하였으며, 실시간으로 구현하기 위해 ADE20K 데이터셋에 37.4%의 mIoU를 달성함과 동시에 8.4(G)의 Flops의 적은 연산량을 보였던 SegFormer_B0 모델을 본 AI 모델 Segmentation 과정에 적용하였다.

ADE20K Pretrained SegFormer_B0 모델은 150가지의 객체들을 픽셀 당 인식하여 길거리 사진을 (그림 4)와 같이 나타낸다.

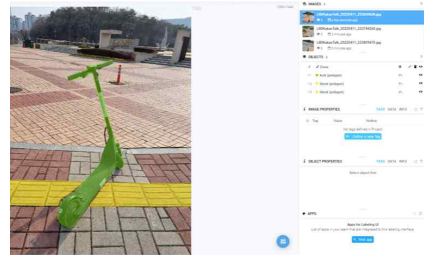


(그림 4) Pre-trained SegFormer_B0 출력값

3.2 Fine-tuning SegFormer

Fine tuning 모델 또한 SegFormer 모델의 가벼운 B0의 인코더를 사용하였으며, 전동킥보드와 점자블록 객체를 인식하기 위해 Segmentation 모델의 Fine tuning 과정을 진행하였다.

실제 전동킥보드와 점자블록이 포함된 사진을 수집하였고, 이에 annotation은 <https://supervise.ly/>에 도구를 이용해 사진을 불러오고 (그림 5)와 같이 Segmentation 라벨링을 하였다.



(그림 5) 전동킥보드, 점자블록 Segmentation labeling 과정

이 과정을 통해 Segformer_B0 모델에 Fine tuning을 진행할 수 있도록 학습 데이터셋을 구축하였다.

3.3 Classification

Classification은 객체 정보와 위치 정보를 담고 있는 Segmentation map을 통해 주차구역과 비주차구역을 분류하는 역할을 한다. 모델에 처음 입력된 사진은 각각 Pre_trained SegFormer 와 Fine_tuning SegFormer의 과정에서 두 Segmentation map을 가지는데, 본 장에서는 이를 요소별 연산(element-wise operation)으로 더하고 $100 \times 100 \times 1$ 크기로 보간법을 한 Segmentation map을 Classification의 입력값으로 사용하였다.

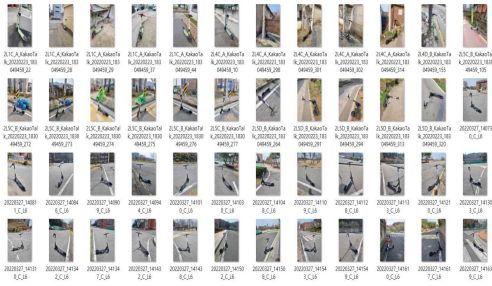
$100 \times 100 \times 1$ 크기의 Segmentation map으로 나타난 이유는 Classification 과정에서 입력 해상도가 작으면 연산량이 크게 줄어드는 장점이 있으며, 또한 본 모델 구성상 Segmentation 과정에서 이미 배경 관련 객체 특징이 추출되어 작은 해상도여도 높은 성능을 보였기 때문이다. 따라서 본 연구에서

는 Segmentation map을 100*100*1의 작은 사이즈로 고정하였다.

최종 주차·비주차구역을 이진 분류하는 모델은 Transformer 구조에 선형적인 계산량을 가지는 S win[8] 모델을 적용하였다.

4. 모델 훈련 및 평가 결과

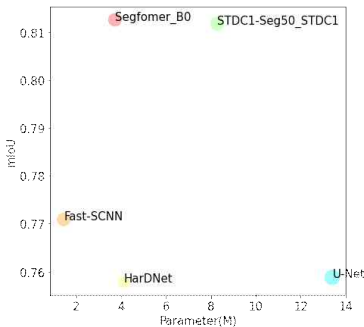
(그림 6)과 같이 대학 캠퍼스 내, 외 반경 2km에 대학이 보유한 자체 공유키패드와 그 외 다른 업체 토크보드의 주차 사진을 통해 훈련용 데이터 사진을 수집하였다.



(그림 6) 훈련용 데이터셋 샘플 이미지

4.1 Fine_tuning SegFormer 결과

전동키패드와 점자블록을 포함하는 총 1,689장의 사진을 수집하였다. 이에 전동키패드와 점자블록 객체의 라벨링 과정을 거치며 SegFormer_B0_Cityscape_1024*512 모델의 fine tuning을 진행하였다. 데이터 셋에 관해 20%는 검증 데이터, 30%는 테스트 데이터, 50%는 훈련 데이터로 나눠 진행하였다.



(그림 7) 모델별 mIoU, Parameter 결과

Segfomer_B0모델은 Batch 4, epoch 8로 fine tuning을 진행하였으며, 비교할 Parameter(M)가 15 이하인(Fast-SCNN[9], HarDNet[10], STDC1-Seg50_STDC1[11], U-Net[12]) Segmentation 모델들도 훈련을 진행하였다. 모델별 mIoU, Parameter (M)의 결과는 (그림 7), <표 2>와 같다.

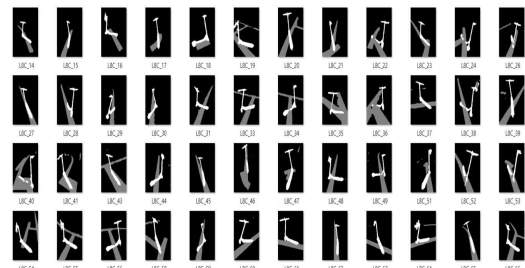
<표 2> 모델별 mIoU와 Parameter(M) 결과

Model	mIoU	Parameter(M)
Segfomer_B0[7]	81.26%	3.72
Fast-SCNN[9]	77.09%	1.44
HarDNet[10]	75.78%	4.12
STDC1-Seg50_STDC1[11]	81.18%	8.28
U-Net[12]	75.88%	13.41

본 훈련 과정에서 Fast-SCNN은 77.09%, HarDNet은 75.78%, STDC1-Seg50_STDC1은 81.18%, U-Net은 75.88%의 mIoU를 달성하였다. 그 중 SegFomer_B0은 3.72 (M)의 적은 파라미터를 구성함과 동시에 81.26% mIoU로 좋은 성능을 보였으며, 특히 UNet과 비교해보았을 때, 파라미터수가 9.69 (M)정도 적었고, mIoU는 5.38% 더 높았다.

모델 중에서 SegFomer_B0은 가장 높은 mIoU를 달성하였으며 Parameter는 두번째로 적게 구성되어 정확도와 효율성에 가장 좋은 균형을 보였다.

Fine_tuning한 SegFomer_B0 모델의 Test data Sampling 결과는 (그림 8)과 같다.

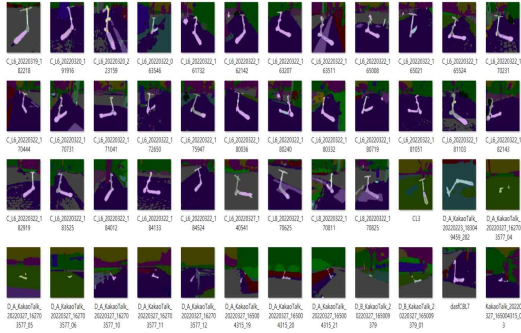


(그림 8) Fine_tuning한 SegFomer_B0 Segmentation 결과

4.2 주차, 비주차 구역 이진 분류 결과

대학 캠퍼스 내에 총 2,817장의 사진을 수집하였으며, Pre_trained SegFormer 와 Fine_tuning Seg

Former의 Segmentation map을 (그림 9)와 같이 100*100*1 크기의 Segmentation map으로 가공하였다.



(그림 9) 100*100*1 크기를 가지는 Segmentation map

주차구역과 비주차구역을 분류하는 기준 별 사진 수는 직접 파일 이름을 수정하여 구분하였으며 <표 3>과 같다.

<표 3> 주차구역, 비주차구역 데이터 세트 구성

Structure of data set	Samples
벽, 빌딩 옆	316
도보 사이드	292
벤치 근처	324
가로등 옆	476
도로 안	323
도보 중앙	432
계단 근처	280
점자블록	268
기타	106
총	2,817

데이터 셋에 관해 20%는 검증 데이터, 30%는 테스트 데이터, 50%는 훈련 데이터로 나누고, Swin 모델은 batch 128, epoch 60, embedding_dim 64, num_mlp 128, num_head 8로 변수를 지정하였다.

일반적인 Swin 모델 훈련 과정과는 다르게 입력값이 1 Channel(100*100*1)이기 때문에 더 적은 계산량으로 학습 과정이 진행되었다. 훈련과정의 Epoch별 loss는 (그림 10)과 같다.



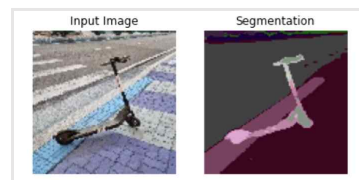
(그림 10) 훈련, 검증 data 별 loss, Epochs 그래프

Epochs 별 train, val Loss는 Epochs가 증가할수록 줄어들었지만, 40 Epoch 이후에 val_loss와 train_loss의 미분값이 0에 가까워지는 것을 볼 수 있다. 이에 Swin 모델은 40 Epoch에서 훈련된 파라미터를 사용하였다. (40 Epoch에서 train_loss는 약 0.26을 달성하였으며, val_loss는 약 0.36을 달성하며 낮은 손실을 보였다.)

$$F1\ Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (5)$$

최종적으로 주차·비주차 구역을 판단하는 AI 모델의 분류 과정을 훈련한 후 테스트 데이터셋에 따른 정량적 평가를 진행해 보았을 때, Test acc = 92.11%, Test loss = 0.33을 보였으며 F1 score는 식 (5)에 따라 91.50%의 성능을 나타내었으며 (그림 11)과 같이 주차구역과 비주차구역을 분류하였다.

```
predict("C:/Users/yonghyeonjo/Workspace/classification/data2/LBC_68.jpg")
점등같은 비주차구역 사진입니다.
예측값은 99 % 비주차구역 사진입니다.
```



(그림 11) 전동킥보드 주차 사진에 제안된 시스템 적용 예시

5. 결 론

본 논문은 전동키토드의 주차 관련 문제를 해결하기 위해 반납 인증사진으로 추정차구역 및 주차 상태를 판단하는 AI 모델을 개발하였다.

시스템 과정은 Pre_trained 모델과 전동키토드, 점자블록 Fine_tuning 모델을 이용하여 주차·비주차에 관련된 객체를 인식하였고, 제안된 Classification 모듈을 통해 전동키토드 주차 상태에 따른 분류를 진행하였다.

본 논문은 어플리케이션으로 사용하기 위해 적은 연산량으로 좋은 성능을 보인 SegFormer_B0 모델을 적용하였으며, Pre_trained SegFormer 와 Fine_tuning SegFormer 과정을 병렬로 진행하였다. 또한 Image Classification 과정에서도 선형적인 계산량을 가지는 Swin 모델을 적용하였으며, Segmentation map을 100*100*1 크기로 제한하며 적은 파라미터를 사용하여 구성하였다.

최종 정량적 평가를 진행하였을 때, SegFormer 모델을 통해 전동키토드, 점자블록을 Fine_tuning 하는 과정에서는 3.72(M)의 적은 파라미터 수로 81.26% mAP인 높은 정확도를 보였으며, Classification 모델은 정확도 92.11%, F1_Score 91.50%로 높은 성능을 보이며 주차구역과 비주차구역을 분류하는 것을 볼 수 있었다. 향후에는 주차·비주차구역에 필요한 다양한 객체를 통해 전이학습과 강화 훈련을 진행함으로써 모델에 성능과 활용성을 높이는 연구를 수행할 것이다.

참고문헌

- [1] The Korea Transport Institute, "A Study Improving Legislation for Promoting Utilization of Personal Mobility and its Safety Management", Research report, Feb. 2020.
- [2] The Korea Research Institute for Human Settlements, "U.S. shared electric scooter management policy and implications", Research report, Nov. 2019.
- [3] "서울특별시 정차·주차위반차량 견인 등에 관한 조례 일부개정조례안", Oct 2020.
- [4] "전동키토드 공유서비스 안전실태조사", Aug 2021.
- [5] Ashish. Vaswani, Noam. Shazzer, Niki. Parmar, Jakob. Uszkoreit, Llion. Jones, Aidan N. Gomez, Lukasz. Kaiser and Lllia. Polosukhin, "Attention Is All You Need" arXiv preprint arXiv:1706.03762v5, Dec. 2027.
- [6] 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, "AN IMAGE IS WORTH 16X16 WORDS : TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE" arXiv:2010.11929v2, Jun 2021.
- [7] Enze. Xie, Wenhui. Wang, Zhiding. Yu, Anima. Anandkumar, Jose M. Alvarez, and Ping. Luo, "SegFormer : Simple and Efficient Design for Semantic Segmentation with Transformers" arXiv:2015.15203v3, Oct 2021.
- [8] Ze. Liu, Yutong. Lin, Yue. Cao, Han. Hu, Yixuan. Wei, Zheng. Zhang, Stephen. Lin, and Baining. Gue, "Swin Transformer : Hierarchical Vision Transformer using Shifted Windows", arXiv preprint arXiv:2103.14030v2, Aug 2021.
- [9] Rudra. PK. Poudel, Stephan. Liwicki, and Roberto. Cipolla, "Fast-SCNN:Fast Semantic Segmentation Network", arXiv:1902.04502v1, Feb 2019.
- [10] Chao, Ping, Chao-Yang Kao, Yu-Shan Rua n, Chien-Hsiang Huang, and Youn-Long Lin. "Hardnet:A low memory traffic network" In Proceedings of the IEEE International Conference on Computer Vision, pp. 3552-3561. Sep 2019.
- [11] Fan, Mingyuan, Shenqi Lai, Junshi Huang, Xiaoming Wei, Zhenhua Chai, Junfeng Luo, and Xiaolin Wei, "Rethinking BiSeNet For Real-time Semantic Segmentation." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9816-9725. Apr 2021.
- [12] Ronneberger O, Fischer P, Brox T, "U-net: Convolutional networks for biomedical image

segmentation” International Conference on Medical image computing and computer-assisted intervention, Springer, Cham, 2015:234-241. May 2015.

[저자 소개]



조 용 현 (Yong-hyeon Jo)
2021년 2월 동신대학교 기계공학과 학사 졸업
2023년 2월 동신대학교 에너지기계시스템공학과 석사 졸업
2023년 3월 ~ 현재 부산대학교 정보컴퓨터공학부 석박사통합과정

email : av1483110@gmail.com



최 진 영 (Jin Young Choi)
1996년 2월 한국과학기술원 기계공학과 학사 졸업
1998년 2월 한국과학기술원 기계공학과 석사 졸업
2007년 2월 한국과학기술원 기계공학과 박사 졸업

email : jiny22choi@dsu.ac.kr