

SegFormer 및 U-Net의 철도 구성요소 객체 분할 성능 비교

이재현⁰, 박창준*, 김남중**, 박준휘***, 곽정환(교신저자)**

⁰국립한국교통대학교 컴퓨터공학과,

*국립한국교통대학교 교통·에너지융합학과,

**국립한국교통대학교 소프트웨어학과,

***국립한국교통대학교 AI·로봇공학과

e-mail: dlwogus1441@gmail.com⁰, jgwak@ut.ac.kr**

The Comparison of Segmentation Performance between SegFormer and U-Net on Railway Components

Jaehyun Lee⁰, Changjoon Park*, Namjung Kim**, Junhwi Park***, Jeonghwan Gwak(Corresponding Author)**

⁰Dept. of Computer Engineering, Korea National University of Transportation,

*Dept. of IT·Energy Convergence, Korea National University of Transportation,

**Dept. of Software, Korea National University of Transportation,

***Dept. of AI·Robotics Engineering, Korea National University of Transportation

● 요약 ●

본 논문에서는 철도 구성요소 모니터링을 위한 효율적인 객체 분할 기법으로 사전학습된 SegFormer 모델의 적용을 제안하고, 객체 분할을 위해 보편적으로 사용되는 U-Net 모델과의 성능 비교 분석을 진행하였다. 철도의 주요 구성요소인 선로, 침목, 고정 장치, 배경을 분할할 수 있도록 라벨링된 데이터셋을 학습에 사용하였다. SegFormer 모델이 대조군인 U-Net보다 성능이 Jaccard Score 기준 5.29% 향상됨에 따라 Vision Transformer 기반의 모델이 기존 CNN 기반 모델의 이미지의 전역적인 문맥을 파악하기 상대적으로 어렵다는 한계를 극복하고, 철도 구성요소 객체 분할에 더욱 효율적인 모델임을 확인한다.

키워드: 객체 분할(Segmentation), 철도 시스템(Railway System), Vision Transformer, SegFormer, U-Net

I. Introduction

철도 시스템에서 열차가 주행하기 위해서는 지속적으로 관리가 된 선로, 선로를 고정하는 여러 가지의 구성요소의 온전한 유지보수가 필요하다. 철도 시스템의 구성요소로는 선로(Railway), 침목(Sleepers), 고정 장치(Fastener)로 구성되며 다양한 요소에서 결함이 발생 되면 대형 사고로 이어질 수 있으므로, 지속적인 추적관리가 필요하다. 철도 시스템에 대한 결함은 열차의 운행, 구성요소의 노후화와 같은 물리적인 요인과 날씨와 같은 외부적인 요인이 추가 된다. 이러한 요인에서 발생하는 구성요소의 결함을 사전에 방지하기 위해 지속적인 유지보수 및 모니터링을 진행함으로써 사고를 사전에 방지하는 것이 가능하다. 하지만, 사람이 막대한 길이의 선로와 선로의 구성요소를 전부 확인하는 것은 물리적인 노동력 및 시간적인 비용이 발생하며, 관리자가 위험에 노출될 수 있다는 한계점이 존재한다. 이러한 점을 개선하기 위해 최근 Convolutional Neural Network(CNN) 기반 객체 분할 기법을 사용하여 철도의 구성요소의 결함을 탐지하는 연구가 진행되고 있다[1]. 하지만, CNN 기반 객체

분할 방법은 선로와 주변 구성요소, 자갈도상 등이 나타나는 전역적인 이미지에서 관련된 핵심 특징을 학습해야 하며, 이에 따라 딥러닝 네트워크를 더욱 깊게 설계해야 한다. 따라서, 국소적인 특징만 추출하는 것은 이미지의 전역적인 문맥을 파악하기 상대적으로 어렵다는 한계를 가지며 더 나아가 기울기 소실과 같은 현상이 발생할 수 있다. 반면, Self-Attention 기법을 통한 이미지 전체 문맥 파악이 가능한 Vision Transformer(ViT)[2] 기반 SegFormer[3]을 CNN 기반 U-Net[4]의 철도 구성요소 객체 분할 성능 비교를 통해 SegFormer가 철도 구성요소 객체 분할에 더 적합한 모델임을 확인한다.

II. Proposed method

Fig. 1은 SegFormer 모델의 아키텍처이다. SegFormer는 계층적인 Encoder 구조로 설계되어 있기에 다양한 해상도에서 Self-Attention을 수행하여 기존 CNN으로 구성된 Encoder보다 더

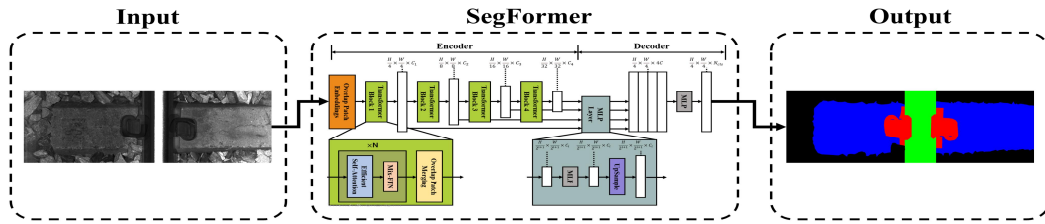


Fig. 1. SegFormer Architecture

큰 수용영역을 얻는다. 이는 CNN이 큰 수용영역을 갖기 위해 모델을 깊게 쌓아 발생하는 기울기 소실 문제, 파라미터 증가에 따른 연산량 증가, 이미지 내 전역적인 문맥을 파악하기 어렵다는 한계들을 보완한다. 또한, SegFormer는 Lightweight ALL-MLP Decoder를 통해 Encoder에서 추출된 다양한 해상도의 특징을 Upsample한 후 Multi-Layer Perceptron(MLP)로 프로젝션하여 다양한 해상도의 특징 채널을 통일시킨다. 따라서, SegFormer는 이미지의 세분화된 특징 유지 및 더 많은 공간적 특징을 학습할 수 있다.

III. Experiment

1. Dataset

Table 1. Configuration of Dataset

train (0.7)	valid (0.2)	test (0.1)	total (1.0)
949	272	136	1,357

실험에 사용된 데이터셋은 철도 기관의 선로 점검차를 통해 수집된 데이터이며, 구성은 Table 1과 같다. 데이터는 총 1,357장으로, 각 데이터를 Train(70%), Validation(20%), Test(10%)로 나누어 학습용 데이터셋을 구성하였다. 또한, 데이터셋의 이미지 크기는 Width는 2048로 동일한 크기를 가지지만, 각각 다른 크기의 Height를 가지기에 이미지 크기는 1024×1024 크기로 재조정하였다.

2. Results

Table 2. Experiment results

model	Mean IoU Score	F1 Score	Jaccard Score
SegFormer	0.9827	0.9827	0.9664
U-Net	0.9539	0.9539	0.9135

Table 2는 ImageNet을 사전학습한 기중치를 통해, 철도 데이터셋의 학습 결과를 나열한 표이다. 하이퍼파라미터의 경우 Epoch은 200, Batch Size는 1로 설정하였고 Early stop은 Mean Intersection over Union + Jaccard Score 기준 patience 6으로 설정하였다. 또한, Optimizer는 SegFormer의 경우 AdamW를, U-Net은 Adam을 사용하였다. Table 2에서 확인할 수 있듯이, CNN기반 모델인 U-Net보다 ViT기반 모델인 SegFormer의 Jaccard Score가 5.29% 더 높음을 확인할 수 있다.

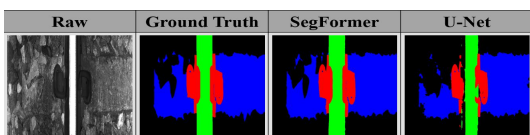


Fig. 2. SegFormer / U-Net Predict Result

Fig. 2는 SegFormer와 U-Net의 테스트 데이터셋에 대한 객체 분할 예측 결과이다. Fig. 2에서 확인할 수 있는 것처럼 SegFormer의 예측 결과를 보면 Ground Truth(GT)와 거의 동일한 결과를 보인다. 반면, U-Net의 예측 결과를 보면 초록색 부분의 레일 및 선로 고정 장치에 대한 객체 분할 결과는 GT와 상이한 모습을 확인할 수 있다. 이를 통해 SegFormer가 철도 구성요소 객체 분할에서 U-Net보다 더 강력한 성능을 보이는 것을 확인하였다.

IV. Conclusions

본 논문에서는 철도 구성요소 객체 분할에 대하여 CNN 기반 U-Net과 ViT 기반 SegFormer의 성능을 비교 분석하였고, SegFormer가 더 효율적인 객체 분할 기법임을 확인했다. 추후 연구에서는 SegFormer에 Knowledge Distillation을 통한 경량화 기법 적용 및 Lightweight ALL-MLP Decoder 고도화를 통해 철도 이미지 내 페트병, 못, 나무 등과 같은 이물질이 포함되어있는 이미지에도 강력한 객체 분할 성능을 유지하는 기법에 대해 꾸준히 연구를 진행할 예정이다.

ACKNOWLEDGEMENT

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2014-3-00077).

REFERENCES

- [1] X. Gibert, V. M. Patel and R. Chellappa, "Material Classification and Semantic Segmentation of Railway Track Images with Deep Convolutional Neural Networks," *IEEE International Conference on Image Processing (ICIP)*, Sep. 2015.
- [2] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *In International Conference on Learning Representations*, May 2021.
- [3] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez and P. Luo, "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," *Conference on Neural Information Processing Systems (NeurIPS)*, Dec. 2021.
- [4] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Oct. 2015.