

도로환경에서 차량 검출을 위한 U-Net 기반 모델 성능 비교

조국한, 송영준

금오공과대학교 전자공학과

rnrgks12@kumoh.ac.kr, yjsong@kumoh.ac.kr

Performance comparison of U-Net-based Models for Vehicle Detection in Road Environment

Guk Han Jo, Young Joon Song

Dept. of Electronic Engineering, Kumoh National Institute of Technology.

요약

본 논문에서는 객체 검출에 많이 사용되는 U-Net을 기반으로 2가지의 모델을 구성하여 차량을 검출하고 성능을 비교하였다. 각각의 모델은 double convolution로 구성된 U-Net 모델, single convolution과 추가적인 depth를 가지는 U-Net 모델이다. 데이터셋은 6500개의 학습 데이터와 500개의 테스트 데이터를 128×128 로 resize하여 사용하였다. 학습 완료 후 mIoU(Mean Intersection over Union)를 이용하여 성능을 비교한 결과 각 모델의 성능 차이는 1%~2%로 근소하다. 하지만 single convolution과 추가적인 depth를 가지는 U-Net 모델의 소요 시간이 epoch당 3초정도 더 빠른 것을 확인할 수 있었다.

I. 서론

최근 머신러닝의 발달로 인해 차량에서 활용될 수 있는 다양한 기술들이 개발되고 있다. 그 중에서도 객체를 인식하는 기술은 자율주행, 사고 예측에 필수적으로 활용되고 있기에 다양한 알고리즘들이 설계 및 개선되고 있다. 예를 들어 YOLO(You Only Look Once)[1], DOTA(Object Detection in Aerial Images)[2], FCN[3], U-Net[4] 등이 대표적으로 사용되고 있으며 이에 맞게 다양한 교통 환경에 대한 데이터셋도 제공되고 있다. 본 논문에서는 BDD100K[5] 데이터셋과 U-Net을 기반으로 2가지 모델을 구성하여 차량을 검출하고 각 모델의 성능을 비교하고자 한다.

II. U-Net 모델 학습과정과 결과

본 논문에서는 BDD100K 데이터셋을 활용하여 차량을 검출한다. BDD100K는 colormap, bitmap, json의 형태로 instance segmentation 검출에 적합한 학습데이터를 제공한다. 본 논문에서 사용된 학습이미지는 6500개, 테스트이미지는 500개로 이를 128×128 로 resize하여 사용한다[6]. 모델은 서로 다른 expanding path, contracting path를 가지는 모델 2개를 구성한다. 첫 번째 모델은 double convolution을 사용한 기존 U-Net의 방법이고 두 번째 모델은 single convolution에 depth를 추가한 방법을 사용한다.

그림 1은 기본 U-Net으로 각 path에서 double convolution을 가지는 모델을 도식화 한 것이다. 그림 1에서 숫자는 필터의 개수를 나타낸다. 학습 횟수는 총 100회를 수행하였으며 batch size는 50으로 설정하였다. Contracting path에서 반복적으로 보이는 convolution, activation function, max pooling을 encoder block이라고 한다면 expanding path에서는 decoder block이라한다[7]. 한 개의 convolution을 가지는 U-Net 모델에서는 각 path에 block을 1회씩 추가하여 학습을 진행한다. 결과적으로 single convolution 기반의 U-Net 모델에서는 depth가 늘어난 형태를 가진다.

표 1은 U-Net을 기반으로 구성한 두 모델의 학습 결과이다. 성능 평가는 mIoU[8]를 이용하였다. 학습 결과 double convolution인 U-Net 모델과 single convolution 모델의 mIoU 차이는 1~2%로 성능에서 큰 차이를 보이지는 않는다. 하지만 epoch 당 소요시간이 3초가량 차이이기 때문에 총 학습시간에서 약 200~300초 정도 차이가 발생하는 것을 알 수 있다. 하지만 추가적인 학습데이터 입력 및 U-Net 모델구조 변경에 따라 성능이 변할 수도 있다.

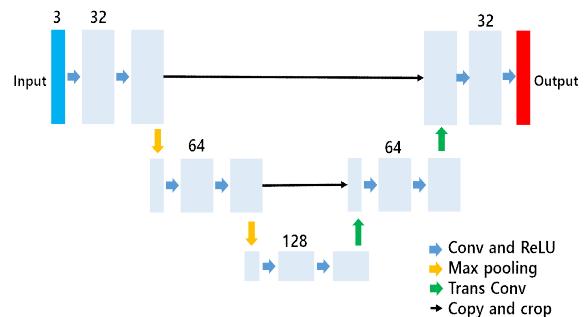


그림 1. U-Net model with double convolution

표 1. Training results of U-Net models

Model	Epoch당 소요시간 (초)	mIoU (%)
Double conv	13~14	54.5
Single conv	10~11	55.2

III. 결론

본 논문에서는 BDD100K 데이터셋과 U-Net 기반 모델을 사용하여 도로환경에서 차량을 검출하고 구성한 모델을 비교하는 연구를 수행하였다.

그 결과 mIoU 차이는 1~2%로 거의 차이가 없었지만 single convolution을 기반으로 구성한 모델의 학습 소요시간이 더 짧은 것을 확인할 수 있었다. 따라서 짧은 학습시간을 요구하는 시스템에서는 single convolution 모델을 효과적으로 활용할 수 있을 것이며 학습 정확도를 향상시키기 위해서는 추후 학습 데이터를 추가하거나 모델의 구조를 개선하는 등 지속적인 연구가 필요하다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2021R1A6A3A13043168)

참 고 문 헌

- [1] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934, April, 2020.
- [2] Xia, Gui-Song, et al. "DOTA: A large-scale dataset for object detection in aerial images." Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City, UT, USA, pp.3974-3893, Nov., 2018.
- [3] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition, pp.3431-3440, UC Berkeley, USA, Nov., 2015.
<https://doi.org/10.1109/CVPR.2015.7298965>
- [4] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, pp.234-241, May, 2015.
- [5] <https://doc.bdd100k.com/download.html>. [accessed : October, 3, 2022]
- [6] Guk-Han Jo, Young-Joon Song "Image Data Generation System for Classification of Traffic Accident Scenarios," The Journal of Korean Institute of Information Technology, Vol.20, No.4, pp.31-38, 2022.
- [7] <https://idiotdeveloper.com/what-is-unet/> [accessed : October, 3, 2022]
- [8] Chen, Liang-Chieh, et al. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." IEEE transactions on pattern analysis and machine intelligence, 40.4, pp.834-848, Apr., 2017.