연구논문

Journal of Drive and Control, Vol.19 No.4 pp.104-109 Dec. 2022
ISSN 2671-7972(print) ISSN 2671-7980(online)
http://dx.doi.org/10.7839/ksfc.2022.19.4.104

승용자율주행을 위한 의미론적 분할 데이터셋 유효성 검증
Validation of Semantic Segmentation Dataset for Autonomous Driving

곽석우[1] · 나호용[1] · 김경수[1] · 송은지[1] · 정세영[1] · 이계원[2] · 정지현[2] · 황성호[1*]
Seoku Gwak[1], Hoyong Na[1], Kyeong Su Kim[1], EunJi Song[1], Seyoung Jeong[1]
Kyewon Lee[1], Jihyun Jeong[1] and Sung-Ho Hwang[1*]

Received: 07 Nov. 2022, Revised: 29 Nov. 2022, Accepted: 29 Nov. 2022

Key Words : Semantic Segmentation(의미론적 분할), Deep learning(딥러닝), Autonomous Driving(자율주행), Dataset(데이터셋), Object Detection(객체 인식)

Abstract: For autonomous driving research using AI, datasets collected from road environments play an important role. In other countries, various datasets such as CityScapes, A2D2, and BDD have already been released, but datasets suitable for the domestic road environment still need to be provided. This paper analyzed and verified the dataset reflecting the Korean driving environment. In order to verify the training dataset, the class imbalance was confirmed by comparing the number of pixels and instances of the dataset. A similar A2D2 dataset was trained with the same deep learning model, ConvNeXt, to compare and verify the constructed dataset. IoU was compared for the same class between two datasets with ConvNeXt and mIoU was compared. In this paper, it was confirmed that the collected dataset reflecting the driving environment of Korea is suitable for learning.

1. 서 론

의미론적 분할은 객체를 이미지의 픽셀단위로 분류하는 알고리즘으로써, 객체 감지와 더불어 자율주행 인지의 핵심이 되는 알고리즘이다. 많은 연구들에서 의미론적 분할 알고리즘의 실시간성 및 정확도 성능을 향상시키기 위해 딥 인공신경망 모델 구조를 개선하고 있고,[1)] LiDAR 센서와의 센서퓨전을 통해 성능을 개선하고자 하는 연구를 진행하고 있다.[2)]

이러한 노력에도 불구하고, 의미론적 분할 성능은 객체 감지 성능 대비해서 양산에 활용할 수 있을 만큼의 성능이 확보되지 못하고 있는 실정이다. 이는 실시간성과 모델 정확성에 대한 요구사항을 만족시켜가면서 성능을 끌어 올리는 것에는 현실적인 한계가 있을 수 있다는 것을 시사한다. 컨벌루션 신경망 기반 의미론적 분할은 야간, 악천후 상황의 왜곡된 영상이 입력될 경우 큰 성능 저하를 보인다.[3)] 또한, 모델 개발에 활용하고 있는 데이터셋들은 대부분 해외에서 구축된 데이터 셋이고[4)] 데이터 규모가 제한적이므로 부족한 데이터를 추가하는 방식이 이루어지고 있지만[5)], 이러한 데이터셋들을 활용한 의미론적 분할 알고리즘은 국내 도로환경에서의 자율주행 로직에 활용하기에 현실적으로 어려운 측면이 있다.

한편, 의미론적 분할을 위한 데이터를 대규모로 확보하기 위해서는 방대한 라벨링 작업이 요구되는데, 이는 비용과 시간이 많이 소요되는 작업이다. 사람이 수작업으로 라벨링을 하기 때문에 사람마다 다른 오차가 발생될 수 있고, 이는 모델 성능에 치명적인 영향을 주게 된다. 이러한 휴먼 에러를 원천적으로 제거하기 위해 가상 시뮬레이션 환경을 활용하여 데이

* Corresponding author: hsh0818@skku.edu
1 Department of Mechanical Engineering, Sungkyunkwan University, Suwon 16419, Korea
2 INFINIQ, 186, Gasan digital 1-ro, Geumcheon-gu, Seoul, Korea

Copyright Ⓒ 2022, KSFC
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.

터셋을 구축하는 경우도 있지만,[6),7)] 데이터들이 실제 환경과 차이가 존재하기 때문에 현실에서 의미론적 분할을 할 경우 성능이 낮게 나오는 경향이 있다.[8)] 가상환경의 한계를 극복하기 위해 현실에서의 데이터셋을 융합해서 학습하는 연구,[8)] Domain adaptation을 활용한 연구들이 있으며,[9)] 데이터 규모의 한계를 극복하기 위해 다중 데이터셋 학습에 관한 연구와 자동 라벨링에 대한 연구들이 존재한다.[10)]

본 연구에서는 위 언급한 한계들을 근본적으로 극복하기 위해 '과학기술정보통신부'가 주관하고 '한국지능정보사회진흥원'이 추진하는 '2022년 인공지능 학습용 데이터 구축사업(The Open AI Dataset Project, AI-Hub, S. Korea)'에서 국내 도로환경에서 의미론적 분할을 위한 대규모 데이터셋을 구축하였고 주/야간 도심 주행상황에 대하여 주간 도심 데이터 21,520장 야간 도심 데이터 17,920장을 검증하였다. 구축한 데이터는 MMSegmentation platform의 ConvNeXt을 활용하여 성능을 검증하였다.[11)] 검증지표는 IoU(Intersection over Union)를 사용하였다.

2. Dataset 비교 및 분석

의미론적 분할 연구를 위해서 CityScapes[12)], IDD(India Driving Dataset), A2D2(Audi Autonomous Driving Dataset)[13)], Mapillary, BDD(Berdeley DeepDrive), nuScene, NightCity 등의 많은 데이터들이 개방되어 있다. CityScapes, IDD, A2D2, Mapillary, nuScene 등의 데이터는 주간 주행에 대한 데이터이며, NightCity는 야간주행에 대한 데이터이다.

Table 1은 각 데이터셋에 대한 클래스 개수와 데이터 규모를 보여준다. 차량의 실외 주행상황에 대한 의미론적 분할 알고리즘의 성능 분석에는 CityScapes 혹은 CamVid 데이터셋이 많이 사용된다. Table 1을

Fig. 1 Class comparison between our dataset and CityScapes dataset

보면 A2D2 데이터셋은 우리가 검증한 데이터셋과 비교했을 때, 다양한 환경에서 데이터가 취득되었다는 점과 클래스 수가 유사한 모습을 보인다. 따라서, 이번 연구를 통해 구축한 데이터셋은 클래스에 대해 분석하였고 CityScapes 데이터셋 및 A2D2 데이터셋들의 클래스에 대해 비교하였다.

데이터들을 대분류, 중분류, 소분류로 분류하였다. 대분류는 이동객체, 고정객체 수준으로의 분류를 의미하며, 중분류는 자동차, 도로, 차선 수준의 분류를 의미한다. 그리고 소분류는 차량이 버스인지, 스쿨버스인지, 경찰차 등의 상세한 분류를 나타낸다.

구축한 데이터셋과 CityScapes는 12개의 클래스가 중복되는 것을 볼 수 있다. 한편, 구축한 데이터에는 roadmark, blue-lane, red-lane 등의 국내 도로 주행환경

Table 1 Dataset analysis

Dataset	# Sequences	# Images	# Classes	Multi Cities	Multi Weathers	Multi Times of Day	Multi Scene types
KITTI	22	14,999	11	No	No	No	Yes
Cityscapes	~50	5000 (+2000)	30	Yes	No	No	No
ApolloScape	4	143,906	28	No	No	No	No
Mapillary	N/A	25K	124	Yes	Yes	Yes	Yes
BDD100K	100,000	120M	19	Yes	Yes	Yes	Yes
A2D2	N/A	41,277	38	Yes	Yes	No	Yes
Ours	538	36,600	33	Yes	Yes	Yes	Yes

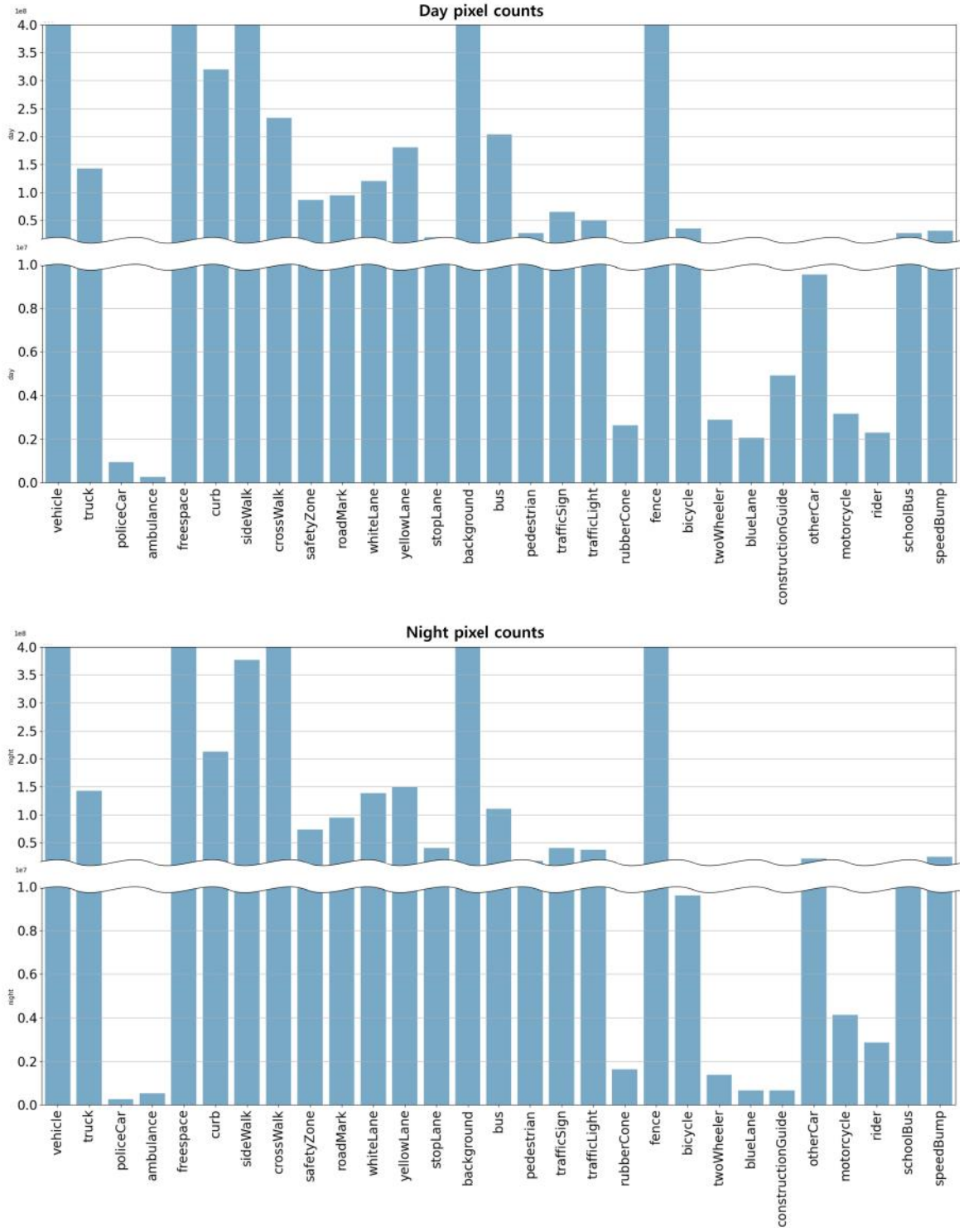

Fig. 2 Day/Night pixel counts

에 특화된 클래스들이 존재한다. red-lane의 경우 전체 데이터셋에서 픽셀 수가 주간 도심의 경우 0개, 야간 도심의 경우는 전체 이미지중 1개의 이미지, 총 495개의 픽셀, waringTriangle의 경우, 주간 도심 3개 이미지, 총 1715개 픽셀, 야간 도심 1개 이미지, 총 3873개 픽셀로 전체 데이터 대비 매우 적은 수치로 클래스를 제거하였다. 또한 야간 도심의 경우 egoVehicle이 라벨링 되어 있으나 주간 도심의 경우는 라벨링이 되어 있지 않아 egoVehicle 클래스 역시 제거하였다.

구축된 데이터셋과 유사한 데이터셋으로 A2D2 데이터셋이 존재한다. A2D2 데이터셋의 경우 19개의

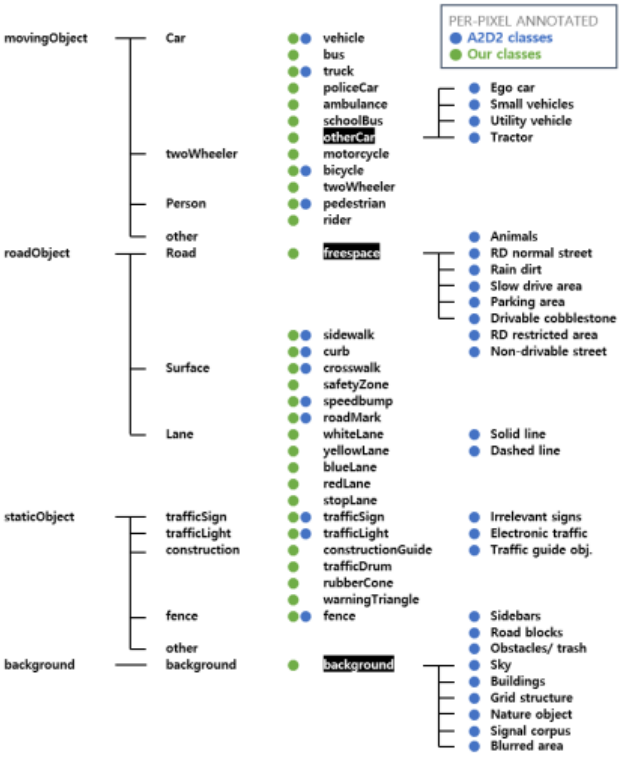

Fig. 3 Class comparison between our dataset and A2D2 dataset

Cityscapes와 유사한 클래스가 있으며 총 38개의 클래스를 갖는다. 따라서 이를 유효성 검증을 진행할 데이터셋의 비교 데이터셋으로 활용하였다. A2D2 데이터셋에서 인접한 동일 클래스의 경우 라벨링을 다르게 하여 구분하였다. 비교군 데이터셋으로 동일한 학습을 진행하기 위해 구분되어 있는 인접한 클래스는 동일한 클래스로 처리하였다.[10)]

3. 검증방법

구축된 데이터셋 검증을 위해 의미론적 분할 학습을 진행하였다. 학습을 진행하는 데 사용한 플랫폼은 MMSegmentation으로, 다양한 의미론적 분할 backbone 네트워크를 활용해 학습하는 기능을 제공한다. 그 중 정확도에 초점을 둔 ConvNeXt 모델을 활용하여 데이터셋의 성능을 검증하였다. 또한 동일 모델을 기존에 의미론적 분할 학습을 위해 주로 사용하는 오픈 데이터셋인 A2D2으로 학습한 결과를 구축한 데이터셋과 비교하였다. A2D2 데이터셋을 학습하기 위해 기존 A2D2의 라벨링 데이터 전처리를 진행하였다. 구축된 데이터셋의 경우 Sky와 Building을 모두 배경으로 지정하였다. 해당 클래스들은 전체 이미지에서 차지하는 영역이 크며 mIoU를 향상시키기 때문에 비교를 위해 학습에서는 클래스를 제외하였다.

Fig. 4 (a) Raw image, (b) Original labeling of A2D2 dataset, (c) Preprocessed labeling of A2D2 dataset

의미론적 분할 학습 성능은 mIoU로 기준을 두었다. 여기서, mIoU는 각 클래스 별로 IoU를 평균낸 값이며 IoU는 Ground Truth(GT) 이미지와 모델 출력 이미지에서 각 클래스로 구분된 영역 간 겹치는 정도를 의미한다.

데이터셋의 학습 성능 검증을 위한 모델로 ConvNeXt을 선정한 이유는 다음과 같다. 정확도 측면에서 State-of-the-Art에 등록된 Cityscapes 데이터셋 기준 의미론적 분할 학습 성능은 HRNetV2+ OCR 네트워크가 mIoU 86.93%으로 가장 높은 것으로 확인하였다. 하지만 Mapillary와 같이 클래스가 세부적으로 분류된 데이터셋을 학습했을 경우 성능이 mIoU 56% 이하로 낮게 나오는 것으로 확인하였다. 우리의 데이터셋은 Fig. 1에서 비교했던 바와 같이 Cityscapes에서 분류된 클래스보다 세분화되어 있기 때문에 성능이 최신 자료에 기재된 것만큼 나오지 않을 것이라 판단하였다. 이에 세분화된 클래스에서 강점을 보이는 ConvNeXt를 검증용 네트워크로 선정하였다.

딥러닝 방식은 같은 네트워크 모델을 사용하더라도 입력 데이터셋의 특징에 매우 큰 영향을 받는다. 따라서 기존 데이터셋인 A2D2을 학습했을 때와 비교하여 우리의 데이터셋을 학습한 결과가 비슷한 mIoU 성능을 보이는지 검증하였다. 여기서 과적합으로 인해 mIoU 값이 높게 나올 수 있기 때문에, 전체 데이터에서 80%는 training set, 10%는 validation set, 나머지 10%는 test set으로 나누어 진행하였다. 학습은 두 데이터셋 모두 160,000 iteration 학습을 진행하였다.

4. 검증결과

본 연구를 통해 구축한 데이터셋과 A2D2 데이터셋에 대해 ConvNeXt 모델을 활용해 IoU 성능은 Table 2, 3과 같다.

Table 2 ConvNeXt methods IoU result on A2D2 dataset

Class	IoU
Car	94.31
Bicycle	71.96
Pedestrain	70.51
Truck	96.75
Small vehicles	72.08
Traffic signal	85.80
Traffic sign	86.59
Utility vehicle	71.68
Sidebars	74.06
Speed bumper	Nan
Curbstone	73.50
Solid line	81.51
Irrelevant signs	67.50
Road blocks	84.89
Tractor	23.57
Non-drivable street	77.60
Zebra crossing	77.32
Obstacles/ trash	59.22
Poles	63.74
RD restricted area	84.54
Animals	42.29
Grid structure	79.75
Signal corpus	64.40
Drivable cobblestone	74.10
Electronic traffic	67.02
Slow drive area	33.20
Nature object	96.33
Parking area	58.89
Sidewalk	83.01
Painted drive. instr.	74.67
Traffic guide obj.	83.89
Dashed line	74.60
RD normal street	97.21
Buildings	91.85
Blurred area	0.00
Ego car	Nan
Sky	Nan
Rain dirt	Nan
mIoU	71.72

Table 3 ConvNeXt methods IoU result on our dataset

Class	IoU
vehicle	96.07
bus	96.95
truck	93.74
policeCar	79.36
ambulance	64.15
schoolBus	93.76
otherCar	77.69
motorcycle	73.70
bicycle	80.16
twoWheeler	57.22
pedestrian	76.23
rider	53.41
freespace	96.43
curb	85.17
sidewalk	87.77
crossWalk	85.95
safetyZone	90.10
speedBump	88.54
roadMark	75.85
whiteLane	77.83
yellowLane	78.74
blueLane	70.67
redLane	Nan
stopLane	72.76
trafficSign	87.38
trafficLight	88.75
constructionGuide	56.39
trafficDrum	52.50
rubberCone	74.42
warningTriangle	Nan
fence	96.07
egoVehicle	Nan
background	99.33
mIoU	79.58

구축한 데이터셋의 배경을 제외한 mIoU는 79.58%, A2D2 데이터셋의 Sky를 제외한 mIoU는 71.72%이다. A2D2의 클래스 개수가 많으며 세세하게 구분되어 있어 전체적인 IoU 성능을 저하 시킬 수 있다. 특히 A2D2 데이터셋은 Tractor, Slow drive area, Blurred area와 같이 취득하기 힘든 클래스를 평균에서 제외한 mIoU는 76.82%로 구축한 데이터셋의 mIoU와 상당히 유사하다. 의미론적 분할 알고리즘으로 검증된 모델인 ConvNeXt를 통해 두 데이터셋을 비교한 결과, 인식률이 차이가 7.86%로 비슷하다.

A2D2 데이터셋과 구축된 데이터셋의 공통적인 클래스의 IoU 결과는 Table 4와 Fig. 5과 같다. A2D2 데이터셋의 speedbump의 경우, 데이터 수가 충분하지

Table 4 ConvNeXt methods IoU result on common classes of A2D2 and our dataset

Class	A2D2	Our Dataset
vehicle	94.31	96.07
truck	96.75	93.74
otherCar	55.78	77.69
bicycle	71.96	80.16
pedestrian	70.51	76.23
freespace	65.85	96.43
sidewalk	83.01	87.77
curb	73.50	85.17
crosswalk	77.32	85.95
speedbump	Nan	88.54
roadMark	74.67	75.85
trafficSign	86.59	87.38
trafficLight	85.80	88.75
fence	63.74	96.07
mIoU	76.91	86.84

Fig. 5 ConvNeXt methods IoU result on common classes of A2D2 and our dataset

않아 IoU 계산에서 배제되었다. A2D2 데이터셋의 otherCar의 경우, Small vehicles, Utiliy vehicle, Tractor를 평균을 내어 구하였고 주행가능영역(freespace)의 경우, RD normal street, Rain dirt, Slow drive area, Parking area, Drivable cobblestone를 평균을 내어 구하였다. A2D2의 otherCar와 주행가능영역은 구축된 데이터셋에 비해 세세하게 구분되어 IoU가 비교적 낮은 경향을 보인다.

5. 결론

본 논문에서는 개방된 기존 데이터셋 분석을 진행하였다. 기존 데이터셋은 대부분 해외 도로 환경을 반영한 데이터셋으로 한국의 도로환경과는 다른 특성을 가지고 있는 부분이 있다. 인공지능 학습용 데이터 구축 사업에서 국내 도로환경을 반영한 의미론적 분할을 위한 대규모 데이터셋을 구축하였고 이를 검증하였다. 해당 데이터셋을 검증하기 위해 이와 유

사한 데이터셋인 A2D2 데이터셋을 비교하였다. ConvNeXt 딥러닝 모델을 활용하여 학습을 진행하였다. A2D2 데이터셋의 ConvNeXt 결과대비 구축된 데이터셋이 딥러닝 모델 학습용으로 유효함을 확인하였다.

추후 연구에서는 구축된 데이터셋의 특정 클래스가 상대적으로 부족하여 생기는 인식률 저하 문제를 해결하고자 데이터 증강기법을 사용할 예정이다. 특히 의미론적 분할 모델을 활용하여 구축된 데이터셋의 활용가능성을 검증했지만 각각의 데이터셋이 보유하는 특정 클래스 종류가 달라 정확한 분석이 어려웠다. 따라서 소분류 되어 있는 클래스를 중분류, 대분류로 묶어 검증된 딥러닝 모델을 활용해 다양한 데이터셋과 비교 분석할 계획이다.

후 기

본 연구는 국토교통부/국토교통과학기술진흥원 교통물류연구사업의 연구비지원 (22TLRP-C152478-04)과 과학기술정보통신부 및 한국지능정보사회진흥원 인공지능 학습용 데이터 구축 사업의 연구결과로 수행된 결과물입니다.

이해관계(CONFLICT OF INTEREST)

저자는 이 논문과 관련하여 이해관계 충돌의 여지가 없음을 명시합니다.

References

1) Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning, p.6105-6114. PMLR.
2) Elmenreich, W. (2002). An introduction to sensor fusion. Vienna University of Technology, Austria, 502, p.1-28
3) Song, T., Jang, H., Ha, N., Yeon, Y., Kwon, K., and Sohn, K. (2019). Deep Multi-task Network for Simultaneous Hazy Image Semantic Segmentation and Dehazing. Journal of Korea Multimedia Society, 22(9), 1000-1010.
4) Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., and Terzopoulos, D. (2022). Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(7), p.3523-3542
5) H. S. Baek, J. H. Shin and S. J. Kim, "Development of AI-Based Condition Monitoring System for Failure Diagnosis of Excavator's Travel Device", Journal of Drive and Control, Vol.18, No.1, pp.24-30, 2021.
6) Lee, J., Gwak, G., Kim, K., Kang, W., Shin, D., and Hwang, S. (2021). Development of virtual simulator and database for deep learning-based object detection. Journal of Drive and Control, 18(4), p.9-18. 2022.
7) Kim, K., Lee, J., Gwak, S., Kang, W., Shin, D., and Hwang, S. (2022). Construction of database for deep learning-based occlusion area detection in the virtual environment. Journal of Drive and Control, 19(3), p.9-15. 2022.
8) Weng, X., Man, Y., Cheng, D., Park, J., O'toole, M., and Kitani, K. (2020). All-in-one drive: A large-scale comprehensive perception dataset with high-density long-range point clouds.
9) Toldo, M., Maracani, A., Michieli, U., and Zanuttigh, P. (2020). Unsupervised domain adaptation in semantic segmentation: a review. Technologies, 8(2), p.35.
10) Chen, Y., Liu, L., Tao, J., Chen, X., Xia, R., Zhang, Q., Xiong, J., Yang, K., and Xie, J. (2021). The image annotation algorithm using convolutional features from intermediate layer of deep learning. Multimedia Tools and Applications, 80(3), p.4237-4261.
11) Contributors, M. (2020). MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark. https://github.com/open-mmlab/mmsegmentation.
12) Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE conference on computer vision and pattern recognition, p.3213-3223.
13) Geyer, J., Kassahun, Y., Mahmudi, M., Ricou, X., Durgesh, R., Chung, A., Hauswald, L., Pham, H., et al. (2020). A2d2: Audi autonomous driving dataset. arXiv preprint arXiv:2004.06320.