보행자 안내 인공지능 모델훈련을 위한 원근변화 데이터 증강

Perspective Transfom Augmentation for Pedestrian Semantic Segmentation

○정 창 현[1], 조 강 현[2*]

[1] 울산대학교 전기전자컴퓨터공학과 (TEL: 052-259-1664, E-mail:chjeong@islab.ulsan.ac.kr)
[2] 울산대학교 전기전자컴퓨터공학과 (TEL: 052-259-1664, E-mail:acejo@ulsan.ac.kr)

Abstract Semantic segmentation is useful because it can be used on many applications. However, to train a semantic segmentation model, huge dataset is necessary. Some dataset is applicable globally as they do not vary by the locale. But some are influenced greatly by the geological or cultural differences. Even when one finds a dataset for their needs, it can be not exactly what they wanted. And the hardest challenge is making the dataset itself. Compared to other domains of computer vision tasks, making semantic segmentation dataset needs a lot more time and effort. Here, when we applied the CityScapes dataset for road scene parsing, it barely worked on the korean pedestrian scenario. So we applied perspective transform augmentation on various models and shared the result here.

Keywords Data augumetation, Semactic segmentation, Pedestrian scene parsing

1. 서론

AlexNet[1] 이후에 CNN(Convolutional Neural Network)을 활용한 많은 컴퓨터 비젼분야의 발전이 있었다. 그중 영역분할(semantic segmentation) 이미지의 모든 픽셀을 구별하는 작업이다. 영역분할(semantic 은 의학 이미지 처리를 통한 인공지능 진단, 자율주행 자동차, 위성, 항공사진 분석 등 다양한 분야에서 사용됐다. CNN 베이스의 컴퓨터 비전 인공지능 모델을 훈련하기 위해서는 많은 양의 데이터셋이 필요하며, 이를 위해서는 많은 인력과 시간이 필요하다. 특히 영역분할은 이미지 내의 모든 픽셀의 레이블을 작업해야 하기에 다른 컴퓨터 비전 데이터셋 보다 더 많은 인력이 필요하다. 이러한 비용 문제로 인해, CNN 모델을 활용한 애플리케이션을 제작할 경우, 기존에 제작된 데이터셋을 활용하는 것이 개발하는 사람들이 첫 번째로 고려해야 할 점이라 할 수 있다.

2. 본론

치매나 인지능력에 문제가 있는 보행자를 보조하는 증강현실 시스템 제작하기 위해 보행자 위주의 장면 분석 모델을 훈련하고자 하였다. DeepLabV3+[2] 모델을 CityScapes* 데이터셋으로 훈련한 모델을 한국의 보행자 환경에 적용해 보았다. 그림 1은 그 결과를 표시하는데, Cityscapes 가 전반적으로 유럽에서 제작된 데이터셋임을 감안했을 때, 훌륭한 성능을 보여주고 있다. 하지만, 붉은색 실선 내의 영역의 레이블 참값은 side_walk 인데, CityScapes 로 훈련된 모델은 이를 감지하지 못하고 있다. CItyScapes 데이터셋이 차량으로 촬영되어, 이미지의 중심영역에는 주로 차량의 도로가 있고, 인도가 화면의 중심에 오는 경우가 없어 이러한 feature를 감지하지 못하는 것이 아닌가 추측한다.

그림 1. DeepLabV3+모델을 CityScapes 데이터셋으로 훈련시킨 결과를 한국 도보 환경에 적용한 결과.

3. 실험

3.1 데이터셋

한국의 AI HUB 사이트에는 테스트웍스에서 2019년에 제작된, 인도 보행자 데이터셋이 공개되어 있다. 이중 도면 마스킹 섹션에 있는 데이터는 총 52,000 장이지만, 이중 레이블링 데이터가 누락되거나 레이블 형식에 오류가 있는 이미지들을 제외하고 총 46,211

* https://www.cityscapes-dataset.com/

장의 이미지가 사용가능한 이미지 이다.

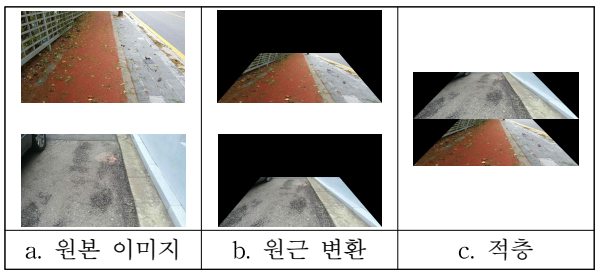
표 1. 노면 객체 클래스 및 속성

대분류	소분류
인도	보도블럭, 시멘트, 우레탄, 아스팔트, 흙_돌_비포장도로, 인도의 파손, 기타
점자블럭	정상, 파손
차도	정상, 횡단보도
이면도로	정상, 과속방지턱, 횡단보도, 이면도로 파손
자전거도로	정상
주의구역	계단, 맨홀, 그레이팅, 보수구역, 가로수영역

6개의 메인 카테고리와 21개의 서브 카테고리로 구성된 레이블에는 문제가 있다. 자전거 도로와 인도_우레탄은 시각적으로 매우 흡사하다. 이는 도로와 이면도로에서도 발생하는 문제이다. 이 문제는 실제 애플리케이션 개발에서는 분명히 고려되어야 할 문제지만, 이번 연구에서는 데이터 증강 문제로만 접근하기 위하여 레이블링 후처리 작업을 진행하지 않고 실험을 진행하였다. 총 42000장의 훈련 이미지와 4000장의 검증 이미지를 사용하였다.

훈련에 사용된 모델은 STDC2[3] 모델로, 애플리케이션 특성상 실시간 영역 구분이 가능한 모델을 사용하였다.

3.2 데이터 증강



a. 원본 이미지 b. 원근 변환 c. 적층

그림 2. 데이터 증강 과정

보행자 보조 시스템을 만들기 위해 필요한 데이터셋은 전방을 주시하는 이미지들인데, 현재 사용 가능한 데이터셋은 시점이 아래를 향하고 있다. 이 문제를 해결하기 위하여 시점 변환 처리를 이미지에 적용 하였다. 적합한한 변환 각도를 구하는 과정에서, 하나의 이미지에만 변환을 적용할 경우 이미지의 많은 공간이 빈 공간이 되는 것을 발견하고 2개의 이미지를 하나의 이미지에 stack 함으로써 보다 효율적인 연산을 적용하였다.

3.3 실험 결과

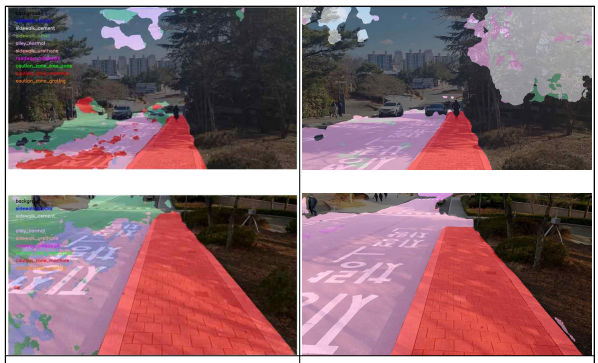
그림 2에서와 같이 이미지를 전처리 한 후 모델을 훈련하고 이에 따른 성능 비교를 하였다. 방법 a는 원본 이미지를 사용하였다. b는 원근변환을 개별이미지에 적용하여 이를 원본 이미지와 함께 훈련하여 총 훈련 이미지 수는 원본의 2배가 되었다. c는 개별이미지에 원근 변환을 적용한 뒤 원근 변환된 이미지를 적층하였다. 이에 c 경우 원본 이미지 수의 절반에 해당하는 이미지가 훈련 데이터셋에 더해졌다.

같은 훈련 횟수를 통해 기존 데이터셋에서 향상된 성능을 보여줌을 확인하였다.

표 2. 데이터 증강 방법별 결과

	a	b	c
훈련 횟수	32,000	32,000	32,000
원근 변환으로 추가된 이미지수	0	42,000	21,000
총 이미지 수	42,000	84,000	63,000
mIoU	47.9%	46.24%	48.38%

위의 방법은 실제 도보 환경에 적용하였을 때 더욱 극적인 성능 변화를 관찰할 수 있었다.



a. 원본 이미지 훈련 b. 원근 변환 훈련

그림 3. 원근 변환 이미지로 훈련된 모델과 원본 이미지로 훈련된 모델의 결과 비교

참고문헌

[1] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton “ImageNet Classification with Deep Convolutional Neural Networks” Advances in Neural Information Processing Systems 25 (NIPS 2012).

[2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”, CVPR 2018

[3] Mingyuan Fan, Shenqi Lai, Junshi Huang, Xiaoming Wei, Zhenhua Chai, Junfeng Luo, Xiaolin Wei, “Rethinking BiSeNet For Real-time Semantic Segmentation”, CVPR 2021