

의미론적 영상 분할 방법을 이용한 딥러닝 기반 차량 파손 영역 검출 시스템

서준혁⁰¹ 윤현주²
금오공과대학교 컴퓨터공학과
ssam2s@kumoh.ac.kr, juyoon@kumoh.ac.kr

Deep Learning-Based Vehicle Damage Area Detection System using Semantic Segmentation

JunHyeok-Seo⁰¹ HyeonJu-Yoon²
Department of Computer Engineering, Kumoh National Institute of Technology

요 약

본 논문에서는 AI-Hub 플랫폼에서 제공하는 4,000개의 차량 파손 영상 데이터를 딥러닝으로 학습하여 차량의 파손 영역을 검출하는 시스템을 제안한다. 검출의 정확도를 높이기 위하여 UNet++ 모델을 통한 의미론적 영상 분할 방법을 사용하였고 이를 통해 입력 영상에서 파손 영역을 픽셀 단위로 검출한다. 학습을 끝낸 모델의 성능 평가에서 평균 IoU(Intersection over Union)를 측정한 결과 약 0.94의 수치를 보였으며 객관적인 성능 평가를 위해 기존의 U-Net 모델을 동일한 환경에서 시험하여 그 성능을 비교, 평가한다.

1. 서 론

오늘날 차량 관리는 대여, 보험 등 다양한 분야에서 필수적으로 요구된다. 차량의 관리에 있어 파손 여부 및 그 부위를 꾸준히 파악하는 것은 하나의 재물(財物)로 귀속되는 차량의 가치가 훼손되는 것을 방지하기 위한 매우 중요한 행위이다. 이러한 이유로 다양한 딥러닝 기술을 이용한 차량 파손 영역 검출 관련 연구[1][2][3]가 꾸준히 진행되고 있다. 하지만 대다수의 연구는 영상에서 파손 부위를 지역화(Localization)하여 검출하는 객체 인식(Object Detection) 방식[2][3]을 사용하는데, 이는 대략적인 파손 부위의 파악에는 용이하지만 정확한 영역을 픽셀 단위로 검출하지는 못하기 때문에 비교적 정확도가 떨어지고 작은 파손 부위는 잘 검출하지 못한다는 한계가 존재한다.

이러한 한계를 해결하기 위해 영상 분할 방식을 이용한 연구도 존재하는데, 대표적으로 국내 카세어링 업체인 쏘카(SOCAR)에서 진행된 연구[1]가 있다. 해당 연구는 육안으로 파손 여부 및 그 영역을 식별하는 것이 가능한 2,000장의 영상을 데이터로 사용하여 U-Net을 통한 의미론적 영상 분할 방법(Semantic Segmentation)으로 차량 파손의 종류와 그 영역을 픽셀 단위로 검출하였다. 영상 분할에 대한 성능 측정 결과 Threshold 0.5를 기준으로 약 0.96의 평균 IoU를 기록했다.

본 논문에서는 쏘카의 선행 연구를 참고하여 의미론적 영상 분할 방식을 사용하되 U-Net[4] 모델의 구조를 수정해 성능을 향상한 UNet++[5] 모델을 사용하여 시스템을 구현한다. 이에 더불어 최종적으로 학습된 모델이 검

출하는 결과에 대해 평균 IoU를 성능 지표로 정하여 같은 환경에서 학습된 U-Net 모델과의 성능을 비교, 평가한다.

본 논문의 구성은 다음과 같다. 2장에서는 학습 및 시험을 위해 준비한 데이터셋을 전처리하는 과정을 설명하고, 3장에서는 딥러닝을 이용하여 해당 데이터셋을 UNet++ 모델에 학습시키는 과정 및 결과를 설명한다. 4장에서는 학습된 모델의 성능 평가를 진행하고 대조 모델과 성능 비교 및 평가를 진행한다. 마지막으로 5장에서는 결론을 맺고 향후 연구 방향에 대해 논의한다.

2. 데이터셋 전처리

본 논문에서는 학습 및 시험을 위한 데이터셋을 구성하기 위해 국내 AI 공개 데이터 플랫폼인 ‘AI-Hub’에서 제공하는 ‘차량 파손 이미지’ 데이터를 확보하였다. 확보한 데이터셋은 2021년에 구축되어 2023년 2월에 최종 갱신된 AI-Hub 차량 파손 영상 데이터이다. 해당 데이터셋에는 차종 및 색상별로 다양한 영상이 존재하며 4가지 분류(스크래치/찌그러짐/파손/이격)에 따른 파손 종류와 파손 부품 및 영역이 라벨링 되어 있다.

해당 데이터셋은 파손 종류가 분류되어 있는데, 이를 모두 하나의 파손으로 검출하게 하기 위해 이 분류를 하나로 묶고 해당 파손 영역들의 좌표를 함께 저장했다. 또한, 클래스가 하나라면 신경망에서 영상을 이진 분류(Binary Classification) 해야 하기에 원본 영상에서 파손에 해당하는 픽셀은 1로, 파손에 해당하지 않는 픽셀은 0으로 만드는 방식으로 동일한 크기의 이진 마스크 영상(Binary Mask Image)을 생성했다.

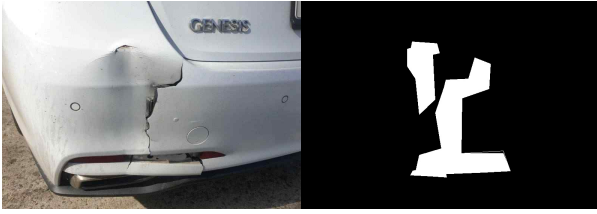


그림 1. 원본 영상과 생성된 이진 마스크 영상

위와 같은 방식으로 총 4,000개의 영상 데이터를 전처리하였고 학습, 검증, 시험 데이터의 비율을 8:1:1의 비율로 구성하여 각각 3,200:400:400의 비율로 분배했다. 모든 영상 데이터는 그림 1과 같이 육안으로 충분히 검출이 가능한 영상부터 그림 2와 같이 육안으로는 구분이 힘든 영상까지 골고루 분포되어있다.



그림 2. 육안으로 파손 영역 식별이 어려운 영상

3. 모델 구성 및 학습

본 논문에서는 영상 분할을 위한 딥러닝 모델로 U-Net 모델을 개선한 UNet++ 모델을 사용한다. 또한, 모델의 전반적인 성능을 향상하기 위해 신경망 내부에서 feature extraction을 수행하는 encoder를 ImageNet을 사전 학습한 EfficientNet-B7 모델로 구성한다.

학습 환경은 CUDA 11.7, CUDNN 8.6.0, RTX 3090의 로컬 환경에서 PyTorch 1.13.1 버전으로 구성하고, 학습을 위한 hyper parameter는 batch 16, learning rate 0.0001으로 설정하였으며 validation loss가 연속해서 3회 이상 감소하지 않으면 과대 적합(Overfitting)을 방지하기 위해 학습을 조기 종료하도록 설정하였다.

모델의 학습 과정에서 어두운 곳에서 찍힌 영상들을 고려하여 밝기에 대한 augmentation을 전체 학습 데이터 중 약 20% 비율로 적용하여 학습을 진행하였다. 학습 결과 48 epoch에 학습이 조기 종료되었으며 training loss는 약 0.03, validation loss는 약 0.26에 도달하였다.

4. 모델 성능 평가

본 논문에서는 실질적인 영상 분할 성능인 IoU를 성능 지표로 정한다. 학습이 끝난 모델에 대해 시험 데이터 400장을 대상으로 평균 IoU를 평가한 결과, 0.9430의 수치를 보였다. 이 값은 400장의 영상에서 실제 정답 픽셀과 모델이 분할한 픽셀의 교집합을 합집합으로 나눈 값을 의미한다. 즉, 정답과 예측의 일치도를 의미하며 94.3%의 평균 일치도를 보였다고 해석할 수 있다.

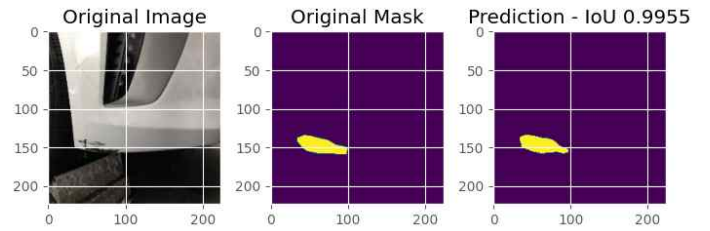


그림 3. 평가 IoU 상위 5개 영상 중 하나

그림 3은 IoU가 가장 높게 평가된 상위 5개 영상 중 한 장이다. 왼쪽부터 차량 파손 영상, 실제 파손 영역, 모델이 예측한 영역에 대한 영상이며 차량의 하단에 생긴 파손을 거의 동일하게 검출해내는 것을 확인했다.

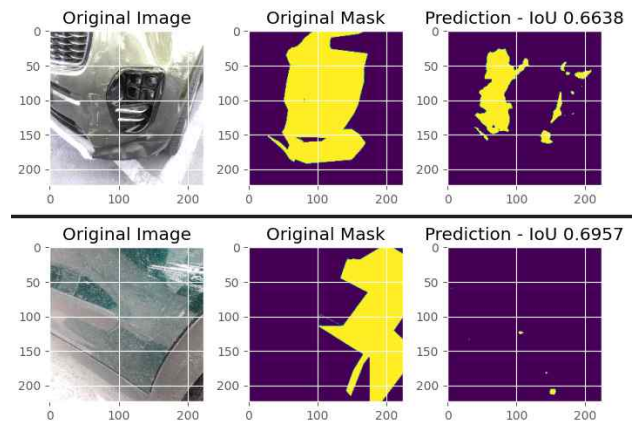


그림 4. IoU가 0.7 이하로 평가된 영상

그림 4는 모델의 성능 평가 과정에서 IoU가 0.7 이하로 평가된 영상 2장이다. 원인을 분석하는 과정에서 두 영상 모두 차량의 외관에 빛이나 주변 풍경이 많이 반사된 것을 확인할 수 있었으며 IoU가 높게 평가됐던 그림 3과 비교해보았을 때 확연히 그 차이를 확인할 수 있었다.

본 논문에서 대조 모델로 선정한 U-Net 모델을 동일한 환경 및 데이터로 학습시키고 평균 IoU를 평가한 결과 0.9232의 수치를 보이며 UNet++ 모델보다 약 0.02 낮았다.

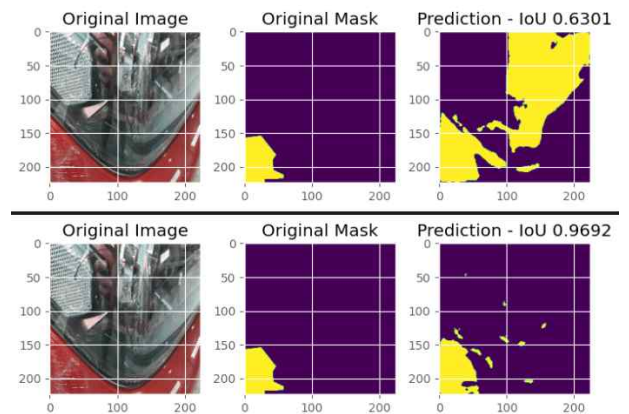


그림 5. 두 모델의 IoU 값의 차이가 가장 큰 영상

그림 5는 두 모델이 예측한 영상 중 평가된 IoU가 가장 크게 차이가 나는 영상이다. 모델의 예측 영상을 시각화한 항목인 ‘Prediction’에서 상단은 U-Net 모델의 예측 결과, 하단은 UNet++ 모델의 예측 결과이며 실제 정답과 비교해보았을 때 확연한 차이를 보이는 것을 확인할 수 있다. U-Net 모델의 예측 결과에서는 차량의 전 조등 덮개에 반사된 부분까지 파손 영역으로 검출했지만, UNet++ 모델의 예측 결과에서는 비교적 실제로 파손된 영역만 검출해냈다.

이와 더불어 선행 연구[1]와 유사한 조건으로 모델을 재구성하여 성능을 비교 및 분석하였다. 앞서 다뤘던 4,000개의 데이터셋에서 육안으로 확실히 식별 가능한 영상 2,000장을 걸러내 학습 및 시험을 진행하였고 그 결과 비교는 다음 표 1과 같다.

사용 모델	데이터 크기	평균 IoU
U-Net	2,000	0.9223
U-Net	4,000	0.9232
UNet++	2,000	0.9328
UNet++	4,000	0.9430

표 1. 모델 및 데이터 수에 따른 평균 IoU

전반적으로 식별이 쉬운 2,000개의 데이터를 사용했을 때 보다, 식별이 어려운 영상을 포함한 4,000개의 데이터를 사용했을 때 좋은 성능을 보였으며 같은 조건에서는 U-Net 모델보다 UNet++ 모델이 더 나은 성능을 보였다. 이를 통해 동일한 환경에서 같은 영상에 대해 차량의 파손 영역을 검출할 때 UNet++ 모델이 더 좋은 결과를 보이는 것을 확인할 수 있었다.

5. 결론

본 논문에서는 차량의 파손 영역을 검출하기 위해 의미론적 영상 분할 딥러닝 모델인 UNet++ 모델에 육안으로 식별이 힘든 영상을 포함한 4,000장의 차량 파손 영상 데이터를 학습시켰다. 학습이 끝난 모델의 성능 평가에서 평균 IoU는 0.9430의 값을 보였고 시각화한 결과에서도 양호한 파손 영역 검출 성능을 보였다.

또한, 데이터 수와 육안으로 식별이 힘든 영상의 포함 여부에 따라 각각 다르게 구성한 2가지 데이터셋에서 U-Net 모델과 UNet++ 모델을 딥러닝을 통해 학습시키고 시험한 4가지 결과를 비교하였다. 전반적으로 식별이 힘든 영상이 포함되어있더라도 데이터의 수가 많을 때 좋은 결과를 보였고 2가지 데이터셋 모두에서 UNet++ 모델이 U-Net 모델보다 좋은 검출 성능을 보였다.

다만 빛이 심하게 반사된 영상들에서 상대적으로 낮은 검출 정확도를 보이는 경우가 발생해 향후 연구에서 이를 해결할 수 있는 영상처리 기술을 전처리 과정 또는 학습 과정에 적용한다면 차량 파손 영역 검출 성능을 향상할 수 있을 것으로 기대된다.

참 고 문 헌

- [1] SOCAR Tech, “Developing car damage detection model using semantic segmentation,” <https://tech.socarcorp.kr/data/2020/02/13/car-damage-segmentation-model.html>, Feb. 2020
- [2] Jong Won Jeon, Hyo Seop Lee, Hee Il Hahn.(2021).Improving the Vehicle Damage Detection Model using YOLOv4.Journal of IKEEE,25(4),750-755.
- [3] H. S. Malik, M. Dwivedi, S. N. Omakar, S. R. Samal, A. Rathi, E. B. Monis, B. Khanna, and A. Tiwari, “Deep learning based car damage classification and detection,” EasyChair preprints, 2020.
- [4] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” arXiv:1505.04597 [cs.CV], 2015.
- [5] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, Jianming Liang, “UNet++: A Nested U-Net Architecture for Medical Image Segmentation.” arXiv:1807.10165 [cs.CV], 2018.