

높은 재현율의 차량 파손 탐지를 위한 2 단계 정제 구조

정상민, 안진현*
명지대학교

jmmm0604@mju.ac.kr, *wlsqus3396@mju.ac.kr

요 약

본 논문은 실제 렌터카 산업 현장에서 핵심적으로 요구되는 차량 파손 탐지 작업을 대상으로 한다. 렌탈/반납 시 촬영 환경의 편차(조명, 각도, 거리, 배경 등)로 인해 손상과 유사한 패턴(반사, 그림자, 오염 등)이 빈번히 발생하며, 단일 단계 탐지 모델만으로 재현율을 높일 경우 오탐이 급증하는 문제가 나타난다.^[1] 이에 본 논문은 높은 재현율을 유지하면서도 오탐을 억제하기 위해, 탐지 기반 후보 생성과 분류 기반 후보 정제를 분리한 2 단계 정제 구조를 제안한다.^[2] 1 단계에서는 파손 부위를 경계 박스로 생성하는 수준의 정확도와 추론 속도를 기준으로 탐지 모델을 선정하고, 손상 가능성이 조금이라도 있는 영역은 후보 박스를 폭넓게 생성하여 미탐을 최소화한다. 2 단계에서는 동일 조건에서 여러 이미지 분류 모델을 비교·평가한 뒤 최적 모델을 선택하여 1 단계 후보 박스를 재판별하고, 정상으로 판단된 후보를 제거함으로써 최종 결과에서 오탐을 정리한다. 또한 모델 크기 및 입력 해상도 변화에 따른 학습·추론 효율과 성능 변화를 함께 분석하여, 현장 적용 관점에서의 실용성을 검토한다.

I. 서 론

현재 렌터카 및 모빌리티 서비스 현장에서는 차량 상태 점검이 주로 육안 검사에 의존한다. 그러나 이 방식은 점검자마다 판단 기준이 다르고, 촬영 환경(조명, 각도, 거리 등)에 따라 동일 부위라도 판독 결과가 달라지는 등 일관성이 떨어진다는 한계가 있다. 특히 경미한 스크래치나 파손은 쉽게 놓치거나 과대 판단되기 때문에 고객과 업체 간 비용 청구 분쟁으로 이어지는 사례가 빈번하며, 결과적으로 점검 시간 증가와 서비스 신뢰도 저하로 연결될 수 있다. 따라서 객관적이고 자동화된 차량 손상 탐지 시스템의 도입이 요구된다.^[3]

본 연구의 목적은 차량의 신규 손상을 자동으로 판별하고, 손상의 위치와 유형을 시각적으로 제시하는 시스템을 구축하는 것이다. 특히 실제 서비스 환경에서는 손상을 놓치는 것(미탐)이 오탐보다 더 치명적이며, 단일 단계 탐지 모델로 재현율을 높일수록 오탐이 급증하는 경향이 있다는 점에 착안하였다. 이에 본 연구는 탐지 기반 후보 생성과 분류 기반 후보 정제를 분리한 2 단계 모델 구조를 설계하여, 높은 재현율을 확보하면서도 최종 결과의 정확도를 향상시키고자 한다.

II. 본론

제안 구조의 핵심은 1단계에서 후보 박스를 적극적으로 생성하여 재현율을 확보하고, 2단계에서 분류를 기반으로 한 필터링을 통해 오탐을 정리하는 데 있다. 이와 같은 2단계 분리는 1단계에서 발생할 수 있는 과잉 탐지를 2단계에서 흡수하도록 설계되었다. 또한 1단계와 2단계의 역할을 명확히 분리함으로써, 현장 환경 변화에 따라 각 단계의 임계값이나 모델을 독립적으로 조정할 수 있다는 장점이 있다.

1단계에서는 실시간 추론 속도가 우수한 YOLO 계열 탐지 모델을 사용하여 이미지 내 잠재적 손상 부위를 폭넓게 탐지한다.^[4] 이때 실제 촬영 이미지에서는 반사, 그림자, 얼룩, 먼지, 물기 등 손상과 유사한 패턴이 빈번히 나타나며, 이러한 요소가 탐지 모델의 오탐을 유발할 수 있다. 따라서 본 연구는 재현율을 극대화하기 위해 세부 손상 클래스를 하나로 통합한 ‘Damage’ 클래스로 학습을 진행하고, 미세한 이상 징후도 우선 후보군으로 포함시키는 접근을 적용한다.^[5] 즉, 1단계는 손상 여부를 확정하기보다 손상 가능성이 있는 영역을 최대한 놓치지 않고 수집하는 역할에 집중한다. 후보

박스는 이후 단계에서 재판별이 가능하도록 충분한 다양성을 갖도록 생성되며, 1단계의 판단 기준은 미탐을 최소화하는 방향으로 설정된다.

2단계는 1단계에서 생성된 후보 박스를 대상으로 분류를 수행하여 정상으로 판단된 후보를 제거함으로써 오탐(FP)을 감소시키는 단계이다.^[6] 2단계에서는 1단계에서 생성한 후보 박스를 입력으로 받아, 각 후보가 실제 손상인지 여부를 판별하고 필요 시 손상 유형까지 세분화한다. 이때 2단계 분류 모델은 동일한 조건에서 여러 이미지 분류 모델을 비교한 뒤, 평가 지표를 기반으로 최종 모델을 선정한다. 비교 과정에서는 학습 조건(배치 크기, 입력 크기, epoch 등)을 통일하여 모델 간 성능 비교의 공정성을 확보한다.

본 연구에서는 비교 실험 결과를 바탕으로 ViT 기반 분류기를 채택하여 1단계 후보 박스를 정제한다.^[7] ViT는 후보 영역이 실제 손상인지, 혹은 빛 반사·오염·그림자 등으로 인해 발생한 유사 패턴인지를 판별한다. 특히 차량 외관의 촬영 이미지에서 흔히 발생하는 경계의 불명확성, 표면 재질 변화, 도색 색상 차이 등으로 인해 손상 여부가 애매한 경우에도, 2단계 분류기가 후보 영역을 재해석함으로써 최종 오탐을 효과적으로 줄일 수 있다. 실제 손상으로 판단된 경우에는 파손, 찌그러짐, 스크래치, 이격의 4가지 세부 유형으로 최종 분류하여, 사용자에게 손상의 위치와 유형을 직관적으로 제공한다.

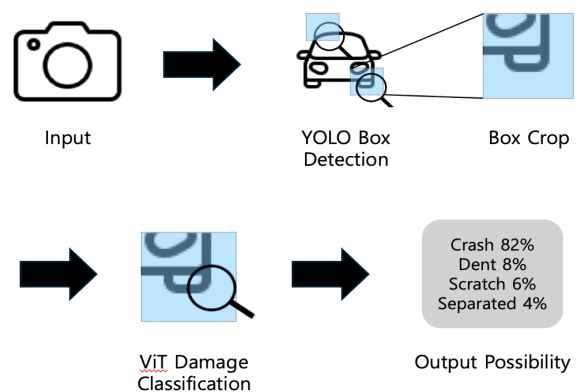


그림 1. 전체 파이프라인 구조

III. 실험

학습에 사용된 데이터셋은 파손, 찌그러짐, 스크래치, 이격의 4개 세부 클래스로 구성된 실제 차량 파손 이미지로 이루어져 있다.^[1]

1단계 탐지 단계에서는 YOLOv8의 모델 크기(n/m)와 입력 해상도(640/1024)를 비교하고, 성능뿐 아니라 학습 소요 시간까지 함께 고려하여 최종적으로 v8m 모델을 선택하였다.^[4] 1024 해상도와 비교했을 때 재현율은 큰 차이가 없었으나, 학습 소요 시간은 크게 증가하는 경향이 관찰되어 효율 측면에서 640 해상도가 더 적합하다고 판단한 결과이다. 또한 입력 해상도에 따른 탐지 결과의 안정성을 함께 확인하여, 실사용 환경에서의 적용 가능성을 고려하였다.

2단계 분류 단계에서는 여러 이미지 분류 모델을 동일 조건에서 학습·평가한 뒤, Accuracy와 Macro F1-score를 기준으로 최종 모델을 선정하였다. 비교 실험 결과 두 지표에서 상대적으로 우수한 성능을 보인 ViT 모델을 채택하여 1단계 후보 박스를 정제하였다.^[5] 이 과정에서 1단계 후보 중 정상으로 판단되는 영역을 제거함으로써, 최종 출력에서 불필요한 검출을 줄이도록 설계하였다.

실험 결과, YOLO 단일 모델만 사용했을 때보다 2단계 정제 구조를 적용했을 때 오탐이 유의미하게 감소하였으며, 사용자 화면에는 최종적으로 정제된 결과만 제공되어 실제 점검 업무의 효율성과 신뢰도를 높일 수 있었다. 또한 후보 정제 이후에도 재현율 저하가 크지 않아, 미탐을 최소화하려는 목표와 오탐 억제 효과를 동시에 확인할 수 있었다.

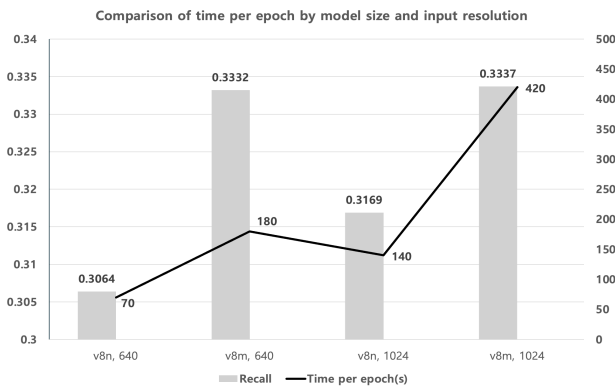


그림 2. YOLOv8 모델 크기 및 해상도에 따른 한 epoch 당 소요시간 비교

표 1. 성능 지표 비교를 통한 ViT 모델 선택

	ViT	ResNet-50	ResNet-101	ResNet-152	EfficientNet
Accuracy	0.6400	0.5625	0.6025	0.5000	0.6075
Macro F1-Score	0.6379	0.5637	0.5985	0.4688	0.6088

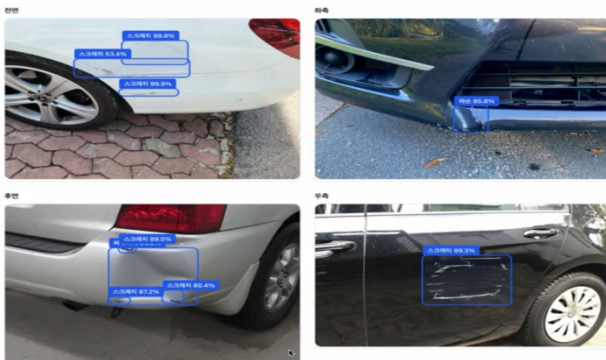


그림 3. 최종 손상 탐지 결과 시각화

IV. 결론

본 논문에서는 렌터카 반납/점검과 같이 미탐 비용이 큰 차량 파손 탐지 환경을 대상으로, 탐지 기반 후보 생성과 분류 기반 후보 정제를 분리한 2단계 정제 구조를 제안하였다. 촬영 조건의 편차(조명, 각도, 거리, 배경 등)로 인해 손상과 유사한 패턴(반사, 그림자, 이물질 등)이 빈번히 나타나는 현장에서 단일 단계 탐지 모델은 재현율을 높일수록 오탐이 급증하는 문제가 발생한다.

이에 1단계에서는 손상 가능성이 조금이라도 존재하는 영역을 폭넓게 후보로 생성해 미탐을 최소화하고, 2단계에서는 1단계 후보를 추가 분류로 재판별하여 정상 후보를 제거함으로써 오탐을 완화한다.

이에 1단계에서는 손상 가능성이 조금이라도 있는 영역을 폭넓게 후보로 생성하여 미탐을 최소화하고, 2단계에서는 1단계 후보를 추가 분류로 재판별해 정상으로 판단된 후보를 제거함으로써 오탐을 완화한다. 이러한 분리 구조를 통해 단일 단계에서 나타나는 재현율-정밀도 간 상충 관계를 구조적으로 완화할 수 있다. 실험에서는 모델 크기 및 입력 해상도에 따른 학습 효율과 성능 변화를 함께 비교하였고, 표준 성능 지표를 기반으로 제안 구조의 유효성을 검증하였다.

ACKNOWLEDGMENT

본 과제(결과물)는 2025 년도 교육부 및 경기도의 재원으로 경기 RISE 센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과입니다. (2025-RISE-09-A15)

This research was supported by the Regional Innovation System & Education(RISE) program through the Gyeonggi RISE Center, funded by the Ministry of Education(MOE) and the Gyeonggi-do, Republic of Korea.(2025-RISE-09-A15)

참 고 문 헌

- [1] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *Proc. ICCV*, 2017, doi: 10.1109/ICCV.2017.324.
- [2] H.-j. Yu and H.-C. Kim, "Vehicle Damage Detection with YOLO and Visual Embedding Fusion," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 29, no. 11, pp. 1481– 1488, 2025.
- [3] X. Wang, W. Li, and Z. Wu, "CarDD: A New Dataset for Vision-based Car Damage Detection," *arXiv:2211.00945*, 2022, doi: 10.48550/arXiv.2211.00945.
- [4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proc. CVPR*, 2016, pp. 779– 788, doi: 10.1109/CVPR.2016.91.
- [5] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving Into High Quality Object Detection," *Proc. CVPR*, 2018, doi: 10.1109/CVPR.2018.00644.