

상황적 데이터 증강 기법을 통한 객체 인식 정확도 향상 연구

황현성, 홍기훈*

서원대학교

subj22@gmail.com, *khong@seowon.ac.kr

A Study on Improving Object Recognition Accuracy Using Contextual Data Augmentation Techniques

Hwang Hyeon Seong, Kihun Hong*

Seowon Univ.

요약

본 연구는 도심 보도에 방치 및 불법 주정차된 공유 전동 킥보드를 안정적으로 탐지하기 위해 YOLOv8m 기반 객체 검출 모델을 구축하고 방치된 도로 상황에 필요한 데이터 증강 기법을 적용한 연구이다. 전동 킥보드 데이터셋 원본과 기본 증강(Basic Aug) 그리고 상황 기반 증강(Proposed) 등 3종 데이터셋을 구성하고 쓰러짐/기울어짐(Fallen/Tilted), 강한 그림자(Shadow), 부분 가림(Occlusion)을 모사하는 맞춤형 증강 모듈을 설계, 적용하였다. 동일한 학습 설정에서 제안하는 방법은 Precision 0.9295, Recall 0.8829, mAP@50-95 0.6442로 가장 우수한 성능을 달성하였다. 결과적으로 도심 환경의 다양한 비정상적 상황을 데이터 수준에서 반영하는 전략이 미탐 감소와 일반화 성능 향상에 효과적임을 보여주고 있다.

* 교신저자

I. 서론

최근 공유 모빌리티 확산으로 전동 킥보드 이용이 일상화되었으나 보도, 횡단보도 인근 방치와 무질서한 주정차는 보행권 침해 및 안전사고 위험을 키우는 사회적 문제로 부각되고 있다. 기존 관리 방식은 GPS 좌표와 앱 기반 규칙에 크게 의존하는데, 실제 현장에서는 위치 오차, 건물 반사, 지형 영향 등으로 정밀한 상태 판단이 어렵고, 결국 인력 단속과 수거에 의존하게 되어 비용과 대응 속도 측면에서 한계가 발생한다. 영상 기반 탐지는 물리적 상태(쓰러짐, 기울어짐, 보도 점유 등)를 직접 확인할 수 있다는 점에서 유효하지만, 도심 영상은 시간대별 명암 변화, 그림자, 부분 가림, 다양한 촬영 각도, 거리 등 복합적인 변수가 존재해 모델이 쉽게 혼동한다. 따라서 실사용을 위해서는 일반적인 학습 데이터만으로는 부족하며, 실패가 잦은 상황을 데이터 차원에서 충분히 재현해 학습시키는 접근이 필요하다. 본 연구는 쓰러짐, 그림자, 가림을 중심으로 현실 장면을 모사하는 증강 방법을 설계하고, 이를 통해 전동 킥보드 탐지의 미탐을 줄이며 안정적 성능을 확보하는 것을 목표로 하였다.

II. 제안하는 상황적 증강 방법

본 연구는 도심 환경에서 방치, 불법 주정차된 공유 전동 킥보드를 안정적으로 탐지하기 위한 객체 검출 시스템을 설계하고, 모델 구조 변경보다 데이터셋 구성과 상황 기반 증강 전략이 탐지 성능에 미치는 영향을 정량적으로 분석하였다. 전체 처리 절차는 도심 전동 킥보드 이미지를 수집하여, 전처리 및 기본 증강, 상황 기반 증강을 포함하여 각각 학습 데이터셋을 구축하고, YOLOv8m 기반으로 학습, 추론, 평가, 시각화 단계로 구성하였다. 이러한 데이터 증강 기법을 통해 탐지 성능을 끌어 올리는 것을 목표로 하였다. 데이터 수집은 전동 킥보드가 많이 배치되는 도심 보도,

건물 출입구 주변, 자전거 도로, 보도와 차도의 경계부 등을 중심으로 선별하였고, 직사광선이 강한 정오, 건물 그림자가 길게 드리우는 오후, 흐린 날 등 다양한 조명 조건을 포함하도록 계획하였다. 총 데이터셋은 1,231 이미지 샘플로 Roboflow 데이터셋 1,031개 샘플[1]과 한국의 전동 킥보드 이미지를 반영하기 위해 웹에서 크롤링한 샘플 200개를 사용하였다. 모델 평가의 신뢰성을 위해 데이터는 학습, 검증, 테스트로 7:2:1 분할했고, 동일 위치에서 촬영된 이미지가 서로 다른 세트에 동시에 포함되지 않도록 하였다. 전처리는 YOLO 입력 형식에 맞춰 letterbox 방식으로 모든 이미지를 640×640으로 맞추되 원본 비율을 유지하고, 바운딩 박스 좌표는 스케일링과 패딩 오프셋을 반영해 재계산하였다. 기본 증강은 과도한 분포 왜곡을 피하기 위해 좌우 반전, 밝기, 대비 조절, 약한 블러 등 최소 수준으로 제한하고, 특히 상하 반전처럼 실제 도로 환경과 맞지 않는 변형은 배제하는 방향으로 운영하였다.

여기에 본 연구의 핵심인 상황 기반 증강을 추가하여, 도심에서 탐지를 어렵게 만드는 쓰러짐, 기울어짐, 강한 그림자, 부분 가림을 데이터 수준에서 재현했다. Fallen, Tilted 모듈은 이미지와 바운딩 박스를 함께 회전 변환하여 쓰러진, 기울어진 킥보드 장면을 생성하고, Shadow 모듈은 다양한 방향과 강도의 그림자 마스크를 합성해 조명 조건이 열악한 상황을 모사하며, Occlusion 모듈은 쓰레기통, 나무 기둥, 도로 표지판 등 가림 물체를 자연스럽게 합성해 부분 가림 상황을 만들었다. 모듈 적용은 확률 기반으로 조합을 제어해 특정 상황만 과도하게 쏠리지 않도록 했고, 최종적으로 원본(Original), 기본 증강 중심(Basic Aug, 2,287장), 기본 증강과 상황 기반 증강을 모두 포함한(Proposed, 약 4,300장) 3종의 데이터셋을 구축하여 증강 방법이 성능에 미치는 영향을 비교하도록 설계했다.

탐지 모델은 Ultralytics 공식 구현인 YOLOv8m을 그대로 사용하고, 세 실험 모두 입력 640×640, 에폭 120, 배치 8로 통일하였다. 초기 학습률

0.01에서 cosine 스케줄로 점진 감소시키며, 옵티마이저는 SGD (Stochastic Gradient Descent)를 사용하여 모멘텀 0.937, weight decay 0.0005로 설정해 안정적인 수렴을 유도했다. 내장 증강 하이퍼 파라미터도 비교의 일부로 간주해 동일하게 유지했고, 회전, 이동, 스케일, HSV 변환 등을 설정하되, MixUp과 Copy Paste는 비현실적 겹침을 줄이고 제안 증강의 효과를 명확히 보기 위해 적극적으로 사용하지 않았다. Mosaic은 데이터 다양성을 제공하지만 작은 객체가 더 작아지는 문제가 있어 학습 초기 일부 구간에서만 제한적으로 사용하고 후반에는 비활성화하는 방식으로 정밀도를 보완했다. 평가는 공통 테스트 세트에서 Precision, Recall, mAP@50-95, F1을 측정하고, 동일 테스트 이미지를 입력했을 때 세 모델의 박스 출력 결과를 한 화면에서 비교할 수 있는 웹 기반 UI를 구현해 정성적 차이도 확인했다.

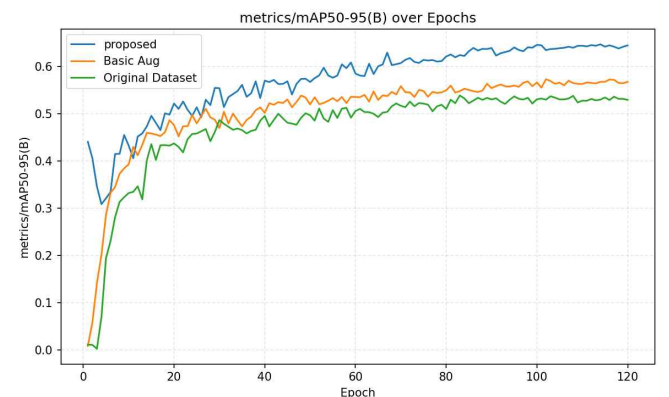


그림 1. 제안 방법의 mAP@50-95 그래프

표 1. 데이터셋 별 성능 비교

데이터셋	Precision	Recall	mAP@50-95
Original Dataset	0.9246	0.6683	0.5292
Basic Aug	0.8563	0.7270	0.5670
Proposed	0.9295	0.8829	0.6442

결과적으로 Original은 Precision 0.9246, Recall 0.6683, mAP@50-95 0.5292로 정밀도는 높지만 미탐이 많은 보수적 탐지 양상을 보였고, Basic Aug는 Recall: 0.7270, mAP@50-95: 0.5670으로 개선되었으나 Precision이 0.8563으로 낮아져 오탐 증가 가능성이 관찰되었다. 반면 Proposed는 Precision: 0.9295, Recall: 0.8829, mAP@50-95: 0.6442로 가장 우수했으며, 특히 Recall이 크게 상승해 방치, 불법 주정차 탐지 목적에서 중요한 얼마나 많이 찾아내는가를 개선했다. 다른 물체에 의해 가려진 상황을 고려하면 mAP@50-95 수치는 우수한 편이며 가려져서 일부만 보이는 객체에 대한 박스의 범위는 추가 연구가 필요해 보인다.

종합하면, 온라인 기본 증강만으로는 쓰러짐, 강한 그림자, 부분 가림 같은 도심 특수 상황을 충분히 재현하기 어렵고, 목표 상황을 직접 모사하는 상황 기반 증강을 데이터셋에 포함할 때 미탐이 유의미하게 감소하며 전반적인 성능이 향상되었다. 이는 전동 킥보드에 국한되지 않고 도심 환경에서 정지, 방치될 수 있는 다른 객체 탐지에도 확장 가능한 데이터 증강 기법이라고 할 수 있을 것이다.

III. 결론

본 연구는 도심 환경에서 방치, 불법 주정차된 공유 전동 킥보드를 안정적으로 탐지하기 위해 객체가 위치한 상황을 기반으로 증강 기법을 제안하여 성능이 개선됨을 실험적으로 확인하였다. 원본 데이터만으로 학습한 경우, 정밀도는 높았지만 쓰러짐, 그림자, 부분 가림과 같은 실제 현장 조

건에서 미탐이 증가해 재현율과 mAP@50-95가 제한되는 문제가 나타났다. 기본 증강은 일정 수준의 성능 향상에 기여했으나 정밀도 하락이 동반될 수 있었고, 현실 세계에서 자주 발생하는 복잡하고 탐지가 어려운 상황을 충분히 반영하지 못하였다. 반면 본 연구에서 제안한 실제 상황에서 많이 발생하는 Fallen, Tilted, Shadow, Occlusion 기반 증강을 포함한 데이터셋은 재현율과 mAP@50-95를 유의미하게 향상시키면서도 정밀도를 안정적으로 유지해, 실 서비스 관점에서 중요한 미탐 감소 효과를 보여주었다. 이는 도심 특수 상황을 목표로 설계한 데이터 중심 접근이 탐지 모델의 일반화 성능을 강화하고, 제한된 원본 데이터에서도 활용 가능성을 높인다는 것을 의미한다. 또한 세 모델의 결과를 동일 화면에서 이미지로 비교할 수 있는 시각화 구현 환경을 구성함으로써 성능 차이를 정성적으로 검증하고, 현장 적용 시 고려해야 할 실패 사례를 구체적으로 확인하여 선별할 수 있다.

향후 연구에서는 더 다양한 도시, 기상, 야간 환경으로 데이터 범위를 확장하고, 킥보드의 상태 분류(정상, 쓰러짐, 보도 점유 등)와 같은 후처리 로직을 결합해 단독 기준에 가까운 판단까지 자동화할 필요가 있다. 더 나아가 영상 스트림 기반 실시간 처리, 경량 모델 적용, 오탐 역제를 위한 추가 학습 전략을 통해 실제 운영 시스템으로의 확장 가능성을 높일 수 있을 것이다. 종합하면 본 연구는 공유 전동 킥보드 관리 문제를 영상 기반으로 해결하기 위한 실증적 근거를 제시했으며, 데이터 설계와 증강 전략이 도심 객체 탐지 성능을 좌우하는 핵심 요인임을 보여 준다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획지원-대학ICT연구센터(ITRC)의 지원(IITP-2026-RS-2020-II201602, 100%)을 받아 수행된 연구임.

참 고 문 헌

- [1] Roboflow, Scooter Computer Vision Model Dataset, <https://university.roboflow.com/computer-vision-dibh7/scooter-31wb6>, 2025.
- [2] Syahrudin E., Utami E., Hartanto A. D., “Augmentation for Accuracy Improvement of YOLOv8 in Blind Navigation System,” Jurnal RES TI (Rekayasa Sistem dan Teknologi Informasi), 8(4), pp. 579 - 588, 2024.
- [3] Gilroy S., Mullins D., Jones E., Parsi A., Glavin M., “E-Scooter Rider Detection and Classification in Dense Urban Environments,” Results in Engineering, vol. 16, 2022.
- [4] Lee S., Oh S., Kim J., “YOLOv5-Based Electric Scooter Crackdown Platform,” Applied Sciences, vol. 15, 3112, 2025.
- [5] 박현식, 하정아, 신성필, 박상규, 홍기만, “영상 AI 기반 공유형 PM(Personal Mobility) 불법주정차 감지 기술 개발,” 한국건설기술연구원, 연구보고서 KICT 2023-107, 2023.
- [6] 우윤희, 유무영, “YOLOv8 및 YOLOv11 세그멘테이션 모델을 활용한 가설구조물 인식 성능 분석,” 대한건축학회논문집, vol. 41, no. 10, pp. 339-348, 2025.
- [7] 장준서, 추동명, 이서아, 노학균, 정관수, “환경 변화에 강건한 교통 CCTV 객체 인식을 위한 데이터 전처리 및 증강 기법 연구,” 2025년 한국컴퓨터정보학회 하계학술대회 논문집, 제33권 2호, pp. 949-952, 2025.