

YOLO 기반 소형 객체 탐지를 위한 타일 기반 학습과 배경 차분 게이팅

장다연^{1,3}, 한일주^{2,3}, 김봉완³, 김광수³

¹성결대학교, ²충남대학교, ³한국전자통신연구원 에어모빌리티연구본부
clavexn@gmail.com, martellato51@gmail.com, {kimbw, enoch}@etri.re.kr

Tile-based Training and Background Subtraction Gating for Small Object Detection based on YOLO

Dayeon Jang^{1,3}, Il-Ju Hahn^{2,3}, Bong Wan Kim³, Kwangsoo Kim³

¹Sungkyul University, ²Chungnam National University,

³Electronics and Telecommunications Research Institute (ETRI)

요약

다양한 위험 요인이 상존하는 건설 현장에서 소형 객체를 안정적으로 탐지하기 위해 본 연구에서는 타일 기반 학습과 배경 차분 게이팅을 결합한 방법을 제안한다. 원거리 고정형 카메라에서는 작업자와 안전모, 반사 조끼 같은 소형 객체의 표현력이 약해 일반적인 리사이즈 과정에서 정보 손실이 크다. 본 연구는 학습 단계에서 이미지를 타일로 분할해 소형 객체의 상대 크기와 특징을 보존하고, 추론 단계에서는 배경 차분으로 움직임이 있는 타일에만 YOLO를 적용한다. 실제 건설 현장 이미지와 영상으로 평가한 결과, 타일 기반 학습은 전체 이미지 학습 대비 지표 전반이 향상되었고, 특히, 병합 기준 평가에서 Precision 22.2% 및 Micro F1 19.6% 상승하였다. 배경 차분 게이팅은 평균 활성 타일 비율 0.256 환경에서 프레임 지연을 60.145ms에서 30.775ms로 감소시켰고 FPS를 16에서 30으로 증가시켰다. 이는 현장 영상에서 움직임의 시간적, 공간적 희소성을 활용해 성능 손실을 최소화 하면서 추론 비용을 실질적으로 절감할 수 있음을 보여준다. 하지만 타일 경계에서 객체 병합 처리가 필요하며, 대형 객체 위주 장면에서는 타일링과 게이팅의 부가 비용 대비 이득이 작아 전체 이미지 기반 추론이 정확도와 속도 모두에서 더 효율적이었다. 따라서 장면 특성에 따라 타일 기반 추론과 전체 이미지 추론을 동적으로 전환하는 전략이 필요하다.

I. 서론

건설 현장에는 대형 장비 이동, 고소 작업, 자재 반출입 등으로 상시 위험 요인이 존재한다. 2025년 고용노동부의 산업 재해 발생 현황에 따르면 건설업이 타 산업군 대비 사고 위험이 가장 높았다[1]. 이러한 환경에서 CCTV 기반 객체 탐지는 안전 관제를 자동화하는 핵심 수단이지만, 원거리 고정형 카메라의 특성상 작업자, 안전모, 반사 조끼 등 소형 객체의 시각적 표현력이 낮아 탐지 성능이 저하되기 쉽다[2].

YOLO 계열 객체 탐지기는 처리 속도와 정확도 측면에서 실용적이지만 고정 입력 크기와 리사이즈 과정에서 소형 객체의 특징이 축소되거나 소실될 수 있다. 학습 영상보다 운영 영상의 크기가 큰 경우 추론 단계에서 이미지를 타일 단위로 나누어 해상도를 국소적으로 올리는 방식으로 정확도를 개선 시킨 연구가 있었으나 경계 중복과 연산량 증가라는 한계가 있었다[3].

본 연구는 두 가지 개선 방안을 제안한다. 첫째, 학습 단계부터 타일 입력을 사용하여 소형 객체의 상대 크기와 특징을 보존하는 타일 기반 학습을 적용한다. 둘째, 배경 차분(background subtraction) 기반 게이팅을 도입하여 전경(foreground)이 검출된 활성 타일에서만 YOLO 추론을 수행하여 불필요한 연산을 줄인다[4]. 배경 차분은 시간에 따른 픽셀 통계를 이용해 배경 모델을 형성하고 현재 프레임과의 차이를 통해 전경 마스크를 얻는 기법이며, 게이팅은 활성 타일에만 YOLO 추론을 수행하고 비활성 타일은 건너뛰는 선택적 실행 정책을 의미한다. 배경 차분 결과로 움직임이 감지된 영상 영역을 전경이라 하고, 전경이 검출된 타일을 활성 타일이라 한다. 본 연구에서는 혼합 가우시안 기반의 MOG2(Mixture of

Gaussians)를 사용하여 프레임마다 전경 마스크를 계산하고 타일의 활성 임계치를 넘는 타일만 추론한다.

타일 기반 학습의 효과는 타일 기준 평가와 병합 기준 평가로 확인하였으며 배경 차분 게이팅의 연산 효율성은 다양한 운영 영상에서 게이팅 적용 전후를 비교해 프레임 지연과 처리 속도에 미치는 영향을 정량적으로 분석하여 검증하였다.

II. 본론

모델 학습에는 웹 검색과 AI Hub에서 수집한 1,511장의 이미지를 사용하였다. 원본 해상도는 188×188부터 5600×8400 px까지 다양하며, 이 중 1,058장을 전체 이미지 학습에 사용하였고, 동일 원본을 한 번 최대 800 px 기준으로 분할하여 얻은 2,702장의 이미지를 타일 기반 학습에 사용하였다. 분할 시에는 원본 바운딩 박스의 포함 비율 30% 이상인 타일만 학습 표본으로 채택하였고, 객체가 존재하지 않는 타일은 학습에서 제외하였다. 두 모델 모두 YOLOv8m을 사용하여 epoch 100 조건에서 학습하였다. 평가는 실제 건설현장의 1920×1080 해상도 CCTV 이미지 100장으로 구성된 테스트 세트에서 수행하였으며, 소형 객체인 작업자, 안전모, 반사 조끼 등의 비중이 충분히 포함되도록 구성하였다.

평가지표는 Precision, Recall, mAP50, mAP50-95, Micro F1, Macro F1을 사용하였다. 모델 평가는 타일 기준 평가와 병합 기준 평가를 함께 사용한다. 타일 기준 평가는 정답 라벨을 타일 단위로 분할하고 타일 예측과 직접 비교하는 방식으로 모델 자체의 소형 객체 인식률을 확인한다. 병합 기준 평가는 타일 예측을 원본 좌표계로 병합한 결과를 정답 라벨

과 비교하는 방식으로 실제 운영 산출물의 품질을 측정한다. 병합 기준은 같은 클래스의 바운딩 박스가 타일 경계에서 2px 이내로 접촉하고 70% 이상 중첩될 때 하나의 객체로 통합하였다.

표 1은 전체 이미지 학습과 타일 기반 학습을 타일 기준과 병합 기준으로 탐지 성능을 비교한 결과를 나타낸다.

표 1. 소형 객체 중심 테스트의 탐지 성능

Metric\Method	Full-image training		Tile-based training	
	Tile	Merged	Tile	Merged
Precision	0.626	0.594	0.846	0.816
Recall	0.453	0.547	0.597	0.722
mAP50	0.526	0.568	0.578	0.619
mAP50-95	0.362	0.352	0.446	0.393
Micro F1	0.526	0.57	0.7	0.766
Macro F1	0.351	0.454	0.496	0.548

소형 객체 비중이 높은 테스트 세트에서 타일 기반 학습은 전 지표에서 일관된 개선을 보였다. 병합 기준에서 타일 기반 학습은 Precision 22.2%, micro F1 19.6% 상승하였다. 타일 기준 평가에서도 유사한 개선이 관찰되어 타일 기반 학습이 소형 객체 표현 학습에 실질적으로 기여함을 확인하였다.

배경 차분은 MOG2(history=250, varThreshold=9)에 morphological opening(3, 3)을 적용하였다. 활성 타일 판정은 타일 픽셀 크기 기준으로 전경 픽셀 비율이 0.004 이상일 때로 정의하였다.

표 2는 연산 효율을 비교하기 위해 배경 차분 게이팅 적용 여부에 따른 평균 지연과 처리 속도를 비교한 결과이며, 실험에 사용된 총 프레임 수는 11,132장이다. Latency는 프레임당 평균 총 지연 시간을 의미하며, Active는 활성 타일 비율로 전체 타일 가운데 전경이 검출된 타일의 비중을 뜻한다. Active 값 0.256은 전체 타일 중 평균 25.6%에서 움직임이 관측되었음을 의미한다. BS는 배경 차분에 소요된 시간이며, YOLO는 활성 타일에 대한 YOLO 추론 시간이다. FPS는 초당 처리 프레임 수를 의미한다.

표 2. 배경 차분 게이팅에 따른 지연과 처리 속도

	Latency	Active	BS(ms)	YOLO(ms)	FPS
gating on	30.775	0.256	6.911	23.828	30
gating off	60.145	1	-	60.104	16

게이팅을 적용하여 지연 시간을 비교한 결과, 프레임당 지연이 평균 29.37ms 감소하고, FPS는 16에서 30으로 개선되었다. 이는 현장 영상에서 움직임의 시간적, 공간적 희소성을 활용하면 성능 손실을 최소화하면서 추론 비용을 실질적으로 절감할 수 있음을 보여준다.

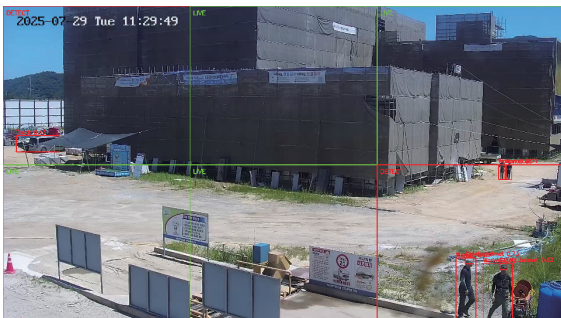


그림 1. 현장 영상에서의 배경 차분 게이팅과 YOLO 탐지 결과

그림 1은 실제 건설 현장 CCTV 영상에 배경 차분 게이팅을 적용해 활성 타일에서만 타일 기반 학습된 YOLOv8m을 수행한 결과로, 사람과 안전모가 검출된 예시를 나타낸다.

III. 결론

본 연구는 원거리 고정형 CCTV 환경에서 소형 객체 검출 성능 저하와 불필요한 연산 증가를 동시에 완화하기 위해 타일 기반 학습과 배경 차분 게이팅을 결합한 방법을 제안하였다. 학습 단계에서 타일 입력을 사용해 소형 객체의 상대 크기와 세부 특징을 보존하고 추론 단계에서는 배경 차분을 통해 활성 타일만 선택적으로 처리하여 연산량을 줄였다.

소형 객체가 많은 장면에서 타일 기반 학습은 병합 기준 전 지표에서 일관된 개선이 확인되었으며, 특히 Precision이 22.2%, Micro F1이 19.6% 상승하였다. 또한, 배경 차분 게이팅을 적용한 결과, 평균 활성 타일 비율이 0.256일 때 프레임당 지연이 60.1ms에서 30.8ms로 감소하고, FPS가 16에서 30으로 향상되었다. 이는 움직임이 있는 부분에만 계산을 집중하면 성능은 유지하면서 추론 비용을 줄일 수 있음을 보여준다.

소형 객체 중심인 원거리 촬영 장면에서는 타일 기반 학습과 배경 차분 게이팅 추론이 유리하다. 반면 대형 객체 비중이 높거나 전체 맥락 보존이 중요한 장면에서는 타일링과 게이팅의 부가 비용에 비해 이득이 작아 전체 이미지 기반 추론이 정확도와 속도 모두에서 더 효율적이었다. 따라서 객체 크기 분포와 활성 타일 비율과 같은 통계를 이용해, 활성 타일 비율이 높아지는 구간에서는 전체 입력 추론으로 전환하고 그렇지 않은 구간에서는 게이팅을 유지하는 동적 전환 전략이 바람직하다.

본 방법은 움직임을 탐지의 전체 조건으로 설정했기 때문에 정지 상태의 위험 대상에는 취약하다. 이를 보완하기 위해 정적 관심 영역에 대한 상시 추론과 일정 간격마다 전체 프레임을 게이팅 없이 처리하는 주기적 전수 추론을 도입할 수 있다. 또한 기상 조건과 카메라 설치 각도 등 환경 변화에 대한 일반화 검증이 아직 제한적이므로, 다양한 현장과 기상 조건을 포함한 실험을 확대하고 데이터 규모와 클래스 구성을 확장하여 적용 범위를 넓힐 계획이다.

ACKNOWLEDGMENT

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음 (과제번호 RS-2020-KA156208).

참고 문헌

- [1] 고용노동부, "2025. 2분기(누적) 산업재해 현황," 2025.09.03. https://www.moel.go.kr/policy/policydata/view.do?bbs_seq=20250900287
- [2] 박용석, 이송연, 이경택, "스마트건설 현장에서 개인 보호장비 검출의 개선 방법," 한국통신학회논문지, vol. 45, no. 12, pp. 2202-2209, 2020. (10.7840/kics.2020.45.12.2202)
- [3] 김광수, 김보완, "학습 영상과 운영 영상의 크기가 다를 때 인식을 저하 개선 방법," 통신 정보 합동 학술 대회 (JCCI) 2025, pp.475-476
- [4] 김영민, 이지영, 윤일로, 한택진, 김철연, "배경 차분과 CNN 기반의 CCTV 객체 검출," 정보과학회 컴퓨팅의 실제 논문지, vol. 24, no. 3, pp. 151-156, 2018. (10.5626/KTCP.2018.24.3.151)