

히트맵 기반 적응형 이미지 슬라이싱을 통한 소형 객체 탐지 : 건설현장 적용 사례

한일주^{1,3}, 장다연^{2,3}, 김광수³, 김봉완^{3*}

¹충남대학교, ²성결대학교, ³한국전자통신연구원 에어모빌리티연구본부
martellato41@o.cnu.ac.kr, clavexn@gmail.com, {enoch, kimbw}@etri.re.kr

Heatmap-Based Adaptive Image Slicing for Small Object Detection : Application to Construction Sites

Il-Ju Hahn, DaYeon Jang, Kwangsoo Kim, Bong Wan Kim*

ChungNam Nat'l Univ., Sungkyul Univ.,

*Electronics and Telecommunications Research Institute (ETRI)

요약

건설현장의 소형 객체는 저해상도와 잡음으로 탐지가 어렵다. 기존 소형 객체 탐지를 위한 방법론은 고정 패치 크기로 긴 추론 시간과 중복 탐지를 유발한다. 본 연구에서는 히트맵 기반 적응형 슬라이싱 프레임워크를 제안한다. CenterNet 기반의 경량 히트맵 네트워크로 객체 중심점 확률 맵을 생성하고, 이를 기반으로 소형 객체 밀집 영역은 정밀 추출과 동시에 대형 객체 영역은 포괄적으로 처리하는 적응형 슬라이싱 전략을 제안한다. 실제 건설현장 데이터에 대하여 소형 객체 탐지 성능 향상이 mAP@0.5:0.95(small) 지표 기준 79% 있었으며, 기존 방법론 대비 적은 슬라이싱 수로 추론 시간을 31% 줄여 정확성과 효율성의 균형을 달성하였다.

I. 서론

최근 컴퓨터 비전 분야의 객체 탐지 기술은 자율주행, 안전 관리, 산업 자동화 등 다양한 영역에서 핵심 기술로 활용되고 있다. 특히 건설현장은 안전사고 예방 측면에서 객체 탐지 기술의 활용 가능성이 매우 높다. 그러나 건설현장의 영상 데이터는 넓은 촬영 범위와 다양한 크기의 객체가 공존하는 특수성을 갖는다. 이 중 소형 객체는 저해상도, 주변 잡음 등으로 인해 탐지가 더욱 어렵다.

기존 연구에서는 소형 객체 탐지를 위해 슬라이딩 윈도우 이미지 슬라이싱 방법론이 사용되었다. Slicing Aided Hyper Inference(SAHI, Akyon et al., 2022)와 같은 정적 슬라이싱 방식은 고정 패치 크기와 겹침으로 인해 긴 추론 시간, 객체 잘림, 중복 탐지 문제를 발생시켰다. 따라서 크기 스케일이 이질적인 객체가 공존하는 대규모 장면에서는 소형 객체 탐지 성능 개선에 한계가 있다.

본 논문에서는 이러한 문제 해결을 위해 '히트맵 기반 적응형 슬라이싱' 프레임워크를 제안한다. CenterNet(Zhou et al., 2019)에서 제안된 방식에 착안하여, 객체의 중심점을 확률적으로 표현하는 히트맵을 생성하고, 이를 기반으로 소형 객체 밀집 영역은 정밀 추출하되 대형 객체 영역은 불필요한 절단 없이 포괄적으로 처리하는 적응형 분할 전략을 사용한다.

본 연구의 기여는 다음과 같다. 첫째, 히트맵 기반 적응형 슬라이싱으로 소형 객체 탐지 성능을 향상시켰다. 둘째, 경량 히트맵 생성 모델로 기존 정적 슬라이싱 방식 대비 효율성을 확보하였다. 셋째, 건설현장 데이터 실험을 통해 산업적 실효성을 입증하였다.

II. 본론

본 논문에서 제안하는 히트맵 기반 적응형 슬라이싱 프레임워크는 두 단계의 탐지 과정으로 구성된다. 첫 번째 단계에서는 입력 이미지에 대해 경

량화된 히트맵 생성 네트워크(HmNet)를 적용하여 객체 중심점의 확률 분포 맵을 생성한다. 생성된 히트맵은 적응형 분할기에 전달되어 객체가 있을 법한 영역을 동적으로 식별하고, 이를 기반으로 이미지를 최적 크기 패치들로 분할한다. 두 번째 단계에서는 분할된 패치와 원본 이미지를 함께 객체 탐지 모델에 입력하여 각각의 탐지 결과를 얻는다. 중복 탐지 결과에 대한 후처리 과정을 통해 최종적으로 통합된 탐지 결과를 산출한다.

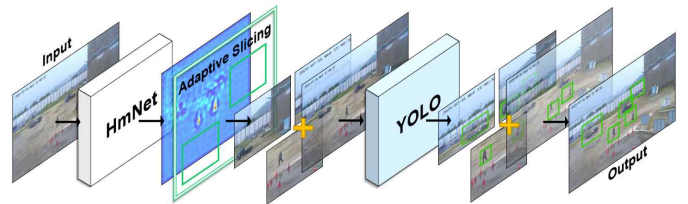


그림 1 히트맵 기반 적응형 슬라이싱 프레임워크

학습 과정에서는 HmNet은 원본 이미지를 대상으로 학습되며, 객체 탐지 모델은 원본 이미지로만 학습한 “기본학습” 방법과 HmNet에서 생성된 분할 패치와 원본 이미지를 모두 활용한 “분할학습” 방법으로 나뉜다. 분할학습을 통해 확대된 소형 객체를 모델이 더 상세하게 학습할 수 있으며 성능 향상을 기대할 수 있다. 추론 단계에서는 객체 탐지 모델의 학습 방식과 무관하게 HmNet이 생성한 분할 패치와 원본 이미지를 동시에 추론하고 그 결과를 병합하는 구조를 적용할 수 있다.

2.1 히트맵 생성 네트워크(HmNet)

HmNet은 객체 탐지 모델 중 하나인 CenterNet의 히트맵 생성 인코더-디코더 구조를 차용하였으며, 실시간 처리 성능을 고려하여 ResNet-18을 백본 네트워크로 채택하였다. 각 객체의 바운딩 박스 중심점을 2D 가우시안 분포로 변환하여 정답(ground truth) 히트맵을 생성하였으며 가우시안 커널의 표준편차는 객체 바운딩 박스 크기에 비례하도록 설정하여 소형

객체는 더 좁은 분포를, 대형 객체는 더 넓은 분포를 갖도록 하였다. HmNet 학습 시 인코더 단계에서는 ResNet-18을 통해 입력 이미지로부터 특징 맵을 추출하고, 디코더 단계에서는 Deformable Convolution을 통해 업샘플링하여 최종적으로 128×128 크기의 히트맵을 출력한다.

2.2 히트맵 기반 적응형 슬라이싱(Adaptive Slicing)

적응형 분할기는 히트맵에서 검출된 피크들을 그룹 단위로 통합한 후 이들을 포함하는 패치를 슬라이싱하는 방식으로 동작한다. 먼저 히트맵에서 상위 k 개의 피크를 검출한 뒤, 각 피크의 주변 평균 값과 비교하여 충분히 두드러진 피크만을 선택한다. 선택된 피크들은 Union-Find 알고리즘을 통해 그룹화된다. 또한 그룹 경계에 위치한 피크들의 Gaussian blob 크기를 추정하여 이에 비례하는 패딩을 적용함으로써, 객체의 크기를 반영한 패치 크기를 결정한다. 최종적으로 히트맵 상의 패치 좌표를 이미지의 패치 좌표로 변환하여 슬라이싱한다. 이러한 방식은 기존의 패치 크기가 고정된 슬라이싱 방식과 달리 객체의 예상 분포에 따라 패치 크기와 위치를 동적으로 조정함으로써, 대형 객체에 대해서는 불필요한 분할을 줄이고 소형 객체에 대해서는 탐지 모델이 해당 부분을 확대하여 학습·추론할 수 있다.

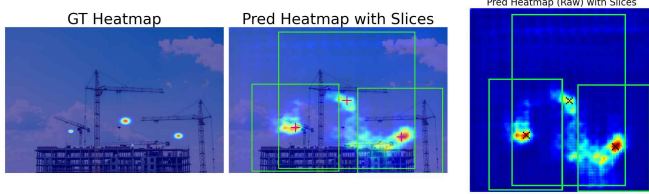


그림 2 적응형 슬라이싱

2.3 객체 탐지 및 결과 병합

하나의 이미지에 대해 추출된 패치들과 원본 이미지는 각각 객체 탐지 모델에 입력되어 추론을 수행한다. 이 과정에서 동일 객체가 여러 패치에서 중복 탐지되거나, 원본 이미지와 패치에서 동시에 탐지될 수 있다. 이러한 중복 탐지 결과를 병합하기 위해서 Non-Maximum Suppression (NMS) 후처리 방법을 사용하여 Intersection over Union(IoU) 임계값 기반으로 중복된 탐지 박스를 제거하였다. 추가적으로 Weighted Boxes Fusion(WBF)을 적용하여 중복된 탐지 결과들의 신뢰도를 가중 평균하여 더 정확한 최종 탐지 박스를 생성하였다. 실험 결과 WBF가 NMS보다 소형 객체에 대해 높은 정확도를 보였다. 최종적으로 병합된 탐지 결과는 원본 이미지 좌표계로 변환되어 출력된다.

2.4 실험 설계 및 결과

본 연구에서 사용한 건설 현장 이미지 데이터는 작업자, 비 작업자, 안전 헬멧, 안전 조끼뿐만 아니라 굴착기, 덤프트럭, 콘크리트 믹서, 지게차, 타워크레인 등 다양한 중장비를 포함한 총 29개 클래스로 구성되어 있다. 데이터의 다양성을 확보하기 위해 세 가지 경로에서 이미지를 수집하였다.

i) 실제 건설 현장 CCTV 영상(2025년 7월 ~ 2025년 8월 수집). ii) AI-HUB의 “건설 현장 위험 상태 판단 데이터”. iii) 이미지 크롤링(예: “dump truck at construction site” 검색어 기반 이미지 검색 결과 활용) 등을 통해 다양한 중장비를 포함한 건설 현장 이미지를 확보하였다. 결과적으로 총 5,883개의 객체를 포함하는 1,511장의 이미지 데이터를 수집하였으며, 188×188부터 최대 5,600×8,400 크기의 다양한 해상도로 구성되어 있다. Train, Validation, Test 데이터를 각각 7:2:1로 분할 하였으며, 객체 크기에 따른 정량적 성능 검증을 위해 COCOeval 프로토콜을 적용하였다. 객체 탐지 모델로는 YOLO11n 모델을 사용하였다. 성능지표의 경우 객체 탐지 성능 평가의 표준지표인 mAP 계열 지표와 F1 score를, 추론 효율성

측정을 위한 지표인 FPS를 사용하였다.

표 1 COCOeval 프로토콜 크기(pixel²)별 객체 수

small : $< 32^2$ px	medium : $32^2 \leq \text{area} < 96^2$ px	large : $\geq 96^2$ px
811	1729	3343

제안 방법과의 성능 비교를 위해 비교 대상을 다음과 같이 설정하였다. 학습에는 YOLO11n을 사용하였고, 추론에는 SAHI를 적용하였다. 이때 SAHI는 512×512 패치 크기를 기준으로, 패치 간 20% 겹침을 허용하는 기본 설정값을 사용하였다. 한편, 제안 방법은 패치와 원본 이미지를 합쳐 총 5,823개의 이미지를 활용하여 “분할학습”을 수행하였고, 추론 단계에서는 히트맵 기반 적응형 슬라이싱을 활용하였다.

표 2 비교 실험 결과

	YOLO11n + SAHI	Proposed
mAP@0.5	0.702	0.572
mAP@0.5:0.95 small	0.217	0.389
mAP@0.5:0.95 medium	0.255	0.288
mAP@0.5:0.95 large	0.570	0.421
Macro F1	0.452	0.506
Micro F1	0.628	0.714
평균 FPS	5.2	7.5
추론 시 생성 패치 수	1564	1018

비교 실험 결과, 제안 방법은 소형 객체 탐지에 있어 SAHI 대비 mAP@0.5:0.95가 약 79% 향상되어 0.217에서 0.389로 개선되었다. 추론 효율성 측면에서도 기존 방법 대비 44%의 FPS 향상을 보였다. 전반적인 mAP 성능은 0.702에서 0.572로 약 19% 감소하였으나, 실시간 추론이 가능한 수준의 효율성을 달성했다는 점과 더불어 소형 객체 탐지에서 가장 좋은 성능을 보였다는 점에서 실제 현장에서의 적용 가능성을 기대할 수 있다.

III. 결론

본 연구에서는 건설 현장과 같이 다양한 크기의 객체가 혼재된 환경에서 기존 방법의 한계를 보완하기 위해 히트맵 기반 적응형 슬라이싱 프레임워크를 제안하였다. 제안된 방법은 경량 히트맵 네트워크를 통해 객체 중심 분포를 추정하고 이를 기반으로 관심 영역을 동적으로 분할한 뒤, 원본 이미지와 패치 이미지를 함께 탐지하여 결과를 통합하는 방식으로 소형 객체 탐지 성능과 추론 효율성을 동시에 확보하였다. 실제 건설 현장 데이터셋을 이용한 실험 결과, 제안 방법은 비교 방법론에 비해 소형 객체에서 높은 mAP를 달성하였다. 또한 불필요한 패치 생성을 줄여 추론 속도를 개선함으로써 효율성도 동시에 개선하였다. 향후 연구에서는 더욱 정교한 관심 영역 제안과 다양한 도메인 데이터셋에 대한 일반화 검증을 통해 실시간 응용 가능성을 넓힐 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

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

참 고 문 헌

- [1] Akyon F. C., Altinuc S. O., and Temizel A., “Slicing Aided Hyper Inference and Fine-tuning for Small Object Detection,” 2022 IEEE International Conference on Image Processing (ICIP), 2022.
- [2] Zhou X., Wang D., and Krähenbühl P., “Objects as Points,” arXiv preprint arXiv:1904.07850, 2019.