X-ray 이미지 기반 위험물품 검출을 위한 딥러닝 모델성능에 관한 연구

김민종*, 채문식

한국원자력연구원

*kmj21@kaeri.re.kr, cmswill@kaeri.re.kr

A Study on Deep Learning Model Performance for Dangerous Object Detection in X-ray Images

Minjong Kim*, Moonsik Chae

Korea Atomic Energy Research Institute

* Corresponding author

요 약

본 연구는 X-ray 이미지 기반 위험 물품 검출을 위한 딥러닝 모델의 성능을 분석하였다. OPIXray, HIXray, XAD 데이터 세트를 활용하여 YOLOV10과 다양한 백본 구조의 Faster R-CNN을 비교하였다. PASCAL VOC Challenge 평가 지표로 성능을 측정한 결과, YOLOV10이 모든 데이터 세트에서 높은 정확도를 나타냈으며. 이는 X-ray 이미지의 특성인 객체의 중첩과 불명확한 경계 등에 대한 YOLOV10의 효과적인 처리 능력을 보여준다. Faster R-CNN 중에서는 ConvNext V2(tiny)가 정확도가 높았으며, CNN 구조가 X-ray 이미지의 지역적 특징 포착에 더 적합함을 알 수 있다. 또한, 경량화된 모델 비교를 통해 X-ray 이미지 특성상 모델 복잡도 증가가 반드시 성능 향상으로 이어지지 않음을 확인하였다. 이 연구를 통해 X-ray 이미지 기반 위험 물품 검출에서의 다양한 최신 딥러닝 모델 결과를 분석하였고, 향후 X-ray 특화 모델 개발 및 데이터 세트 확장을 위해 활용되기를 기대한다.

1. 서 론

현대 사회에서 테러 위협의 증가, 국제 교역량의 증가로 인해 불법 물품 반입 시도가 늘고 있어 보안의 중요성이 점점 증가하며, 특히 X-ray 검색 기술의 중요성이 점점 커지고 있다. X-ray 기반의 위험 물품 분류 기술은 공항, 항만, 국경 등에서 불법 물품이나 위험 물질을 탐지하는 데 핵심적인 임무를 수행하고 있기 때문이다. 최근에는 인공지능의 딥러닝을 이용한 객체검출에 관한 연구가 진행 중이나, 현재 AI 기술의 완성도는 아직 시장의 요구 수준에 미치지 못하고 있다. 검사원들의 판단을 완벽하게 대체하기에는 기술적 한계가 존재하며, 이로 인해 완전 자동화에는 어려움이 따른다. AI 기술을 이용한 완전 자동화를 위해서는 아직 해결해야 할 과제들이 남아있다. 첫 번째는 딥러닝 모델 학습 데이터 확보의 한계점이다. X-ray 객체검출 분야의 연구가 전 세계적으로 활발히 진행되고 있음에도 불구하고, 기술 유출에 대한 우려로 인해 데이터 공유에 제약이 있다. 두 번째로는 X-ray 이미지 특징에 맞는 딥러닝 모델 개발이다. X-ray 이미지는 가시광선 이미지와는 다른 특성이 있고, 이에 적합한 딥러닝 모델의 개발이 필요하다. X-ray 이미지는 대부분 그레이스케일로 표현되어 컬러 정보가 부재하므로, 일반적인 이미지 인식 태스크에서 많이 사용되는 컬러 기반 특징(color-based features)을 활용하기에는 한계가 있다. 본 논문에서는 X-ray 이미지 기반 위험 물품 검출 기술의 자동화 수준을 높이기 위해, 공개된 다양한 X-ray 데이터 세트에서의 딥러닝 모델성능을 분석하였다.

2. X-ray 데이터 세트 구성

본 연구에서는 중국의 베이항 대학교 (Beihang University)의 왕텐보(Tianbowang) 교수 연구팀이 제공한 공개 X-ray 이미지 데이터 세트를 활용하였다. 해당 데이터 세트는 학습용과 평가용으로 나뉘어 있으며, X-ray 이미지에서의 위해 물품에 대한 경계박스(Bounding Box) 좌표가 표기되어 있다. 데이터 세트에는 단일 물체뿐만 아니라 중첩된 물체, 부분적으로 가려진 물체 등 검출이 어려운 상황도 포함되어 있어, 모델의 강건성을 평가할 수 있다. 딥러닝 모델 학습을 위해 각 데이터 세트를 JSON 파일 형식으로 변환하였으며, 이 과정에서 OPIXray[1], HIXray[2], XAD[3] 데이터 세트를 활용하였다. OPIXray 데이터 세트는 총 8,885로 오클루전(occlusion) 레벨에 따른 영향을 연구하기 위해 테스트 세트가 나뉘어 있다. HIXray 데이터 세트는 고품질의 X-ray 이미지 45,364개로 구성되어 있으며, 총 102,928개의 일반적인 금지 물품을 포함하고 있다. XAD 데이터 세트는 총 5,587개의 X-ray 이미지로 구성되어 있으며, 금지 물품이 4개의 카테고리로 분류되어 있다.

3. X-ray 이미지 기반 딥러닝 모델 성능비교

객체검출 기법은 일반적으로 1단계와 2단계로 분류된다. 1단계 방식은 속도가 빠르지만, 정확도가 상대적으로 낮고, 2단계 방식은 정확도는 높지만, 속도가 느린 특징을 갖고 있다. 본 연구에서는 1단계 검출 기법 중 최신 모델인 YOLOV10과 2단계 검출 기법 중 대표적인 Faster R-CNN[4]을 선정하였다. YOLO(You Only Look Once)는 이미지를 그리드로 나누고 각 그리드 셀에서 경계상자와 클래스 확률을 예측하는 방식으로 동작한다. X-ray 이미지는 객체의 내부 구조까지 투과하는 성질을 갖고 있으므로, 객체의 경계 포착이 중요하며, YOLO 방식이 이미지를 그리드로 나누어 국소적으로 특징을 추출하는데 적합할 것으로 판단하였다. Faster R-CNN의 경우, 다양한 backbone 구조를 적용하며 각 딥러닝 아키텍처별로 X-ray 이미지에 대한 객체검출 성능을 비교 분석하였다. 이 알고리즘은 크게 특징추출(Feature Extraction), 영역제안 네트워크(Region Proposal Network, RPN), 그리고 멀티태스크학습(Multi-task Learning)의 세 단계로 구성된다. 대표적인 CNN 구조인 ResNet50과 ResNet101, 최신 Transformer 기반 모델인 Swin TransformerV2, 그리고 최적화된

CNN 구조인 ConvNextV2의 성능을 비교하였다. 또한, 경량화 모델인 MobileNetV3와 MobileOne을 통해 정확도와 속도 간의 trade-off를 검토하였다. 이러한 비교 분석을 통해 X-ray 이미지 분석에 가장 적합한 딥러닝 모델을 분석하였다.

4. X-ray 이미지 기반 객체 검출 알고리즘의 성능 비교 및 분석

X-ray 이미지 기반 객체 검출 알고리즘의 성능 평가를 위해 컴퓨터 비전 분야에서 널리 사용되는 PASCAL VOC Challenge의 평가 방식을 채택하였다. 실험 결과에 따르면, 본 연구에서 사용된 X-ray 이미지 데이터 세트에서는 1단계 검출 방식인 YOLOV10이 가장 우수한 성능을 보였다. 이는 일반적인 컴퓨터 비전 분야에서 2단계 검출 방식이 더 높은 성능을 나타내는 경향과는 대조적인 결과로, X-ray 이미지 기반 객체 검출의 특수성을 나타내고 있다. 2단계 검출 방식인 Faster R-CNN 기반 모델 중에서는 ConvNextV2가 backbone으로 사용되었을 때 가장 우수한 성능을 보였다. X-ray 이미지는 객체의 특징이 겹치거나 객체를 나타내는 이미지의 특징 채널이 부족한 경향이 있다. 따라서 해당 데이터 세트에서는 지역적인 특징을 효과적으로 포착하는 CNN 모델이 Transformer 모델보다 더 적합한 것으로 나타났다. ResNet50과 ResNet101은 모델의 크기와 관계없이 유사한 성능을 나타내었다. 이는 모델의 복잡도를 높이는 것이 성능 향상으로 직결되지 않음을 시사한다. 최신의 CNN 모델 중 하나로, 최적화된 구조로 되어 있기 때문으로 해석할 수 있다. 또한, 경량화된 모델인 MobileOne과 MobileNetV3는 ResNet 기반 모델과 유사한 성능을 보였다. 이는 X-ray 이미지의 특성상 모델의 크기와 복잡도를 줄여도 객체 검출 성능이 크게 저하되지 않음을 나타낸다.

5. 결론

본 논문에서는 X-ray 이미지 기반 위험물품 검출을 위한 다양한 딥러닝 모델의 성능을 비교 분석하였다. 실험 결과, 1단계 검출 방식인 YOLO V10이 가장 우수한 성능을 보였으며, 이는 X-ray 이미지 기반 객체검출의 특수성을 보여준다. 2단계 검출 방식 중에서는 ConvNextV2가 가장 좋은 성능을 나타냈다. 모델의 복잡도 증가가 반드시 성능 향상으로 이어지지 않음을 발견하였다. 또한, 경량화 모델도 상대적으로 좋은 성능을 보여 실제 응용 가능성을 보여주었다. 향후 연구에서는 이러한 발견을 바탕으로 X-ray 이미지에 특화된 모델 개발과 데이터 세트 확장을 통해 성능을 더욱 개선할 예정이다.

표1. Faster R-CNN 기반 및 YOLO 딥러닝 모델의 X-ray 이미지 데이터 세트별 mAP(mean average precision) 성능 비교.

Faster R-CNN 기반 모델			
Methods	OPIXray	HiXray	XADXray
MobileOne(S0)[6]	0.358	0.437	0.558
Swin transformerV2[7]	0.347	0.437	0.577
ResNet50[8]	0.369	0.441	0.542
ResNet101[8]	0.368	0.432	0.549
MobileNetV3(50)[9]	0.315	0.423	0.51
ConvNextV2(tiny)[10]	0.406	0.456	0.597
YOLO 기반 모델			
YOLOV10	**0.411**	**0.55**	**0.646**

ACKNOWLEDGMENT

본 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 연구개발특구진흥재단의 지원을 받아 수행된 연구임. (2024-DD-RD-0367)

참 고 문 헌

[1] Wei, Y., Tao, R., Wu, Z., Ma, Y., Zhang, L., & Liu, X. (2020, October). Occluded prohibited items detection: An x-ray security inspection benchmark and de-occlusion attention module. In Proceedings of the 28th ACM international conference on multimedia (pp. 138-146).

[2] Tao, R., Wei, Y., Jiang, X., Li, H., Qin, H., Wang, J., ... & Liu, X. (2021). Towards real-world X-ray security inspection: A high-quality benchmark and lateral inhibition module for prohibited items detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10923-10932).

[3] Liu, A., Guo, J., Wang, J., Liang, S., Tao, R., Zhou, W., ... & Tao, D. (2023). {X-Adv}: Physical adversarial object attacks against x-ray prohibited item detection. In 32nd USENIX Security Symposium (USENIX Security 23) (pp. 3781-3798).

[4] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks

[5] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458.

[6] Vasu, P. K. A., Gabriel, J., Zhu, J., Tuzel, O., & Ranjan, A. (2023). Mobileone: An improved one millisecond mobile backbone. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7907-7917).

[7] Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., ... & Guo, B. (2022). Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12009-12019).

[8] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[9] Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for mobilenetv3. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 1314-1324).

[10] Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., & Xie, S. (2023). Convnext v2: Co-designing and scaling convnets with masked autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 16133-16142).