

비전 LLM을 활용한 차세대 X-ray 보안 검색 시스템 개발

김민중^{1*}, 박형달²

¹한국원자력연구원, ²(주)에이리스

*kmj21@kaeri.re.kr

Development of Next-Generation X-ray Security Screening System Using Vision LLM

Minjong Kim^{1*}, Hyeongdal Park²

¹Korea Atomic Energy Research Institute, ²AIRISS CO., Ltd.

* Corresponding author

요약

현재의 X-ray 보안 검색 시스템은 검사원의 전문성과 폐쇄형 집합 기반 객체 탐지 모델에 크게 의존하고 있으며, 이들은 복잡한 은닉 시나리오와 새로운 형태의 위협에 대응하는 데 한계를 보인다. 본 논문은 X-ray 보안 이미지를 이해하고 추론할 수 있는 비전-언어 모델(VLM) 학습을 위한 고품질 데이터셋 구축 프레임워크를 제안한다. 제안하는 프레임워크는 세 가지 핵심 구성요소로 이루어진다: (1) 다양한 재질의 X선 투과 특성을 시뮬레이션하여 위험물품의 사실적인 X-ray 이미지를 생성하는 이미지 생성기, (2) 실제 X-ray 영상 물리학을 반영한 덧셈 합성을 통해 복잡한 은닉 시나리오를 구현하는 합성 모듈, (3) 구조화된 설명과 단계별 논리적 추론 과정을 생성하는 다층적 캡션 및 추론 체인 생성기이다. 생성된 데이터셋은 이미지-캡션 쌍, 추론 체인, 종합적인 합성 파라미터를 포함하는 구조화된 JSON 메타데이터로 저장된다. 본 프레임워크는 현재 AI 기반 보안 시스템의 중요한 한계를 해결하여, VLM이 단순한 패턴 인식을 넘어 맥락적 이해와 논리적 추론으로 나아갈 수 있도록 한다. 이러한 접근법은 변화하는 위협과 정교한 은닉 기법에 대응할 수 있는 차세대 적응형 보안 검색 시스템 개발의 기반을 제공한다.

I. 서론

X-ray 보안 검색은 테러 및 불법 물품 반입을 막는 국가 안보의 최전선에 있지만, 그 판독 과정은 여전히 검사원의 높은 숙련도와 집중력에 크게 의존한다. 특히 수하물 내에 여러 물품이 복잡하게 얹혀 있거나 위험물품이 의도적으로 다른 물체 뒤에 숨겨진 경우, 판독의 난이도는 급격히 증가하며 이는 인적 오류의 주요 원인이 된다. 이러한 문제를 해결하기 위해 인공지능(AI) 기반 자동 검출 시스템이 도입되고 있으나 [1-2], 현재의 기술은 대부분 사전 정의된 위험물품의 시각적 패턴을 학습하여 식별하는 폐쇄형 집합(closed-set) 패러다임에 머물러 있다. 이 방식은 학습 데이터에 포함된 특정 형태와 구도의 물체는 잘 찾아내지만, 다음과 같은 근본적인 한계를 가진다. 기존 모델은 '칼'의 형태는 인식할 수 있지만, '금속 케이스 안에 들어있는 칼날'과 같이 객체 간의 공간적, 기능적 관계를 이해하지 못한다 [3]. 이는 정교한 은닉 시도에 매우 취약하여 맥락적 이해의 부재가 있다. 추론 능력이 부족하여 3D 프린팅 총기나 새로운 형태의 폭발물처럼 학습 데이터에 없던 위협이 등장했을 때 대응하기 어렵다. X-ray 스캐너의 종류나 설정에 따라 이미지의 특성이 달라질 때 성능이 저하되는 문제가 있으며, 이는 모델이 객체의 본질이 아닌 특정 이미지 스타일에 과적합되기 때문이다. 이러한 한계를 극복하기 위해, 본 연구는 단순히 이미지를 '인식'하는 것을 넘어 '이해'하고 '추론'하는 비전-언어 모델(VLM)의 도입을 제안한다. VLM은 이미지 인코더와 대규모 언어 모델(LLM)을 결합하여, 이미지 내 시각적 정보와 언어적 설명을 연계하여 학습한다. 이에 본 연구는 VLM 훈련의 핵심인 고품질 데이터셋 구축에 초점을 맞춘다. 특히, 실제와 유사한 은닉 시나리오를 반영한 합성 X-ray 이미지와, 모델의 추론 능력을 극대화하는 상세한 설명 캡션(caption)을 생

성하는 프레임워크를 제안하고자 한다. 이를 통해 차세대 지능형 X-ray 보안 검색 시스템 개발의 초석을 마련하는 것을 목표로 한다.

II. 본론

본 연구에서 제안하는 X-ray 비전-언어 모델(VLM) 데이터셋 구축 프레임워크는 VLM이 X-ray 이미지의 복잡한 맥락을 효과적으로 학습하는 데 필요한 고품질의 이미지-캡션 쌍을 체계적으로 생성하는 것을 목표로 한다. 프레임워크는 그림1에 나타난 것처럼 (1) 개별 위험물품 이미지 생성, (2) 복합 은닉 시나리오 합성, (3) 멀티모달 캡션 및 추론 데이터 생성의 세 가지 핵심 단계로 구성된다. 이 모든 과정은 통합된 사용자 인터페이스(UI)를 통해 제어되며, 연구자가 다양한 조건의 데이터셋을 효율적으로 구축할 수 있도록 지원한다.

2.1 개별 위험물품 이미지 생성

개별 위험물품 이미지 생성은 X-ray 투과 특성을 정확히 모사하는 것에서 시작된다. 실제 X-ray 스캐너에서 물체의 밀도와 두께에 따라 X선의 투과율이 달라지는 원리를 그레이스케일 이미지로 재현하기 위해, 각 물체를 구성하는 재질별로 서로 다른 픽셀 값을 할당하는 방식을 채택하였다. 칼, 총기, 가위, 망치 등 주요 위험물품의 기본 형태를 다각형과 사각형 같은 기초 도형의 조합으로 구현하였으며, 이는 실제 X-ray 이미지에서 관찰되는 물체의 단순화된 실루엣을 효과적으로 표현한다. 밀도 모사 과정에서는 객체의 각 부분에 재질 특성을 반영한 그레이스케일 값을 세심하게 할당하였다. 예를 들어, 칼의 경우 금속 재질의 날 부분은 180-220 범위의 높은 픽셀 값을 부여하여 X선이 투과하기 어려운 고밀도 영역임

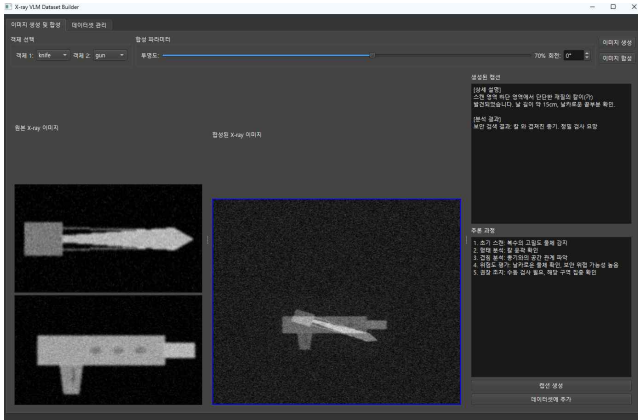


그림 1. 비전 LLM을 위한 위험물품 X-Ray 데이터셋 생성 시스템

을 표현하였고, 플라스틱이나 목재로 이루어진 손잡이 부분은 120-150 범위의 상대적으로 낮은 값을 할당하여 X선이 더 쉽게 투과하는 저밀도 영역임을 나타내었다.

2.2 복합 은닉 시나리오 합성

개별적으로 생성된 위험물품 이미지들은 실제 수하물 검색 상황에서 발생하는 복잡한 은닉 및 중첩 시나리오를 재현하기 위해 하나의 통합된 이미지로 합성된다. 이 과정은 단순한 이미지 중첩을 넘어서 X-ray 영상의 고유한 물리적 특성을 반영하도록 설계되었다. X-ray 이미지에서는 물체가 겹쳐질 때의 영역에 대한 특징을 살린 처리를 해주었다. 다양한 은닉 환경을 구현하기 위해 시스템은 여러 조작 가능한 파라미터를 제공한다. 회전 기능을 통해 위험물품이 다양한 각도로 배치된 상황을 만들 수 있으며, 투명도 조절 기능으로 위험물품이 다른 물체에 의해 부분적으로 혹은 심하게 가려진 시나리오를 생성할 수 있다. 이러한 파라미터들의 조합은 실제 보안 위협 상황에서 나타날 수 있는 정교한 은닉 기법들을 효과적으로 재현한다. 실제 X-ray 장비의 환경적 특성도 충실히 재현하였다. 센서에서 발생하는 전자적 노이즈를 정규 분포 기반의 무작위 패턴으로 모사하여 이미지 전반에 적용하였으며, 이는 실제 스캐너에서 관찰되는 자연스러운 노이즈 특성을 반영한다. 또한 이미지 대비 조절을 통해 다양한 제조사의 스캐너나 서로 다른 운영 환경에서 획득된 이미지들의 특성 차이를 시물레이션할 수 있도록 하였다. 이러한 변형들은 학습된 모델이 다양한 실제 환경에서도 강건하게 작동할 수 있는 일반화 능력을 갖추도록 돕는다.

2.3 추론 과정 데이터 생성

본 프레임워크의 가장 혁신적인 측면은 VLM이 위협 판단에 이르는 논리적 과정을 학습할 수 있도록 단계별 추론 과정 데이터를 생성하는 것이다. 이는 최근 LLM 분야에서 주목받고 있는 Chain-of-Thought 프롬프팅 기법을 [4] 시각-언어 멀티모달 영역으로 확장한 것으로, 모델이 마치 숙련된 보안 검색 전문가처럼 체계적인 사고 과정을 거쳐 결론에 도달하도록 유도한다. 추론 과정은 5단계의 논리적 흐름으로 구성된다. 첫 번째 단계인 '초기 스캔'에서는 이미지 전체에서 관찰되는 고밀도 물체들의 존재를 확인한다. 두 번째 '형태 분석' 단계에서는 개별 객체의 윤곽과 특징을 식별하며, 세 번째 '겹침 분석' 단계에서는 객체들 간의 공간적 관계와 은닉 패턴을 파악한다. 네 번째 '위험도 평가' 단계에서는 식별된 객체들의 위험성과 은닉 의도를 종합적으로 평가하고, 마지막 '권장 조치' 단계에서는 구체적인 대응 방안을 제시한다. 이러한 추론 데이터는 모델의 판독 결과에 대한 설명 가능성을 확보하는 데 핵심적인 역할을 한다. 기존의 블랙박스 AI 모델들이 가진 한계, 즉 '왜 그런 판단을 내렸는지' 설명할 수 없는

문제를 극복하여, 보안 검색과 같은 고위험 분야에서 요구되는 투명성과 신뢰성을 제공한다. 또한 이러한 단계적 추론 과정의 학습은 모델이 새로운 유형의 위협이나 예상치 못한 은닉 패턴에 직면했을 때도 논리적인 분석 과정을 통해 합리적인 판단을 내릴 수 있는 일반화 능력을 향상시킨다.

III. 결론

본 논문에서는 X-ray 보안 검색을 위한 비전-언어 모델(VLM) 학습용 데이터셋 구축 프레임워크를 제안하였다. 기존의 폐쇄형 집합 기반 객체 탐지 모델이 가진 한계를 극복하기 위해, 실제 X-ray 영상의 물리적 특성을 정확히 모사한 합성 이미지 생성과 함께 다층적 캡션 및 추론 과정 데이터를 체계적으로 생성하는 통합 시스템을 개발하였다. 단순한 패턴 인식을 넘어 맥락적 이해와 논리적 추론이 가능한 차세대 X-ray 보안 검색 시스템 개발의 토대를 제공한다는 점에 있다. 특히 새로운 형태의 위협이나 정교한 은닉 시도에 대응할 수 있는 적응형 AI 시스템 구현에 필수적인 학습 데이터 구축 방법론을 제시하였다. 향후 연구 과제로는 실제 X-ray 이미지와의 도메인 적응 기법 적용, 추론을 위한 데이터생성 모델링을 통한 더욱 사실적인 이미지 생성이 있다. 이후 구축된 데이터셋을 활용한 VLM 모델의 실제 성능 검증과 현장 적용 가능성 평가가 필요하다. 본 프레임워크가 국가 안보 강화와 공공 안전 향상에 기여하는 실용적인 AI 시스템 개발로 이어지기를 기대한다.

ACKNOWLEDGMENT

본 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 연구개발특구진흥재단의 지원을 받아 수행된 연구임. (2025-DD-RD-0367)

참 고 문 헌

- [1] Miao, C., Xie, L., Wan, F., Su, C., Liu, H., Jiao, J., & Ye, Q. (2019). Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2119-2128).
- [2] Liang, K. J., Sigman, J. B., Spell, G. P., Strellis, D., Chang, W., Liu, F., ... & Carin, L. (2019). Toward automatic threat recognition for airport X-ray baggage screening with deep convolutional object detection. arXiv preprint arXiv:1912.06329.
- [3] Gao, C., Killeen, B. D., Hu, Y., Grupp, R. B., Taylor, R. H., Armand, M., & Unberath, M. (2023). Synthetic data accelerates the development of generalizable learning-based algorithms for X-ray image analysis. Nature Machine Intelligence, 5(3), 294-308.
- [4] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35, 24824-24837.