

국방분야 기술도면 분석을 위한 비전-언어 모델 최적 활용 방안 연구

서영진*, 박경준

국방기술품질원, 대구경북과학기술원

mulle0514@dtqa.re.kr, kjp@dgist.ac.kr

Optimal Utilization of VLM for Technical Drawing Analysis in the Defense Field

Young-Jin Seo*, Kyung-Joon Park

Defence Agency for Technology and Quality, Daegu Gyeongbuk Institute of Science&Technology

요 약

본 연구는 국방 분야에서 무기체계 기술 도면을 분석하기 위한 로컬 형태의 멀티모달 LLM(VLM) 모델 성능을 연구하였다. 다양한 로컬 VLM 모델을 대상으로, 도면 데이터셋을 활용하여 특정 목적의 질문을 하고 답변 정확도를 평가하였다. 평가 결과 질문의 의도마다 최적의 답변 성능을 모델이 다름을 확인하였다. 이는 기술 도면을 VLM을 통해 분석할 때, 질문 유형에 따라 모델을 다르게 선택하여 혼합 운용하는 것이 제한된 컴퓨팅 자원을 효과적으로 사용할 수 있음을 의미한다.

I. 서 론

4차 산업혁명 시대를 맞아 국방 분야에서도 데이터 분석과 인공지능 등의 신기술 활용이 활발하다. 한국군은 국방혁신 4.0 기본계획을 통해 AI 기반의 국방 역량 강화를 추진하고 있다[1]. 이러한 흐름에서 무기체계 개발 및 생산 과정에서 작성되는 기술 도면을 AI를 활용하여 이해하고 분석하는 능력이 중요해졌다. 기술 도면에는 구성 품목, 치수, 공차 등 군수품을 생산하는 핵심 정보가 담겨 있어, 이를 AI를 활용하여 많은 양의 도면을 기계가 해석할 수 있다면 국방 분야의 개발·제조 효율 향상과 지식 축적에 큰 역할을 할 수 있다.

기존 연구에서는 객체 검출 알고리즘과 광학문자판독(OCR) 기술을 조합하여 도면의 정보를 추출하였다. 예를 들어, 이미지 내 물체 인식을 위해 YOLO와 같은 딥러닝 기반 탐지 모델을 활용하고, 문자 영역의 치수 표기를 읽기 위해 OCR 엔진을 적용하는 방식이다[2]. 그러나 이와 같은 개별 모델의 조합 방식은 도면 내 복잡한 맥락을 종합적으로 이해하기 어렵고, 작업 공정 또한 복잡하다. 특히 YOLO 기본모델은 기술 도면의 기본 객체 탐지에는 낮은 성능을 보인다[3]. 최근 등장한 멀티모달 대형언어 모델은 하나의 모델에서 이미지와 텍스트를 동시에 처리하여, 객체 인식부터 문자 정보 추출, 도면 해석까지 통합적으로 수행하는 능력을 보이고 있다.

다만, 국방 분야는 보안상 외부 클라우드에 의존한 AI 서비스 활용이 제한되므로, 외부 인터넷 연결 없이 온프레미스 환경에서 운용할 수 있는 로컬 AI 모델이 필요하다. 이에 따라 도면 해석 성능이 우수하면서도 폐쇄망에서 구동 가능한 Vision-Language Model(VLM) 모델을 선정하는 것은 매우 중요하다.

II. 본론

본 연구에서는 멀티모달 AI 기술을 폐쇄망에서 기술 도면 분석에 적용했을 때 가장 최적의 조합을 찾기 위한 실험을 수행하였다. 국방망 환경을 고려하여 인터넷 연결 없이 로컬 PC 또는 서버에 설치·구동할 수 있는 네

가지 VLM 모델을 선정하였으며, 해당 모델들의 도면 분석 성능을 비교하였다. 실험에 사용된 모델은 Llama-4-Maverick-17B-128E, Qwen2.5-VL-32B, Pixtral-12B, C4AI-Aya-Vision-32B의 4종이다. 이들은 모두 오픈소스로 공개된 거대 멀티모달 모델로서, 이미지-텍스트 쌍 학습을 통해 시각 정보를 자연어로 변환하거나, 질의 응답 할 수 있는 능력을 가지고 있다. 실험 데이터로는 품목 및 치수 정보가 포함된 도면 30장과, 도면 이해 능력을 판단할 수 있는 질문 답변 데이터셋 30개를 사용하였다.

질문은 크게 세 가지 유형으로 구성된다: (1) 품목 인식 : 도면에 표현된 구성품이나 부품의 명칭, 개수 등을 식별하는 질의, (2) 치수 정보 파악 : 도면의 주기에서 길이·각도 등의 수치를 정확히 읽어내는 질의, (3) 도면 내용 이해 : 도면 전체를 해석하여 용도나 작동 원리 등 의미를 파악하는 종합적인 질의.

각 모델은 동일한 도면 이미지와 질의를 입력으로 받아 한국어 답변 능력을 평가하였다. 품목 인식 능력은 질문에 대해 정확한 품목 명칭을 언급했는지로 채점하였고, 치수 정보 추출 정확도는 도면의 실제 표시값과 모델이 제시한 값의 일치 여부를 F1 score로 평가하였다. 도면 이해 능력은 모델의 서술형 답변을 검토하여, 의도한 정답과 부합하는지를 평가하여 점수를 부여하였다.

그 결과 각 평가 항목별로 최고 성능을 보인 모델은 다음과 같다. 품목 식별 정확도 측면에서는 Qwen2.5-VL-32B 모델이 가장 높은 정확도를 나타냈다. 한편, Pixtral-12B 모델은 치수 정보를 추출하는 능력이 가장 좋았다. 도면의 전체적인 의미 파악 및 응용 질문에 대해서는 C4AI-Aya-Vision-32B 모델이 가장 뛰어난 이해력을 보였다. 그리고 Llama-4-Maverick-17B-128E 모델은 세부 성능에서는 일부 모델들보다 성능이 낮았지만, 모든 항목에서 비교적 좋은 성능을 보여 특정 기능에 치우치지 않는 강점이 있음을 확인하였다. 이러한 실험 결과는 단일 VLM 모델로 모든 종류의 도면 질의에 최적의 답변을 하기 어렵다는 점을 시사한다. 대신 질의 내용의 유형에 따라 가장 적합한 VLM 모델을 선택하여 사용하는 조합 전략이 성능 향상에 유리함을 확인하였다.

III. 결론

본 연구를 통해 멀티모달 대형언어모델이 국방 기술도면의 자동 이해·분석에 유용하게 활용될 수 있음을 실험적으로 확인하였다. 최신 VLM 모델들은 복잡한 도면에서 다양한 정보를 통합적으로 추출해낼 수 있었지만, 모델마다 특화된 장점이 있음을 확인하였다. 따라서 실제 응용 시에는 질문 또는 작업의 유형에 따라 최적의 모델을 선택하여 혼합 운영하는 방안이 효과적이다. 이러한 모델 최적 적용 전략은 제한된 컴퓨팅 자원을 효율적으로 활용하면서도 전반적인 분석 정확도를 높일 수 있는 현실적인 대안이 될 수 있다.

특히 국방망처럼 폐쇄된 환경에서는 다양한 로컬 AI 모델을 확보하고 운영하는 것이 중요하므로, 본 연구의 결과는 실무적으로 큰 의미를 가진다. 향후에는 더 다양한 도면 자료와 질문 세트를 통해 VLM의 성능을 지속적으로 평가하고, 실제 국방 업무환경에 적용하여 그 효과성을 실증하는 연구를 수행할 수 있을 것으로 기대한다.

참 고 문 헌

- [1] KK Lee, KR Kim "The ROK Army's innovation strategy linked to Defense Innovation 4.0" The Quarterly journal of defense policy studies, Vol. 39, No. 4, pp.65-99, 2023
- [2] Agritania, M Mahira, M Mi'radj Isnaini. "Development of an engineering drawing detection and extraction algorithm for quality inspection using deep neural networks." CIRP 132, pp.135-140, 2025
- [3] An, Ning, et al. "Improved basic elements detection algorithm for bridge engineering design drawings based on YOLOv5." AI EDAM 38, 2024