

한국 민화 데이터 셋 구축을 위한 AI 보조 파이프라인

송민준, 김수민, 김준성, 이흥노
광주과학기술원

paulmjsong@gm.gist.ac.kr, smkim6927@gm.gist.ac.kr, wnstjd123@gm.gist.ac.kr,
heungno@gist.ac.kr

AI-Assisted Dataset Creation Pipeline for Korean Minhwa

Minjun Song, Sumin Kim, Junsung Kim, Heung-No Lee
Gwangju Institute of Science and Technology

요 약

본 논문은 한국 민화와 같이 상징적 모티브가 핵심인 예술을 시각-언어적으로 이해하고 추론하기 위한 데이터 셋 구축 파이프라인을 제안한다. 기존의 미술 관련 데이터 셋과 공공 문화유산 아카이브는 이미지 접근성과 기본 메타데이터 제공에는 유용하나, 작품 내부 요소가 지니는 상징적 의미와 문화적 근거를 정형화된 주석으로 구조화하는 데에는 한계를 보인다. 본 연구는 AI 보조(AI-assisted) 방식을 도입하여 평가용 데이터 셋 구축 과정을 부분적으로 자동화하였다. 구체적으로, 한국민족문화대백과사전과 e 박물관 자료를 대상으로 세 종류의 모델을 활용해 문장을 자동 라벨링하고, 모델 간 라벨이 불일치하는 항목에 한해 사람이 검토하는 Human-in-the-Loop 방식을 적용하였다. 이를 통해 주석 부족으로 인한 수작업 비용을 줄이고, 상징 정보가 포함된 문장을 효율적으로 선별할 수 있었다. 최종적으로 한국 민화의 '시각적 묘사'와 '상징적 해석'을 포함하는 문장을 추출하여 총 174 장의 이미지와 537 개 문장으로 구성된 경량 평가용 시각-언어 데이터 셋을 구축하였다. 본 데이터 셋은 향후 멀티모달 모델의 민화 상징성 이해 능력을 평가하는 데 활용될 수 있으며, 상징 해석 주석이 부족한 도메인에서 AI 기반 의미 추론을 가능하게 하는 최초의 구조적 시도라는 의의를 갖는다.

I. 서 론

한국 민화는 현실 세계를 사실적으로 재현하기보다 삶의 염원과 기원을 상징적 개체로 압축해 전달하는 시각 문화로 이해되어 왔다. 민화의 조형은 원근법의 결여, 비례의 과격, 형태의 자유로운 변형 등 '가시적 현상계의 공간 질서'를 흔드는 표현을 통해 단순한 묘사를 넘어 고유한 세계관과 정서를 드러낸다 [1]. 민화는 삶의 평탄함과 복락을 기원하는 실전적 매개체로 기능해 왔다는 해석이 축적되어 왔으며, 재액을 물리치는 벽사와 복을 불러들이는 길상에 있다 [2].

이러한 상징성은 작품 내부의 요소들이 문화적 규범, 세시풍속, 민간 신앙과 결합하는 과정에서 “왜 그렸는가”라는 질문에 대한 해답을 구성한다. 까치호랑이 민화는 호랑이, 까치, 소나무 등의 도상이 ‘불안과 안녕’, ‘불길과 길상’이라는 대립항을 조직하고, 일상적 안녕을 기원하는 동시에 불운을 물리치려는 서사 구조를 형성한다 [3]. 이때 상징적 해석은 단순한 묘사와 달리 시각 정보에서 관찰되지 않는 잠재적 의미를 추론하는 절차이기 때문에, AI 모델 연구에서 고난도 과제로 분류된다.

그러나 기존 예술 분야 시각-언어 데이터 셋에서는 이러한 상징적 연결 관계가 충분히 다뤄지지 못했다. 한국디자인진흥원의 “디자인 DB”에는 민화 이미지 기반의 자료 접근성이 높이에 기여하였으나, 상징 의미를 체계적으로 서술하고 정형화하는 것을 전제로 설계되어 있지는 않다 [4]. 또한 전국 박물관 유물 정보 오픈 API는 다수 박물관의 유물 목록과 상세 정보를

제공하는 중요한 인프라이지만 키워드 중심 조회나 목록 제공을 넘어 개체가 함축되는 상징 의미 수준의 해석 주석을 표준 형태로 제공하지는 않는다.

관련 AI 연구 역시 주로 데이터 생성 및 증강이나 객체 인식 기반의 설명 생성에 초점을 맞춰왔으며 [5], 민화의 상징 체계를 구조화해 시각-언어적으로 연결하는 주석 설계 및 데이터 구축은 상대적으로 미진하다. 이러한 한계는 비단 민화에만 국한되지 않는다. 서양 미술을 포함한 주요 미술 데이터 셋들 역시 스타일, 장르, 표면적 장면 내용 기술에 머무르는 경우가 많으며, 은유나 상징성을 명시적으로 서술하는 경우는 드물다 [6, 7]. 최근 퍼즐, 은유, 언어유희 성격의 예술을 대상으로 요소 식별부터 상징 추론까지 요구하는 멀티모달 벤치마크가 제안된 바 있는데, 이는 상징 예술의 이해가 별도의 데이터 설계와 과제 정의를 필요로 함을 시사한다 [5]. 즉, 구조화된 해석 주석이 부족한 문화예술 도메인에서 상징성 이해는 기술적으로 미개척 영역이라는 문제가 명확히 존재한다.

이에 본 논문은 상징적 서술을 중심으로 하는 예술 데이터 셋 구축 파이프라인을 제안한다. 제안 파이프라인은 해석 주석 부족 환경을 극복하기 위해 AI 모델의 자동 라벨링과 인간 검수를 결합하여 효율성을 극대화하였다. 구체적으로 세 가지 모델을 앙상블 형태로 활용하여 문장 단위 라벨링을 수행하고, 불일치 항목만 수작업으로 확정하였다. 본 연구의 주요 기여는 다음과 같다. (1) 상징 예술 이해를 위한 효율적인 데이터 셋 생성 파이프라인을 제안한다. (2) 상징 추론 모델의 경량 평가를 위한 한국 민화 시각-언어 데이터 셋을 구축한다.

(3) 해석 주석 부족 민화 도메인에서 상징 해석을 위한 멀티모달 연구 가능성을 제시한다.

II. 데이터 셋

데이터 셋 구축을 위해 한국민족문화대백과사전과 e 박물관 등 공개 문화유산 자원에서 민화 작품 이미지와 설명 텍스트를 수집하였다. 수집 과정에서는 공식 오픈 API 를 우선적으로 활용하였으며, API 로 제공되지 않는 항목에 한해 필요한 경우 웹 스크래핑을 병행하였다. 그 결과 한국민족문화대백과사전에서 114 건, e 뮤지엄에서 827 건의 문서를 확보하였다. 수집된 텍스트는 전처리 단계에서 정규표현식 패턴 제거 및 정규화를 거쳤으며, Python NLTK 라이브러리를 이용해 문장 단위로 분리하여 각 문장을 하나의 주석 단위로 처리하였다.

본 연구의 최종 목적은 모델이 민화의 상징성을 얼마나 정확하게 설명하는지 측정하는 평가 지표를 구축하는 것이며, 이를 위해 데이터 셋에는 ‘시각적 묘사’와 ‘상징적 해석’에 해당하는 문장만을 포함시켰다. 여기서 ‘상징적 해석’에 해당하는 문장은 민화의 시각적 객체가 문화적 규범·민속적 관습·길상 관념 등에 의해 추상적 의미로 해석되는 내용을 포함하는 서술을 의미하며, 단순 묘사나 역사적 배경 설명과는 구별된다. 반면 역사적 배경, 화법, 작품 설명, 메타데이터 등은 모델 평가의 노이즈가 되므로 제거 대상이다. 하지만 이를 ‘기타’로 묶어 ‘시각적 묘사-상징적 해석-기타’라는 3 분류 체계로 처리할 경우 모델이 ‘기타’ 범주의 복잡한 맥락을 ‘해석’으로 혼동하는 문제가 관찰되었다. 범주를 지나치게 세분화할 경우에도 범주 간 미세한 차이로 인해 의사결정이 불안정해지는 문제가 발생하였다.

이러한 분류의 모호성과 모델의 인지적 과부하를 최소화하기 위해, 본 연구는 분류 모델이 환각을 일으키지 않고 필요한 정보를 안정적으로 추출할 수 있도록 5 개의 범주 체계를 새롭게 설계하였다. 구체적인 범주는 (1) 주제 및 구성, (2) 상징 및 의미, (3) 예술적 기법, (4) 역사적 배경, (5) 목록 및 메타데이터이다. 이러한 설계는 모델이 버려야 할 정보를 구체적으로 인지하게 함으로써 데이터의 순도를 높이고 지나친 세분화로 인한 혼란을 방지한다. 최종적으로 (1)번과 (2)번 문장만을 데이터 셋으로 채택함으로써, 모델이 시각적 객체와 상징적 가치를 연결하는 능력만을 정교하게 평가할 수 있도록 하였다.

효율적이고 정확한 라벨링을 위해 서로 다른 구조적 특성을 가진 다음의 세 가지 모델을 앙상블 형태로 활용하였다. (1) GPT-4.1 을 활용한 LLM 추론 능력 기반 분류, (2) text-embedding-ada-002 를 활용한 임베딩 유사도 기반 분류, (3) DeBERTa-v3-base-xnli-multilingual 을 활용한 다국어 NLI 기반 분류이다. 세 모델은 추론 능력, 의미 유사도, 논리적 합의 판단이라는 서로 다른 강점을 가지므로 조합이 유의미하며, 이질적 판단 기준을 통한 교차 검증 효과가 발생한다. 세 모델의 예측 라벨이 모두 동일한 경우, 해당 라벨을 확정하였다. 이 중 ‘상징 및 의미’ 또는 ‘주제 및 구성’으로 분류된 문장은 최종 데이터 셋에 포함하였으며, 분석 목적과 거리가 먼 ‘예술적 기법’, ‘역사적 배경’, ‘목록 및 메타데이터’ 문장은 제외하였다. 세 모델의 예측이 불일치한 문장에 한해서만 사람이 직접 개입하여 최종 라벨을 확정하는 Human-in-the-Loop 방식을 통해 비용 효율성과 데이터 품질을 동시에 확보하였다.

최종 구축된 데이터 셋은 174 장의 이미지와 537 개의 문장으로 구성된다. 수집된 문장은 본 연구에서

정의한 기준에 따라 시각적 묘사 270 개와 상징적 해석 267 개로 분류되었다. 본 데이터 셋은 모델의 상징 추론 능력을 검증하기 위해 설계되었다. 한국 민화 영역에서 이미지와 매핑된 ‘시각적 묘사’ 및 ‘상징적 해석’을 명확히 구분하여 구조화한 최초의 시각-언어 데이터 셋이라는 의미를 갖는다.

III. 결론

본 연구는 한국 민화와 같이 해석 주석이 부족한 예술 도메인의 한계를 극복하기 위해 AI 보조 기반의 데이터 셋 구축 파이프라인을 제안하였다. 문장 단위 범주 체계를 설계하고 다중 모델을 앙상블 형태로 활용하여 고품질 평가용 데이터 셋을 구축하였다. 본 데이터 셋은 기존 메타데이터의 단순 정보 나열을 넘어서 시각적 요소와 그 이면에 담긴 문화적 상징을 연결했다는 점에서 가치를 지닌다. 향후 연구에서는 데이터의 규모를 확장하고 주석의 신뢰도를 추가 검증함으로써, 상징적 예술에 대한 인공지능의 심층적 이해와 설명 생성 성능을 높이는 데 기여하고자 한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2026-RS-2021-II211835) 그리고 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (RS-2025-22932973)

참 고 문 헌

- [1] 임두빈. (2013). 민화 (民畵) 의 미학적 분석을 통한 정신미 (精神美) 연구. 예술과 미디어, 12(4), 35-50.
- [2] 한유진. (2015). 민화에 나타난 주술성에 관한 연구: 벽사와 길상을 중심으로:(A) study on the shamanistic aspects of folk painting: Focusing on driving out evil spirits and bringing in auspicious signs.
- [3] 조운한. (2014). 민화, 까치호랑이의 의미구조에 관한 연구. 일러스트레이션 포럼, 15(40), 69-78.
- [4] Korea Institute of Design Promotion. (2023). Design DB_Korean Design Motifs_Folk Paintings_20230710 [Data set]. Open Data Portal. <https://www.data.go.kr/data/15116448/fileData.do>
- [5] Zhang, T., Feng, T., Ni, Y., Cao, M., Liu, R., Avestimehr, K., ... & Avestimehr, S. (2025, July). Creating a lens of chinese culture: A multimodal dataset for chinese pun rebus art understanding. In Findings of the Association for Computational Linguistics: ACL 2025 (pp. 22473-22487).
- [6] Garcia, N., & Vogiatzis, G. (2018). How to read paintings: semantic art understanding with multi-modal retrieval. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops (pp. 0-0).
- [7] Stefanini, M., Cornia, M., Baraldi, L., Corsini, M., & Cucchiara, R. (2019, September). Artpedia: A new visual-semantic dataset with visual and contextual sentences in the artistic domain. In International Conference on Image Analysis and Processing (pp. 729-740). Cham: Springer International Publishing.