

# YOLOv8과 CLIP을 이용한 한국 전통 도자기 인식 및 분류 프레임워크 연구

표수현\*, 황영하\*, 최동언\*, 권용재\*, 강주호\*, 백호기 교수\*\*

## A Study on a Framework for Recognizing and Classifying Korean Traditional Pottery Using YOLOv8 and CLIP

Suhyeon Pyo\*, Yeongha Hwang\*, Dongeon Choi\*, Yongjae Kwon\*, Juho Kang\*, Prof. Hoki Baek\*\*

### 요약

본 연구는 한국 전통 도자기의 유형을 자동으로 인식하고, 역사적 고증 정보를 검증하기 위한 객체 인식 및 분류 프레임워크를 제안한다. 제안된 프레임워크는 YOLOv8 기반 객체 검출기를 통해 이미지 내의 도자기 영역을 탐지하고, 검출된 영역을 crop한 후 CLIP 모델을 이용하여 청자, 백자, 기타의 1차 분류를 수행한다. 이후 CLIP zero-shot detection 방식을 적용하여 도자기의 분류를 세분화한다.

실험 결과, 제안된 시스템은 한국 문화유산 이미지 내 전통 도자기의 유형을 비교적 높은 정확도로 구분하였으며, 향후 문화유산 디지털 고증 및 시각지능 기반 분석 연구에 유용하게 활용될 가능성을 보여준다.

### Abstract

This study proposes an object recognition and classification framework designed to automatically identify and verify the historical authenticity of traditional Korean pottery. The proposed system employs a YOLOv8-based object detector to locate and crop jar regions within images. Subsequently, a CLIP model is fine-tuned to perform primary classification into three categories: celadon, white porcelain, and others. Using CLIP's zero-shot detection, each pottery is further categorized into detailed subtypes such as Goryeo celadon, Buncheong ware, Joseon white porcelain, blue-and-white porcelain, iron-painted porcelain, and copper-red porcelain. The results demonstrate that the proposed framework effectively distinguishes types of traditional Korean pottery and holds potential for applications in digital verification of cultural heritage artifacts.

### Key words

YOLOv8, CLIP, K-culture, Object Recognition, Korean Ceramics

---

\*경북대학교 컴퓨터학부, pyo9911@knu.ac.kr

\*경북대학교 컴퓨터학부, jech003@knu.ac.kr

\*경북대학교 컴퓨터학부, ehddjs1128@knu.ac.kr

\*경북대학교 컴퓨터학부, dydwo512@knu.ac.kr

\*경북대학교 컴퓨터학부, kangjuho322@knu.ac.kr

\*\*경북대학교 컴퓨터학부, nelyouyou@knu.ac.kr (교신저자)

※"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업의 연구결과로 수행되었음"(2021-0-01082)

## I. 서 론

한국의 전통 도자기는 시대적·문화적 특징을 반영하는 중요한 유물로, 형태나 문양의 세부 특징을 통해 제작 시기와 역사적 배경을 추정할 수 있다. 그러나 이러한 분류 및 고증은 주로 전문 연구자의 육안 검증에 의존하고 있어, 자동화된 인공지능 기반 분석 기술의 필요성이 커지고 있다.

최근 인공지능 기반 객체 인식 기술이 빠르게 발전함에 따라, 문화유산 이미지를자동으로 인식하고 분류하는 연구가 활발히 진행되고 있다. 특히 YOLO 모델은 실시간 객체 탐지에서, CLIP 모델은 시각-언어 결합 학습 분야에서 뛰어난 성능을 보여주고 있어 문화유산 분류에도 적용할 여지가 있다.

기존 연구는 주로 일반 도자기나 회화류 분류에 집중되어 있었으나, 형태적 변이가 크고 시각적 패턴이 복잡한 도자기 유물에 대한 세분류 연구는 상대적으로 부족하였다.

이에 본 연구는 YOLOv8과 CLIP을 결합한 시각-언어 기반 전통 도자기 인식 프레임워크를 제안하여, 한국 전통 도자기의 자동 분류 및 디지털 고증을 위한 기반 기술을 마련하고자 한다.

표 1. 카테고리별 데이터 구성

Table 1. Dataset composition by category

구분	청자	백자	기타	총합
학습	20	35	16	71
검증	18	13	4	35
합계	38	48	20	106

## II. 한국전통 도자기 인식/분류 프레임워크

제안된 프레임워크는 객체 검출 모듈(YOLOv8), Crop 전처리 모듈, CLIP 기반 분류 모듈, 그리고 Zero-shot 세분류 모듈로 구성된다.

YOLOv8기반 객체 검출 모듈에서는 도자기 탐지를 위해 Ultralytics의 YOLOv8n 모델을 기반으로 실험을 진행하였다. 본 연구에서는 106장의 이미지를 수집하여, Label-studio를 사용하여 각 이미지에

Bounding Box 라벨링을 수행하고, 이미지 크기를 640\*640으로 정규화 하였다. YOLO모델 학습에는 Learning rate = 0.01, Epochs = 100, Batch Size = 16으로 적용하였다.

CLIP Fine-tuning 기반 분류에서는 YOLOv8로 검출된 도자기 이미지가 crop되어 CLIP모델의 시각 인코더에 입력된다. 전처리 단계에서 각 이미지를 224\*224로 resize하고 CLIP의 표준 normalization (mean, std)를 적용하였다. Fine-tuning은 {celadon', 'white\_porcelain', 'other\_pottery} 세 카테고리에 대해서 Epochs = 20, Batch Size = 32로 진행하였다.

본 연구의 CLIP의 파인튜닝에선 부족한 데이터셋을 고려하여, 과잉 적합 방지를 위해 Feature Extraction 기반의 전이 학습 전략을 채택했다. 기존 CLIP vision encoder의 가중치를 고정한 상태에서, 새로운 경량의 classification head를 추가해 이를 학습시키는 방식이다.

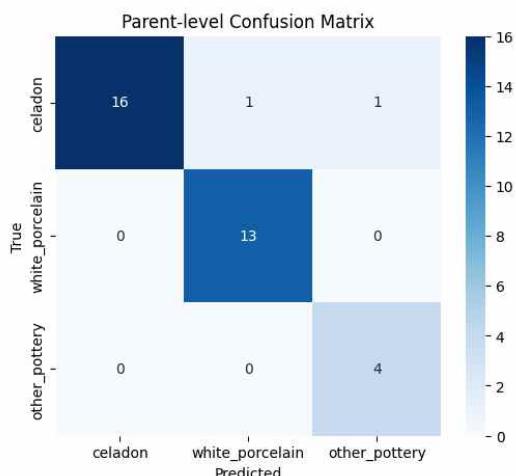


그림 1. 검증 데이터셋에 대한 CLIP Fine-tuning 기반 분류 결과 (Parent-level Confusion Matrix)

Fig. 1. Classification Results on the Validation Dataset Based on CLIP Fine-tuning (Parent-level Confusion Matrix)

표 2. 검증 데이터셋에 대한 CLIP Fine-tuning 기반 분류 성능

Table 2. Classification Performance on the Validation Dataset Based on CLIP Fine-tuning

	Precision	Recall	F1-score	Support
celadon	1.00	0.98	0.94	18
white_pocelain	0.93	1.00	0.96	13
other_pottery	0.80	1.00	0.89	4
accuracy			0.94	35
macro_avg	0.91	0.96	0.93	35
weighted_avg	0.95	0.94	0.94	35

표 3. 프롬프트에 따른 CLIP Zero-shot detection 결과

정확성 (Input : 조선백자 이미지)

Table 3. Accuracy of CLIP Zero-shot Detection

Depending on Prompts

(Input : Image of Joseon\_white\_porcelain)

언어	내용	조선 백자	청화 백자	철화 백자	동화 백자	결과
한 글	단어 만	0.2035	0.2046	0.2112	0.1055	철화 백자
	묘사	0.2087	0.2065	0.2080	0.2088	동화 백자
영 어	단어 만	0.3299	0.2467	0.2784	0.2611	조선 백자
	묘사	0.3756	0.2912	0.3306	0.3083	조선 백자

CLIP Zero-shot 세분화 분류에서는 표3. 와 같이 프롬프트의 내용과 언어가 분류 정확도에 영향을 끼친다. 따라서 영어로 자세하게 묘사된 프롬프트를 사용하여 분류하였다. 자세한 프롬프트 내용은 아래와 같다.

고려청자 : “Goryeo dynasty inlaid celadon pottery with jade green glaze, Korean celadon bowl with sanggam inlay technique”

분청사기 : “Joseon dynasty buncheong ware with stamped decoration, traditional Korean buncheong bowl with grey-green glaze”

조선백자 : “Joseon dynasty white porcelain jar, unpainted, smooth white glaze, traditional Korean baekja”

청화백자 : “blue and white porcelain from the Joseon dynasty”

철화백자 : “Joseon dynasty iron-painted white porcelain with black painted decoration, traditional Korean jar or bowl”

진사백자 : “Joseon dynasty polychrome painted (red color painted) porcelain called dongchae baekja”

기타 : “Korean earthenware pottery that is not celadon or white porcelain”



그림 2. CLIP zero-shot detection 기반 분류 결과

Fig. 2. Classification Results Based on CLIP

Zero-shot-Detection

### III. 한계 및 향후 연구

앞서 제안한 YOLOv8-CLIP 기반 도자기 인식 및 분류 프레임워크에는 다음과 같은 한계가 존재한다.

첫째, 수집 가능한 문화유산 이미지의 수가 매우 제한적이어서, 데이터셋 규모가 작고 일부 유형(예: 진사백자, 철화백자 등)은 샘플 수가 충분하지 않았다. 이에 따라 모델의 일반화 성능이 떨어질 가능성이 존재한다.

둘째, 분류 대상이 ‘청자-백자-기타’ 및 소수의 세부 항목으로 한정되어 있어, 실제 전통 도자기의 다양한 형태와 문양을 모두 포괄하기에는 부족하다.

셋째, CLIP 모델의 zero-shot 분류 성능은 프롬프트 설계에 따라 편차가 존재하였으며, 한국어-영어 간 표현 차이로 인해 세밀한 구분이 어려운 경우가 있었다.

향후 연구에서는 문화유산 아카이브, 박물관 데이터 등 외부 자료를 추가 확보하여 데이터셋을 확장하고, 또한 한국어 멀티모달 표현을 지원하할 수 있는 모델을 적용하면 분류 정밀도와 문화적 맥락 이해 측면에서 개선이 가능할 것으로 기대된다.

### IV. 결 론

본 연구는 YOLOv8과 CLIP을 결합한 한국 전통 도자기 인식 및 분류 프레임워크를 제안하였다.

YOLOv8의 객체 검출과 CLIP의 시각-언어 표현 능력을 융합함으로써, 전통 도자기의 시각적 특징을 유형별로 세밀하게 구분할 수 있었다.

향후에는 한국어 프롬프트 지원이 가능한 CLIP 모델 및 다국어 임베딩 모델을 활용한다면, 한국어 기반 문화유산 인식 모델로 확장할 수 있을 것으로 기대된다. 또한 도자기 외에도 불상, 석탑, 도자기 문양 등 다양한 문화유산 객체에 대한 3D 복원 및 디지털 고증 연구로 발전시켜 나갈 가능성이 있다.

### 참 고 문 헌

- [1] J. Redmon et al., “You Only Look Once: Unified, Real-Time Object Detection,” Proc. CVPR, 2016.
- [2] A. Radford et al., “Learning Transferable Visual Models From Natural Language Supervision,” arXiv preprint arXiv:2103.00020, 2021.
- [3] 이대호, 정인택, 최준우, 김주영, 김주연, 홍진혁. (2021-06-23). 딥러닝을 활용한 도자기 이미지 분류 이해. 한국정보과학회 학술발표논문집, 제주.
- [4] 강우진, 전준렬, 김송이, 박도윤, 이지현, 홍진혁. (2019-06-26). Mask R-CNN을 활용한 한국 문화 유산(도자기와 토기) 자동 검출 연구. 한국정보과학회 학술발표논문집, 제주.
- [5] Cardarelli, Lorenzo. 2025. "PyPotteryLens: An Open-Source Deep Learning Framework for Automated Digitisation of Archaeological Pottery Documentation." Digital Applications in Archaeology and Cultural Heritage 38 (e00452).