VaseVQA-3D: Benchmarking 3D VLMs on Ancient Greek Pottery

Nonghai Zhang1* Zeyu Zhang1*† Jiazi Wang2*

Yang Zhao3 Hao Tang1‡

1Peking University 2Beijing Jiaotong University 3La Trobe University

*Equal contribution. †Project lead. ‡Corresponding author.

ICLR International Conference On Learning Representations

github.com/AIGeeksGroup/VaseVQA-3D aigeeksgroup.github.io/VaseVQA-3D

LA TROBE UNIVERSITY

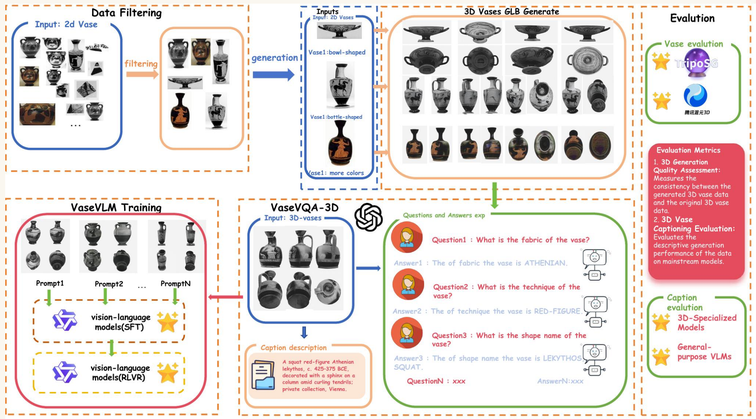
Empowering Cultural Heritage Preservation with AI !

Motivation & Problem Statement

VLMs face critical gaps in 3D cultural heritage understanding:

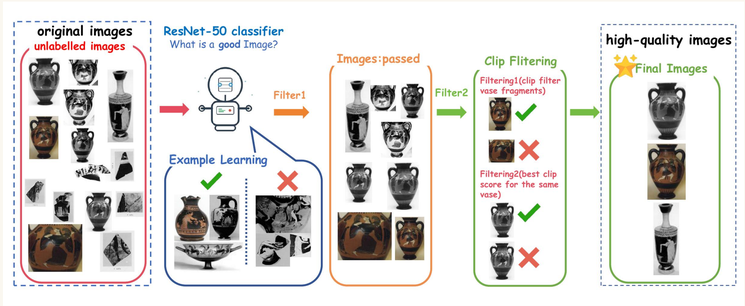
- Data Scarcity: No 3D VQA dataset for archaeological artifacts. Ancient Greek pottery represents rare long-tail data (>2,000 yrs history, 300 yrs of research).
- Domain Gap: General VLMs lack expertise in archaeological semantics: Fabric · Technique · Shape · Dating · Decoration · Attribution.
- 3D Reasoning: Archaeological significance lies in spatial features — symmetry, proportions, morphology — invisible in 2D fragmented views.

Complete Data Construction Pipeline

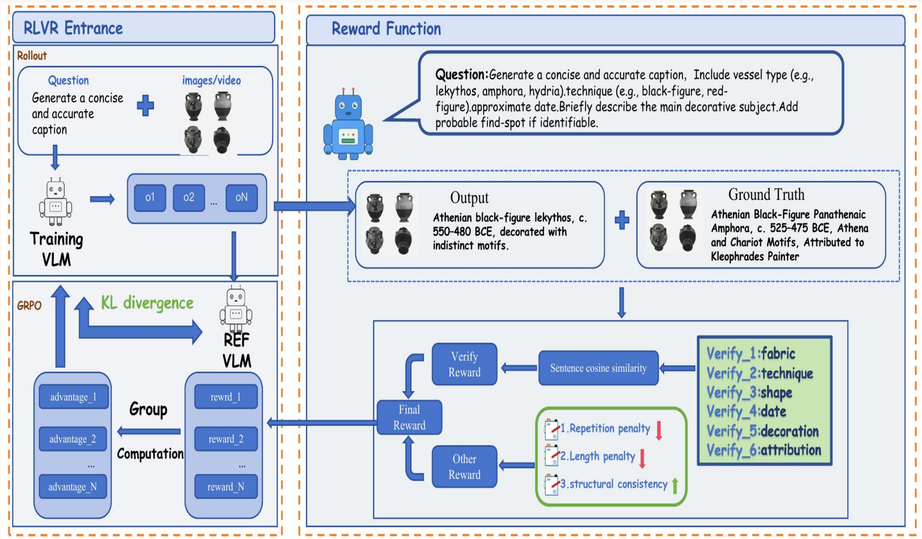


Pipeline: 30K images → 3,880 filtered → 664 3D models → 4,460 QA pairs → VaseVLM

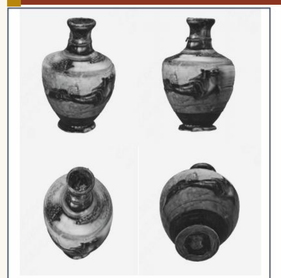
ResNet-50 + CLIP Dual Filtering Pipeline



RLVR Framework — Multi-dimensional Reward

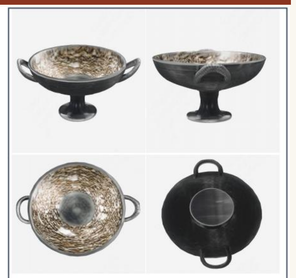


3D Dataset Examples



Athenian black-figure lekythos, c. 575-525 BCE, adorned with dogs and owls; National Museum, Copenhagen.

Squat Athenian red-figure lekythos, c. 425-375 BCE, depicting a woman and Eros; Cleveland Museum of Art.



Athenian Red-Figure Cup by Epiktetos, c. 525-475 BCE, featuring symposium scene with reclining woman.

Athenian Black-Figure Lekythos, c. 500-450 BCE, Beldam Workshop, Ivy and Berry Motif, Nola Provenance.

Athenian black-figure lekythos, c. 500-450 BCE, adorned with ivy leaf and berry motifs; National Museum, Warsaw.

Athenian Black-Figure Panathenaic Amphora, c. 525-475 BCE, Athena and Chariot Motifs, Attributed to Kleophrades Painter

★ Comprehensive VLM Evaluation on VaseVQA-3D

Method	FID↓	CLIP↑	R@10↑	R@5↑	R@1↑	Lex.↑
3D-Specialized Models						
DiffuRank	0.421	0.798	16.67%	8.33%	2.08%	0.274
Cap3D	0.445	0.792	14.58%	7.29%	1.56%	0.267
LLaVA3D	0.494	0.784	10.42%	5.21%	1.04%	0.238
Closed-source VLMs						
Gemini-2.5-flash	0.325	0.736	28.57%	17.58%	2.20%	0.210
Claude-4-sonnet	0.353	0.676	23.96%	10.42%	3.12%	0.188
Gemini-2.5-Pro	0.397	0.680	22.92%	14.58%	3.12%	0.162
GPT-4.1	0.501	0.644	25.00%	10.42%	3.12%	0.128
Claude-3.5-sonnet	0.455	0.643	15.62%	8.33%	2.08%	0.116
Doubao-1.5-vision	0.504	0.606	14.58%	4.17%	1.04%	0.074
GPT-4o	0.582	0.520	13.54%	6.25%	2.08%	0.104
Claude-3.7-sonnet	0.600	0.339	13.54%	6.25%	1.04%	0.101
InternVL	0.376	0.771	10.42%	8.33%	2.08%	0.252
Qwen2.5-VL-7B	0.334	0.775	18.75%	9.38%	2.08%	0.217
Qwen2.5-VL-3B	0.358	0.782	9.38%	6.25%	1.04%	0.259
VaseVL	0.493	0.790	10.40%	6.25%	2.08%	0.255
VaseVLM — Ours (bold = best)						
VaseVLM-3B-SFT	0.359	0.788	17.71%	8.33%	2.08%	0.223
VaseVLM-3B-RL	0.363	0.789	17.71%	10.42%	2.08%	0.245
VaseVLM-7B-SFT	0.332	0.779	20.83%	10.42%	3.12%	0.272
VaseVLM-7B-RL	**0.328**	**0.792**	**21.24%**	**11.12%**	**3.52%**	**0.276**

★ **VaseVLM-7B-RL achieves best R@1 (+12.8% vs. Best) and Lex. Sim. (+6.6% vs. Best).**

Example: Red-Figure vs Black-Figure Classification

GT:"Athenian Red-Figure Cup, c. 500–450 BCE, depicting a youth wreathing an altar; Detroit Institute of Arts."

Gemini 2.5 Flash (R@1): "Athenian bl sack-figure kylix, c. 550-500 BCE, with figural decoration; Attica."

VaseVLM-7B-RL (R@1): "Athenian red-figure cup, c. 500–450 BCE, depicting a youth at an altar; Detroit Institute of Arts."