Beyond Keywords:

Retrieving Blue-and-White Ceramics in Dutch Paintings with Knowledge-Augmented CLIP

Yang Zhao · School of Information Studies, Syracuse University

NKOS Workshop · DCMI 2026 · Seoul (virtual)

The Problem: metadata describes the painting, not what is in it

- Museum records cover title, artist, date, medium etc.
- The descriptions cannot list every object depicted inside a painting.
- Many relevant works are catalogued only as "still life": the ceramics they contain are never mentioned.
- This is a structural limit of text-based knowledge organization applied to visual collections.

Art historian

Searches "kraak", "wan-li" → fewer than 5 hits. Searches "porcelain" → many hits, mostly irrelevant.

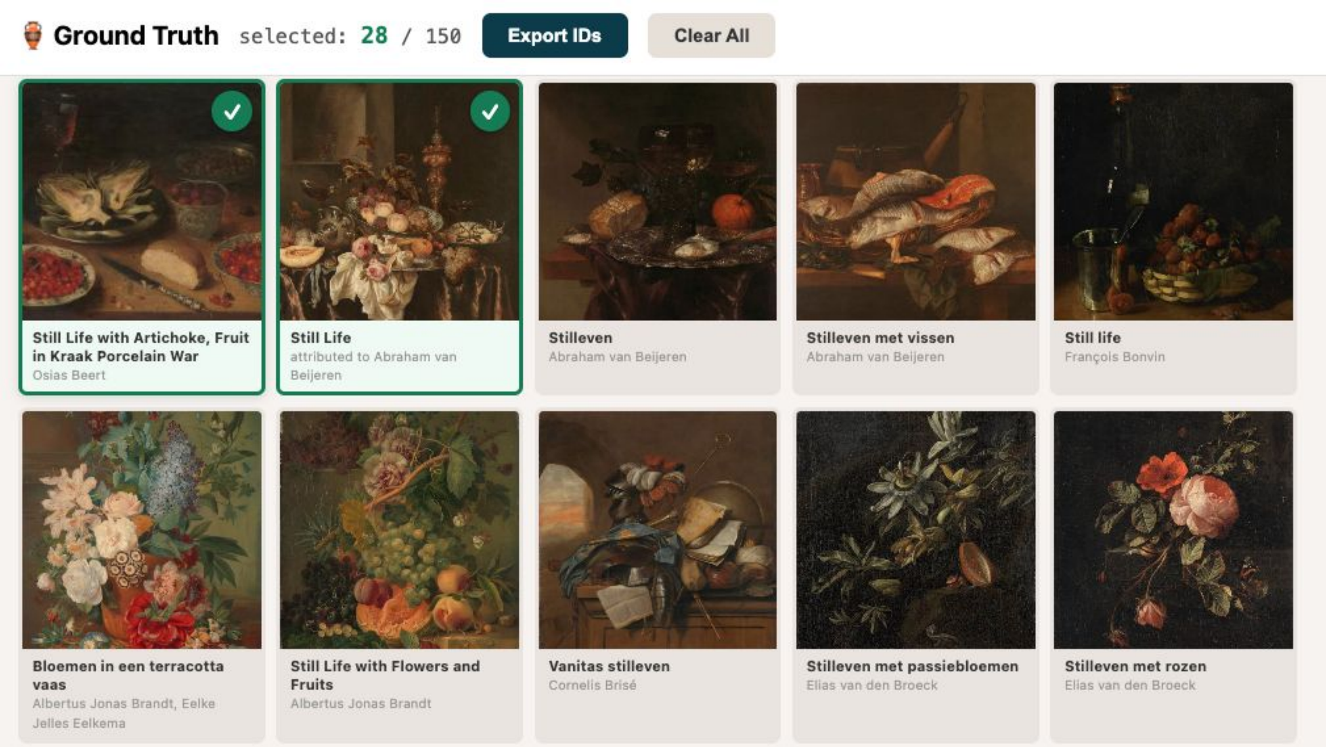
Museum visitor

Recognizes a blue-and-white plate in a painting but has no vocabulary to find similar works.

Users recognize what they seek when they see it — whether or not they can name it.

Corpus: still lifes from Rijksmuseum

- 150 digitized oil paintings from the Rijksmuseum Linked Art API, predominantly Dutch Golden Age still lifes.
- 28 manually identified as containing blue-and-white ceramics.



Three ceramic plates, one look

Chinese

Kraak porcelain

before 1613; Export ware, Sino-Dutch trade

Japanese

Arita/Imari Ware

c. 1600 - c. 1699; Imitated kraak, then diverged

Dutch

Delft blue earthenware

c. 1660 - c. 1670; Local imitation of imports

All three share near-identical blue-and-white decoration. So the retrieval target should be the visual class "blue-and-white".

Three retrieval approaches

1

Metadata keyword search

Phrase-matching on the catalogue at 3 expertise levels: general ("porcelain"), basic ("blue and white"), expert ("kraak", "wan-li", "Delft blue").

2

Zero-shot CLIP

Encode natural-language prompts, rank paintings by visual similarity. No domain knowledge required.

3

Knowledge-augmented CLIP

Retrieve visually similar reference images, append their scene annotations to the prompt, re-encode into an enriched query.

Reference set = 20 annotated paintings from external open-access collections. No reference image overlaps the test corpus.

The knowledge base:
20 paintings & annotations

- Each reference painting paired with two short scene-context annotations (30–50 words).
- Annotations describe visually observable features: vessel type, compositional placement, surrounding objects.
- Curated from external open-access museum collections, kept fully separate from the test corpus.

Example annotation

"A wide blue-and-white dish sits at the left edge of the table, holding fruit. Behind it, a pewter jug and a half-peeled lemon. The dish is angled toward the viewer, its rim decorated with panelled borders."

Describe the vessel in its painted scene, not as an isolated object.

Key design choice: scene descriptions beat object descriptions

Scene-context

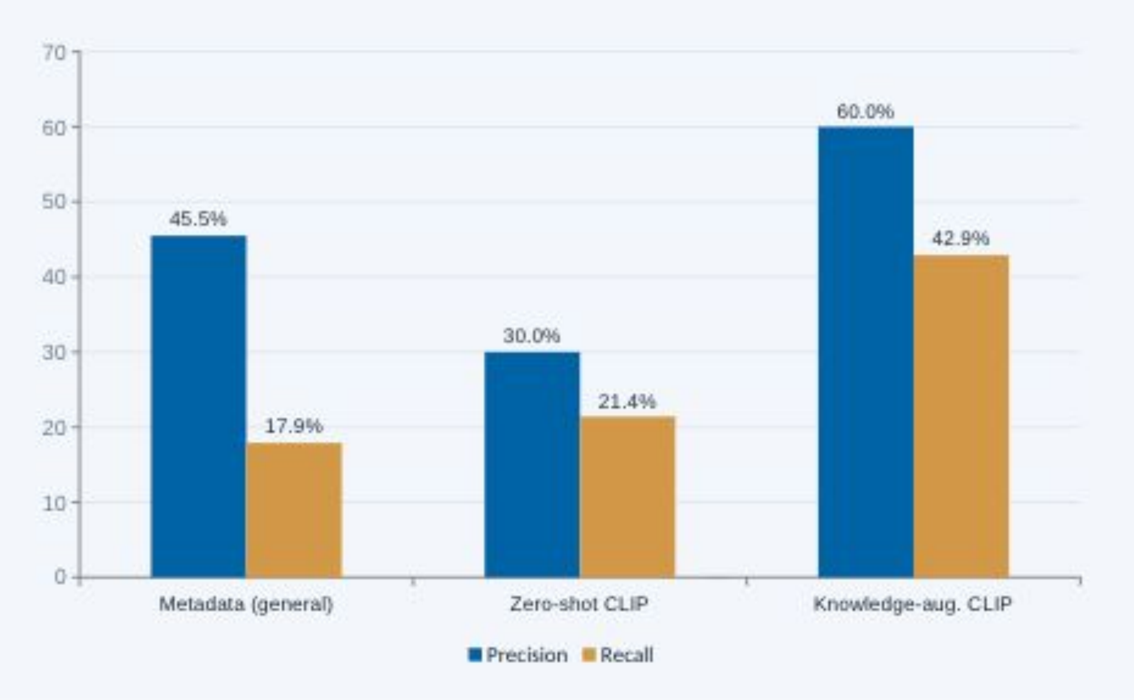
- Where the vessel sits, what surrounds it, how it is angled.
- Matches how CLIP reads a whole painting.
- Better results.

Object-feature

- Glaze type, decorative motif, isolated ceramic features.
- Overlaps semantically with unrelated content (e.g. flower arrangements).
- Produces false positives & even worse than zero-shot.

Annotation type matters more than annotation volume.

Results: knowledge augmentation doubles precision and recall



With only 20 annotated paintings

+7

ground-truth paintings found that zero-shot misses

while losing only 1.

Expert-term metadata reaches 100% precision but only 10.7% recall (most paintings simply don't have related descriptions).

Live demo

Type a query in plain language → no catalogue vocabulary required.

https://huggingface.co/spaces/YangZhao2026/blue-n-white

The argument in action :
This is a VISUAL problem. You recognize the motif when you SEE it.

Ablation: fewer, better annotations win

Adding 10 object-feature descriptions of ceramic photographs made retrieval worse: even below zero-shot CLIP.

20 scene-context (paintings only)

Best

Descriptions match how CLIP reads a whole painted scene.

30 items (+10 object-photos)

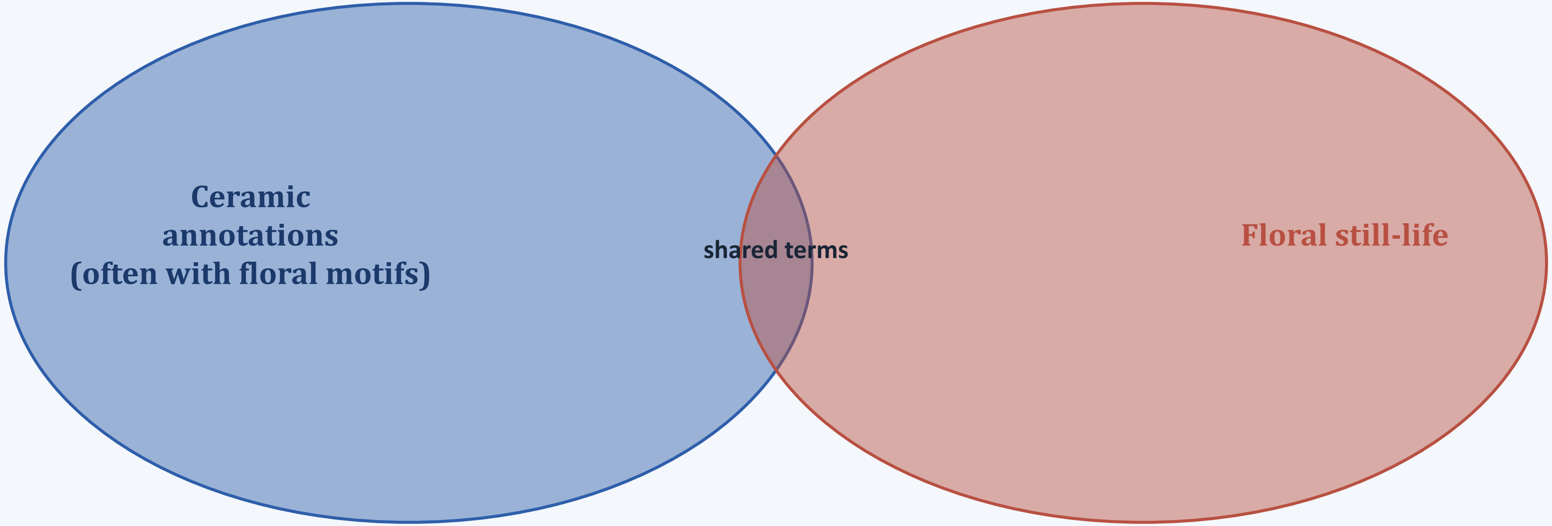
Worse than zero-shot

Isolated ceramic features overlap with flower arrangements → false positives.

Knowledge organization still matters: it just shifts from controlled terms to contextual scene descriptions.

Why the ceramic photos hurt: vocabulary overlap

Where annotations of ceramics contain floral patterns, they share terms with floral still-life, and CLIP cannot tell them apart.



When metadata claims more certainty than the image allows

- *“This example includes heavily wrought and gilded silver trays and ewers, **Japanese Kakiemon and Imari porcelain bowls**…”*

- A painting shows a motif, not a physical object, so even a specialist cannot know which vessel the painter saw. This example reads as Kraak, not Kakiemon/Imari (see below), yet is described as the latter. That is not simply "wrong": early Imari copied Kraak. A better description would be “Kraak-style ceramic bowls”.

© Rijksmuseum

Still Life with Silver (detail)
Alexandre François Desportes, French
1720s © The Met

The problem is structural, not one institution’s slip:
No single label can be ground truth when the visual evidence is inherently ambiguous.

The challenge is epistemically informative

- Retrieval mistakes map the boundary of what computational vision can resolve in a painted image.
- The visual ambiguity between vessel types or genres is real, not just a matter of modelling.
- A "wrong" answer often points to a genuine hard case a human would also hesitate on.

Limitations & future work

Narrow scope
One object class, one collection.

Proper evaluation
Retrieval benchmarks, larger ground-truth annotation, multi-model comparison.

Generalization
The approach applies wherever visual collections hold specific content metadata cannot cover.

References

Balauca, A., et al. (2024). Taming CLIP for fine-grained and structured visual understanding of museum exhibits. arXiv:2409.01690.

Economou, M., & Kist, C. (2024). Practice architectures for bridging the semantic gap in museum documentation. Museum Management and Curatorship, 1–20. https://doi.org/10.1080/09647775.2024.2426799

Fiebrink, R., & Sologub, P. (2024). Sketchy collections: Exploring digital museum collections by drawing via CLIP. NeurIPS Creative AI.

Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems (NeurIPS).

Mahowald, J., & Lee, B. C. G. (2024). Integrating visual and textual inputs for searching large-scale map collections with CLIP. In Proceedings of the Computational Humanities Research Conference 2024 (CHR 2024), Aarhus, Denmark. https://ceur-ws.org/Vol-3834/paper17.pdf

Meyer, L. S., Engel Aaen, J., Tranberg, A. R., Kun, P., Freiberger, M., Risi, S., & Løvlie, A. S. (2024). Algorithmic ways of seeing: Using object detection to facilitate art exploration. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1–18). ACM. https://doi.org/10.1145/3613904.3642157

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In M. Meila & T. Zhang (Eds.), Proceedings of the 38th International Conference on Machine Learning (PMLR 139, pp. 8748–8763). https://proceedings.mlr.press/v139/radford21a.html

Smits, T., & Wevers, M. (2023). A multimodal turn in digital humanities: Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections. Digital Scholarship in the Humanities, 38(3), 1267–1280. https://doi.org/10.1093/llc/fqad008

Villaespesa, E., & Murphy, O. (2021). This is not an apple! Benefits and challenges of applying computer vision to museum collections. Museum Management and Curatorship, 36(4), 362–383. https://doi.org/10.1080/09647775.2021.1873827

Thank you

Acknowledgement

This work began as a course project. Special thanks to Prof. Bei Yu for her feedback and guidance.

Questions or suggestions?

Reach me at

Email: yangz24@illinois.edu (Work completed at Syracuse University.)

Demo: https://huggingface.co/spaces/YangZhao2026/blue-n-white

Code & data: https://github.com/yzhao-is/Blue-n-White