

Retrieval-Augmented Generation for Symbolic Folk-Art Interpretation: Comparing Dense and Graph-Based Retrieval Using Korean Minhwa

Minjun Song^{1,2}, Sumin Kim¹, Junsung Kim¹, Yebeen Lee¹, Heung-No Lee^{1,2,†}

¹Gwangju Institute of Science and Technology (GIST), ²LiberVance
{paulmjsong, smkim6927, wnstjd123, leeyeebeen_ug}@gm.gist.ac.kr
heungno@gist.ac.kr [†]Corresponding author

Abstract

Existing artwork-oriented retrieval-augmented generation (RAG) systems typically support visual understanding through contextual information such as artist metadata and historical background. Symbolic folk-art interpretation instead requires connecting visible motifs to culturally situated meanings. We study semiotic retrieval for this task in Korean *minhwa*, comparing dense semantic and graph-structured retrieval. We introduce SemioticRAG, which organizes symbolic knowledge through Barthesian Form, Concept, JointConcept, and Myth structures. Blind human evaluation shows that DenseRAG performs best overall, with SemioticRAG ranking second. Qualitative analysis indicates that SemioticRAG produces more explicit pathway-based symbolic compositions, while depending on Form-level grounding and graph construction quality. Overall, explicit semiotic structure changes how culturally grounded meanings are composed rather than uniformly improving interpretive quality.

1 Introduction

Recent advances in multimodal language models (LMs) and retrieval-augmented generation (RAG) have significantly improved automatic visual understanding and description generation for artworks. Existing approaches to artwork understanding commonly enhance generation through retrieval of contextual information such as artist biographies, historical background, and curatorial metadata. Systems such as ArtRAG (Wang et al., 2025) demonstrate that multi-perspective retrieval can improve the richness and factual grounding of generated art descriptions by augmenting visual reasoning with external contextual knowledge. While effective for institutional and canonical artworks with extensive documentation, these approaches do not directly address settings where symbolic interpretation, rather

than factual contextualization, constitutes the primary interpretive task.

Korean *minhwa* presents one such case. Unlike canonical fine art traditions, *minhwa* often lacks extensive archival metadata regarding authorship, provenance, or formal historical documentation. Instead, its cultural significance emerges through symbolic motifs associated with folk beliefs, Confucian virtues, prosperity, protection, and cosmological worldviews (Han, 2015; Im, 2013). Motifs such as tigers, cranes, peaches, and bookshelves frequently convey layered symbolic meanings beyond their literal depiction, functioning as visual carriers of auspicious wishes, moral values, and apotropaic beliefs (Han, 2015; Cho, 2014b). Consequently, the task of interpreting *minhwa* differs fundamentally from generic artwork description: meaningful interpretation requires modeling culturally situated symbolic signification rather than merely retrieving descriptive background information.

Using Korean *minhwa* as our empirical testbed, we study how retrieval structure affects symbolic folk-art interpretation through a framework we call SemioticRAG. Motivated by Barthesian semiotics, SemioticRAG organizes symbolic knowledge as a knowledge graph (KG) connecting Form, Concept, JointConcept, and Myth nodes through successive layers of signification. Given an input image, the system extracts visual entities using a multimodal LM, retrieves corresponding Form nodes via similarity matching, and traverses the graph to obtain candidate symbolic pathways, which are then used to condition generation. This formulation provides a structured way to compare symbolic graph-based retrieval with dense semantic retrieval under a shared generation setting, rather than introducing a standalone model optimization objective.

We evaluate SemioticRAG against direct multimodal generation, dense semantic retrieval, and a prior art-oriented RAG baseline. Human evaluation considers visual fidelity, coverage, interpretive

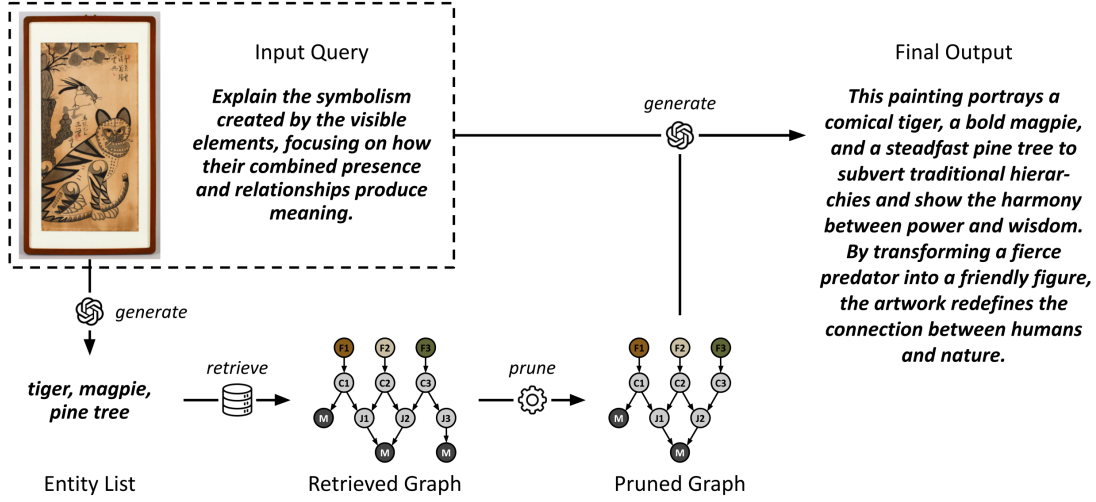


Figure 1: Overview of the SemioticRAG retrieval and generation pipeline. Given an input *minhwa* image, the system first extracts visually salient entities and uses them to retrieve candidate symbolic pathways from the semiotic KG. Retrieved symbolic context is then pruned through coverage-aware ranking before being supplied, together with the original image and an interpretation prompt, to a multimodal LM for symbolic interpretation generation. Image source: *까치호랑이*, Samcheok Municipal Museum (삼척시립 2291), [eMuseum](#), KOGL Type 1.

soundness, and cultural authenticity. Experimental results show that dense semantic retrieval is the strongest overall approach, consistently outperforming graph-structured retrieval across all four criteria, while SemioticRAG ranks second overall. Qualitative analysis indicates that SemioticRAG produces more explicit pathway-based symbolic compositions, but this behavior depends on Form-level grounding and graph construction quality. Overall, the findings suggest that explicit symbolic structure changes the mode of interpretation rather than uniformly improving interpretive quality.

Our contributions are as follows:

- We formulate semiotic retrieval for symbolic folk-art interpretation, distinguishing it from conventional artwork description and contextual retrieval tasks.
- We propose SemioticRAG, a graph-structured retrieval framework grounded in Barthesian semiotics for explicit symbolic traversal and interpretation.
- We compare graph-structured symbolic retrieval and dense semantic retrieval under a unified framework for Korean *minhwa* interpretation.
- We analyze how retrieval structure affects compositional interpretation, highlighting differences between flexible semantic retrieval and explicit pathway-based symbolic composition.

2 Related Work

This section reviews prior work on retrieval structure and evidence integration, knowledge-augmented artwork understanding, and semiotic interpretation relevant to retrieval-augmented folk-art analysis.

2.1 Retrieval Structure and Evidence Integration in RAG

Retrieval-augmented generation (RAG) (Lewis et al., 2020) has become a widely adopted framework for incorporating external knowledge into language generation systems. An important design choice concerns how retrieved evidence is organized and integrated before generation. Graph-structured approaches organize information through explicit relational structure and enable multi-hop traversal over connected knowledge representations.

KG²RAG (Zhu et al., 2025) shows that graph-based expansion can improve contextual completeness by reorganizing retrieved evidence through KG structure. GraphRAG (Edge et al., 2025) similarly demonstrates how graph-organized retrieval can support structured aggregation of dispersed information for complex generation tasks. GNN-RAG (Mavromatis and Karypis, 2025) demonstrates that learned reasoning over graph neighborhoods can strengthen multi-hop inference through structured selection of relational paths, while KG-

RAG (Sanmartin, 2024) further illustrates how explicit graph representations can improve controllability by grounding retrieval in structured relational states. More recent retrieval frameworks such as LightRAG (Guo et al., 2025) and MS-RAG (You et al., 2025) further explore lightweight and multi-semantic retrieval formulations for improving retrieval efficiency and semantic coverage. ReGraphRAG (Kim et al., 2025), however, highlights that incomplete or fragmented KGs can limit the benefits of structured retrieval.

Beyond retrieval structure, prior work has also shown that downstream generation depends on how retrieved evidence is transformed and integrated before use. Retrieved documents may be normalized into a common representation to reduce inconsistency (Ranaldi et al., 2026), while DialecticRAG (Ranaldi et al., 2025) explicitly compares and reconciles conflicting evidence before generation. These studies motivate a broader view of RAG in which retrieval quality depends not only on what is retrieved, but also on how evidence is represented and organized for generation. Our work investigates this question in symbolic folk-art interpretation by comparing dense passage retrieval with graph-structured semiotic retrieval.

2.2 Knowledge-Augmented Art Interpretation

External knowledge has also been used to improve art and cultural heritage understanding by enriching visual representations with contextual information. Existing approaches typically incorporate metadata such as artist biographies, historical context, stylistic categories, and curatorial documentation to support more informed interpretation of artworks.

Earlier work on knowledge-aware art description generation (Bai et al., 2021) explored knowledge-aware artwork description generation by integrating artistic attributes and contextual information into multimodal generation frameworks. KALE (Jiang et al., 2024) integrates structured cultural metadata into a vision-language framework by aligning visual features with associated textual and structured information. ArtRAG (Wang et al., 2025) constructs an art-specific KG from curated textual sources and retrieves relevant subgraphs for contextual grounding during generation. ArtSeek (Fanelli et al., 2025) uses multimodal retrieval and external contextual signals to support artwork understanding and explanation generation.

These methods largely frame interpretation as contextual augmentation through artist informa-

tion, historical context, stylistic metadata, or related documentation. That framing is less aligned with Korean *minhwa*, where symbolic meaning often arises from culturally embedded motifs and compositional relations rather than artist-centered provenance or institutional metadata (Im, 2013; Han, 2015). Motifs such as tigers, magpies, and pine trees, for example, may jointly participate in culturally established associations involving protection, auspiciousness, and everyday wishes (Cho, 2014b).

Our work differs from prior art-oriented RAG systems in two ways. First, it distinguishes symbolic folk-art interpretation from general artwork description. Second, rather than retrieving broad contextual information, SemioticRAG organizes retrieval around symbolic signification pathways intended to model how visual motifs acquire culturally situated meaning.

2.3 Semiotics and Symbolic Interpretation

Our work is grounded in structural semiotic theory, particularly Barthes’ account of signification as a multi-order process involving denotation, connotation, and myth (Barthes, 1972, 1977a,b). The relation between signifier and signified was formalized in Saussurean semiotics (Saussure, 2011), while Peircean semiotics emphasizes interpretive multiplicity and context-dependent meaning-making (Peirce, 1931–1958). Broader traditions in structural and cultural semiotics similarly emphasize that symbolic meaning emerges relationally and culturally rather than through purely literal description (Eco, 1979).

Despite this theoretical foundation, relatively little work has explored how multi-level signification structures can be operationalized within modern RAG systems. Our study operationalizes this perspective through explicit semiotic retrieval pathways and examines how they differ from dense semantic retrieval in supporting layered symbolic interpretation.

3 Task Definition and Formulation

This section defines the symbolic folk-art interpretation setting investigated in this work and introduces the semiotic representation used to organize retrieval. Rather than formalizing symbolic interpretation as a deterministic reasoning task, we examine how different retrieval structures support culturally grounded interpretation of visual motifs.

3.1 Task Definition

Existing retrieval-augmented artwork understanding systems primarily focus on contextual augmentation, where external information such as artist metadata, historical context, and stylistic descriptions is used to support visual interpretation. In contrast, this work investigates symbolic interpretation, where the objective is to model how visual motifs acquire culturally grounded meaning through structured symbolic associations rather than direct descriptive alignment.

In Korean *minhwa*, visual motifs often acquire meaning through relations with other motifs and established cultural conventions. A single entity may play multiple symbolic roles depending on context and narrative structure. Symbolic interpretation is therefore not purely descriptive, but an interpretive process linking visual entities to higher-level cultural abstractions.

We frame symbolic folk-art interpretation as a RAG setting in which retrieved symbolic associations are used to support culturally grounded interpretation of visual motifs. The objective is not merely to describe visible content or retrieve contextual metadata, but to generate interpretations that connect observed motifs to broader symbolic and cultural meanings.

Because symbolic interpretation is inherently ambiguous and culturally contextual, the task does not assume a single deterministic target interpretation. Instead, system behavior is evaluated operationally through human judgments of visual fidelity, symbolic coverage, interpretive soundness, and cultural authenticity.

3.2 Semiotic Representation

To organize symbolic associations computationally, we model symbolic knowledge using a layered semiotic schema inspired by Barthes. The graph contains four node types:

- **Form:** denotative symbolic forms associated with visually identifiable entities
- **Concept:** abstract symbolic ideas produced through connotative interpretation
- **JointConcept:** intermediary compositional structures representing combinations of multiple concepts that collectively participate in higher-order symbolic interpretation

- **Myth:** higher-order cultural narratives or ideologies that contextualize symbolic associations within broader folk belief systems and cultural worldviews

This representation models symbolic interpretation as a structured transition from concrete visual perception toward increasingly abstract cultural meaning. Importantly, these mappings are not assumed to be deterministic or one-to-one. A single Form may connect to multiple Concept nodes, while multiple symbolic pathways may contribute to different higher-order interpretations.

3.3 Problem Formulation

Let I denote an input artwork image depicting Korean *minhwa*. The goal of symbolic interpretation is to generate a natural-language explanation Y that describes both the visible content and its culturally grounded symbolic meanings.

We assume a set of visually salient entities $E = \{e_1, e_2, \dots, e_n\}$ extracted from I . These entities correspond to identifiable motifs such as animals, plants, or objects that participate in symbolic interpretation.

To support interpretation, we consider a semiotic KG $G = (V, R)$, where V contains symbolic nodes of type Form, Concept, JointConcept, and Myth, and R encodes symbolic relations between them. Given E , retrieval-based systems in this work construct or retrieve a symbolic context G' that links visual entities to higher-level cultural meanings.

Conditioned on the input image I and retrieved symbolic context G' , the model generates an interpretation Y . Outputs are evaluated in terms of visual grounding, symbolic coverage, interpretive coherence, and cultural appropriateness.

In this formulation, the focus of the paper is not to define a single optimal symbolic interpretation, but to study how different retrieval structures influence the quality and nature of generated interpretations.

4 Methodology

We study different retrieval and knowledge integration strategies for symbolic interpretation of folk-art images by comparing five systems: Direct, DenseRAG, SemioticRAG, SemioticRAG-YOLO, and ArtRAG.

The following section focuses on the construction and operationalization of the SemioticRAG

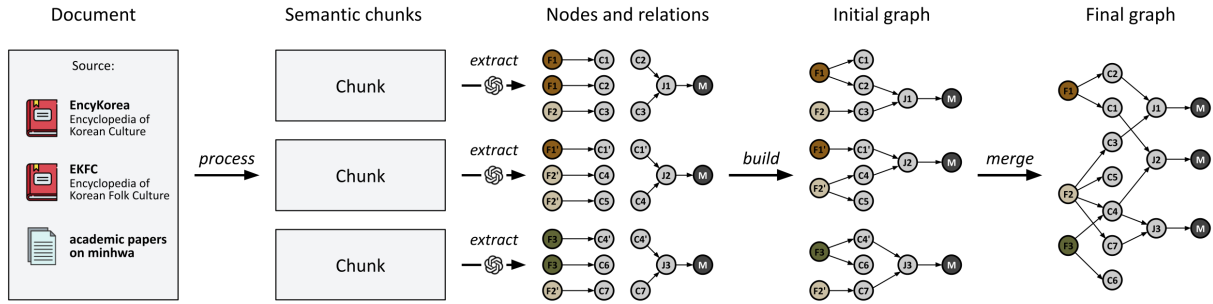


Figure 2: Semiotic KG construction pipeline. Source documents are segmented into semantic chunks, after which an LM extracts symbolic nodes and relationships according to a predefined semiotic schema. The extracted structures are used to construct the graph, followed by graph-cleaning steps that merge duplicate or overlapping nodes.

framework, which represents the graph-structured symbolic retrieval condition in our comparison.

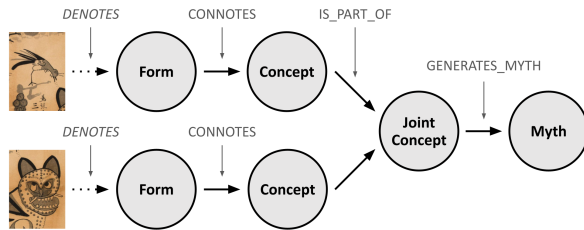


Figure 3: Semiotic graph schema used in SemioticRAG. Visual entities in an image are associated with Form nodes through denotative matching. Form nodes connect to higher-level Concept nodes through symbolic associations, while multiple Concept nodes may combine into a JointConcept representing a larger symbolic composition. JointConcept nodes then connect to Myth nodes representing broader cultural meanings. Artwork details: source as in Figure 1 (KOGIL Type 1).

4.1 Semiotic KG Construction

SemioticRAG relies on a symbolic KG constructed from cultural and scholarly texts describing Korean *minhwa*. The goal of the graph is not to encode exhaustive cultural knowledge, but to organize recurring symbolic associations that can support retrieval during interpretation. The construction pipeline is illustrated in Figure 2.

Corpus Collection and Preprocessing The corpus consists of cultural-heritage reference sources and academic literature on Korean *minhwa*, focusing on symbolic motifs and interpretive traditions. Documents are segmented into semantically coherent chunks to preserve local contextual meaning while enabling robust extraction of symbolic relations. Corpus details are provided in Appendix B.

Semiotic Graph Schema The graph follows the layered representation introduced in Section 3,

consisting of Form, Concept, JointConcept, and Myth nodes. These represent increasing levels of abstraction in symbolic interpretation.

Directed relations connect these layers: CONNOTES links forms to concepts, IS_PART_OF links concepts to compositional structures, and GENERATES_MYTH links concepts or compositions to myth-level meanings.

LM-Based Symbolic Extraction An LM is used to extract symbolic entities and relations from each text chunk according to the predefined schema. The extraction process is constrained to focus on reusable symbolic knowledge rather than artwork-specific descriptions or provenance meta-data. Prompt details are provided in Appendix A.1.

Graph Assembly Extracted nodes and relations are merged into a global graph. Duplicate nodes are consolidated based on type and name, and redundant edges are merged to form a compact symbolic knowledge structure. The resulting graph serves as a structured retrieval resource for symbolic associations. Statistics are provided in Appendix C.

4.2 Semiotic Retrieval and Generation

Given an input image, symbolic interpretation is produced by retrieving relevant subgraphs from the semiotic KG and conditioning a multimodal LM on the retrieved context. The overall inference pipeline is illustrated in Figure 1.

Entity Extraction and Form Retrieval Visual entities are extracted from the input image using a multimodal LM. These entities are matched to Form nodes in the graph using similarity-based retrieval, which serves as the starting point for symbolic context retrieval. Prompt details are provided in Appendix A.2.

Graph Traversal and Pruning Starting from retrieved Form nodes, symbolic associations are obtained by traversing graph edges that connect visual forms to higher-level conceptual nodes. This produces candidate symbolic paths linking visual entities to culturally grounded meanings through intermediate signification structures.

Rather than treating this as a deterministic reasoning procedure, we frame graph traversal as a structured retrieval mechanism that assembles symbolic context for downstream generation.

Candidate symbolic contexts are ranked using aggregated evidence from supporting paths and normalized to reduce bias toward highly connected regions. The top-ranked symbolic subgraphs are retained after merging overlapping paths. Details are provided in Appendix D.

Interpretation Generation The retrieved symbolic subgraph is provided as structured context to a multimodal LM along with the input image. The model generates a natural-language interpretation conditioned on both visual and retrieved symbolic information. Prompt details are provided in Appendix A.3.

5 Experiments

We evaluate and compare RAG pipelines for symbolic interpretation of *minhwa*, focusing on how different retrieval structures influence culturally grounded meaning construction. The objective is not to optimize for a single correct description, but to study how semantic and symbolic retrieval affect visual grounding, interpretive composition, and cultural alignment in generated explanations.

5.1 Experimental Setup

We evaluate different retrieval and generation pipelines for symbolic interpretation of *minhwa* under a blind human-ranking protocol. Each system output is ranked across four criteria: (1) visual fidelity, (2) coverage, (3) interpretive soundness, and (4) cultural authenticity.

Each criterion is defined as follows: visual fidelity measures whether the generated description accurately reflects observable visual elements in the image; coverage measures how comprehensively the description accounts for salient visual and symbolic components; interpretive soundness measures the internal consistency and plausibility of inferred symbolic relationships; and cultural authenticity measures the extent to which the interpretation

aligns with established symbolic conventions and culturally grounded meanings in Korean *minhwa*.

We do not report standard automatic generation metrics (e.g., n-gram overlap or embedding-based similarity scores), as the task involves open-ended symbolic interpretation with multiple valid responses and no unique ground-truth description. In particular, evaluation requires assessing culturally grounded meaning, interpretive structure, and visual-symbolic alignment, which are not reliably captured by lexical or embedding similarity measures. We therefore adopt a human ranking protocol as the primary evaluation method.

We compare five systems:

- **Direct**: direct multimodal generation using gpt-4o-mini-2024-07-18 (OpenAI, 2024a)
- **DenseRAG**: flat dense retrieval using Faiss (Douze et al., 2026)
- **SemioticRAG**: the proposed graph-structured retrieval framework
- **SemioticRAG-YOLO**: the proposed graph-structured retrieval using YOLOv12-M for entity extraction (Tian et al., 2025)
- **ArtRAG**: multi-perspective art-description retrieval baseline (Wang et al., 2025)

For fair comparison, the ArtRAG baseline is reconstructed using the same *minhwa* corpus employed for SemioticRAG while following its original multi-perspective retrieval design. For SemioticRAG-YOLO, we train a YOLOv12-M model on a manually annotated *minhwa* object detection dataset consisting of 540 images. Training details are provided in Appendix E.

For all retrieval-based systems, we set the top- K Form retrieval limit to 1 per detected entity and limit the number of retained Myth candidates to 5. All experiments use seed 42 with a maximum generation length of 1024 tokens. Additional implementation details are provided in Appendix F.

The evaluation set consists of 55 images, yielding 275 generated outputs across the five systems. Human evaluation is conducted under a blind ranking protocol with 7 annotators (2 domain experts and 5 non-experts). Because the task involves inherently ordinal judgments of interpretive quality, we report ranking-based statistics and pairwise win rates rather than automatic generation metrics. Additional details on evaluation method are provided in Appendix G.

5.2 Evaluation Metrics

We report three complementary evaluation views: (1) overall ranking statistics based on mean rank, (2) criterion-specific average ranks for each evaluation criterion, and (3) pairwise win-rate comparisons between systems.

5.3 Quantitative Results

Table 1 reports the ranking results across all systems. DenseRAG achieves the best overall performance with the lowest mean rank (2.1250), followed by SemioticRAG, SemioticRAG-YOLO, Direct generation, and ArtRAG.

This ordering is stable across all evaluation criteria, including visual fidelity, coverage, interpretive soundness, and cultural authenticity. DenseRAG remains the top-ranked system in each criterion, while SemioticRAG consistently ranks second, indicating that graph-structured symbolic retrieval provides a uniform improvement over direct generation while not surpassing dense semantic retrieval. SemioticRAG-YOLO performs slightly below SemioticRAG, suggesting sensitivity to upstream entity extraction quality. Direct generation and ArtRAG consistently rank lower, with ArtRAG performing worst overall.

Pairwise comparisons in Figure 4 reinforce this pattern. DenseRAG wins most head-to-head comparisons across all systems, while SemioticRAG consistently outperforms both Direct generation and ArtRAG. Overall, the results indicate a stable performance hierarchy in which dense retrieval provides the strongest interpretive support, graph-based retrieval offers consistent intermediate gains, and non-retrieval or alternative retrieval design perform weakest.

5.4 Qualitative Analysis

The qualitative examples help explain why DenseRAG achieves the strongest aggregate ranking despite SemioticRAG’s explicit symbolic structure. DenseRAG retrieves source passages, preserving local qualifiers, co-occurring motifs, and alternative symbolic associations. SemioticRAG instead abstracts this information into canonical Form, Concept, JointConcept, and Myth nodes and conditions generation on a pruned set of connected symbolic paths. This makes higher-order composition explicit, but also makes retrieval more dependent on the accuracy of Form-level grounding and the validity of the constructed relations and

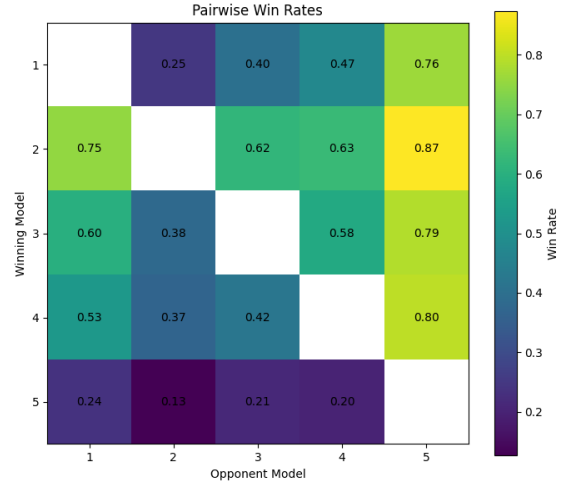


Figure 4: Pairwise win-rate matrix under blind human evaluation. Each cell indicates the proportion of comparisons in which the row system was preferred over the column system. System indices correspond to: (1) Direct, (2) DenseRAG, (3) SemioticRAG, (4) SemioticRAG-YOLO, and (5) ArtRAG.

selected paths. The following examples illustrate these differences under progressively more difficult conditions, from accurate grounding to upstream detection errors and construction-time graph errors.

Figure 7 illustrates the intended behavior of structured retrieval when the initial visual anchors are accurate. The LM-based extractor misidentifies the phoenixes and paulownia as cranes, pine, and bamboo, causing DenseRAG to produce a plausible but incorrectly grounded crane-and-pine interpretation. SemioticRAG partially recovers the phoenix from the original image, indicating that the final multimodal generator can supplement retrieved evidence, although the resulting interpretation remains comparatively generic. SemioticRAG-YOLO correctly identifies the phoenixes, paulownia, lingzhi, clouds, and moon, enabling more specific associations involving longevity, prosperity, immortality, harmony, and protection. This example shows that explicit symbolic traversal can support composition toward higher-order cultural meanings when the retrieved Forms are accurate. At the same time, the improvement begins with stronger visual grounding rather than graph structure alone.

Figure 8 examines a milder grounding error in which the intended pine is identified only as a generic tree. DenseRAG nevertheless produces pine-specific associations involving integrity, longevity, and noble character, suggesting that passage-level retrieval can remain useful under an

System	Fidelity	Coverage	Soundness	Authenticity	Mean	Median	Std Dev
Direct	2.9156	3.0325	3.0584	3.4416	3.1120	3.0	1.2166
DenseRAG	2.3052	1.9286	2.0714	2.1948	2.1250	<u>2.0</u>	1.1668
SemioticRAG	2.7532	<u>2.5000</u>	<u>2.7532</u>	<u>2.6039</u>	<u>2.6526</u>	<u>2.0</u>	1.3401
SemioticRAG-YOLO	2.8571	2.8766	2.9351	2.8831	2.8880	3.0	1.2900
ArtRAG	4.1688	4.6623	4.1818	3.8766	4.2224	5.0	1.1045

Table 1: Average human ranking results across all evaluation criteria. The table reports average ranks for visual fidelity, coverage, interpretive soundness, and cultural authenticity, together with overall average rank and standard deviation for each system. Lower rank indicates better performance.

underspecified but semantically related motif description. SemioticRAG stays closer to the generic tree anchor and extends it toward broader narratives involving reclusive life, social order, prosperity, and descendants. When the pine is correctly identified by YOLO, the graph-based variant retrieves more specific associations with permanence and moral virtue. This contrast suggests that dense retrieval is relatively tolerant of partial motif descriptions because the retrieved passages retain surrounding context and related associations, whereas SemioticRAG depends more directly on the selected Form node. Importantly, DenseRAG also combines the motifs coherently, so the distinction is better understood as contextual semantic synthesis versus more explicit pathway-based abstraction, rather than simply flat versus layered interpretation.

Figure 9 shows how an upstream grounding error can redirect graph-based interpretation. With LM-based extraction, SemioticRAG includes the rabbit that DenseRAG omits, but connects it to relatively weak associations involving instability, survival, and social hierarchy, showing that greater motif coverage does not necessarily imply a better symbolic interpretation. The clearer failure occurs in SemioticRAG-YOLO, where the rabbit is misclassified as a carp. This incorrect visual anchor redirects retrieval toward carp-related associations of ambition, success, and upward advancement, producing an internally coherent but visually misgrounded interpretation. The graph therefore does not create the original error; rather, it propagates and elaborates that error through culturally plausible symbolic relations. Once an incorrect motif becomes the starting Form, downstream coherence can make the resulting interpretation appear plausible even though the symbolic pathway is grounded in the wrong visual evidence.

Figure 10 illustrates a different failure mode in which the relevant visual motif is correctly identi-

fied but the graph contains an overgeneralized association. DenseRAG and SemioticRAG both receive peony, leaf, bud, and mountain from the LM-based extractor. DenseRAG remains centered on conventional peony symbolism and broader landscape associations, whereas SemioticRAG additionally describes the leaves as having "dynamic brushwork" and connects them to vitality and cultural identity. This stylistic claim is not supported by the image. It originates from corpus material describing dynamically painted leaves in a particular context, which LM-based KG construction converts into a reusable leaf-related association without preserving that original scope. The correct leaf anchor therefore activates an inappropriate relation in another painting. SemioticRAG-YOLO, which anchors retrieval only on peony in this example, avoids this particular error, exposing a construction-time graph failure rather than an upstream detection failure.

Taken together, these examples help explain DenseRAG’s aggregate advantage while clarifying what SemioticRAG contributes. Dense retrieval preserves passage-level qualifiers and alternative associations, making it relatively tolerant of underspecified motif descriptions. SemioticRAG provides more explicit composition from visual Forms toward higher-order cultural meanings, but introduces stronger dependence on both the selected anchors and the relations encoded during graph construction. Two distinct failure channels therefore emerge: (1) upstream grounding errors can redirect traversal, while (2) construction-time scope errors can generalize context-specific corpus associations into inappropriate reusable relations. SemioticRAG can consequently produce interpretations that are coherent and structurally explicit yet visually misgrounded or overgeneralized. The observed tradeoff is therefore between passage-level contextual tolerance and explicit but anchor- and graph-dependent symbolic composition.

6 Conclusion

We study RAG for symbolic folk-art interpretation by comparing dense and graph-structured retrieval. We introduce a semiotic retrieval formulation modeling relationships between visual motifs and culturally situated meanings using Form, Concept, JointConcept, and Myth structures.

Experimental results show that dense semantic retrieval remains the strongest overall approach, consistently outperforming graph-structured retrieval across evaluation criteria. SemioticRAG, however, produces more structured and layered symbolic compositions by traversing intermediate conceptual relationships, which can improve motif coverage and enable multi-step associations between visual elements and higher-level cultural narratives. This behavior is nevertheless sensitive to both initial entity detection quality and the connectivity of the symbolic graph, making it less robust under incomplete or noisy grounding.

Overall, the findings suggest that, within the setting studied here, explicit symbolic structure does not uniformly improve interpretive quality. Rather, structured retrieval primarily changes how meaning is composed, shifting interpretation from relatively flat aggregation of symbolic attributes toward pathway-based construction of culturally grounded associations.

Limitations

First, our study is limited to Korean *minhwa*, which constrains the cultural and symbolic scope of our findings. While *minhwa* provides a rich and structured setting for studying visual symbolism, it remains unclear how well the observed behaviors and retrieval dynamics generalize to other artistic traditions with different levels of symbolic codification or interpretive ambiguity.

Second, our evaluation relies on human judgments and pairwise comparisons, which, while informative, may introduce subjectivity and annotator bias. Although we report consistent trends across evaluators, finer-grained distinctions between systems may depend on task framing and evaluator familiarity with the domain.

Third, the comparison is restricted to a fixed set of baselines (Direct, DenseRAG, SemioticRAG, SemioticRAG-YOLO, and ArtRAG). While these systems capture both dense semantic and graph-based symbolic retrieval paradigms, the study does not explore other strong alternatives such

as instruction-tuned retrievers, multimodal fusion models, or RAG systems with learned routing strategies.

Finally, the semiotic KG construction depends on automated extraction using LMs, which introduces potential noise and incompleteness. However, our paper does not systematically quantify graph quality or isolate its effect on downstream performance, leaving open the extent to which performance differences are driven by graph construction versus retrieval strategy.

Ethics, Risks, and AI Assistance Disclosure

Generative AI assistants were used for language refinement and editing support during manuscript preparation. All technical designs, formal definitions, experimental settings, analyses, and interpretations were developed, verified, and validated by the authors. The authors take full responsibility for the content of this paper and its claims.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-22932973), the IITP (Institute of Information & Communications Technology Planning & Evaluation)–ITRC (Information Technology Research Center) grant funded by the Korea government (Ministry of Science and ICT) (IITP-2026-RS-2021-II211835), and the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.RS2026-25507979, AI Data Training Verification and Reliability Assurance Technology). We also appreciate the high-performance GPU computing support of HPC-AI Open Infrastructure via GIST SCENT.

We thank Professor Junghee Kim (School of Humanities and Social Sciences, GIST) and Professor Eunseok Lee (Department of Media and Visual Communications, Hannam University) for serving as domain-expert evaluators in our human evaluation. We also thank Professor Eunsung Song (Department of AI Convergence, GIST) for facilitating the participation of Professor Eunseok Lee.

References

Academy of Korean Studies. 2026. Encyclopedia of Korean culture. <https://encykorea.aks.ac.kr>. Accessed: 2026-05-25.

- Zechen Bai, Yuta Nakashima, and Noa Garcia. 2021. [Explain me the painting: Multi-topic knowledgeable art description generation](#). In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5402–5412, Montreal, QC, Canada. IEEE.
- Roland Barthes. 1972. *Mythologies*. Hill and Wang, New York, NY.
- Roland Barthes. 1977a. *Elements of Semiology*. Farrar, Straus and Giroux, New York, NY.
- Roland Barthes. 1977b. *Image-Music-Text*. Farrar, Straus and Giroux, New York, NY.
- Esther Cho. 2014a. [A study on the symbolic meaning of fish and crab painting in the second half joseon dynasty](#). *JANGSEOGAK*, (31):262–296.
- Un Han Cho. 2014b. [Folk painting, the study for meaning creation of magpies and a tiger](#). *Journal of The Korean Society of Illustration Research*, 15(40):69–78.
- Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2026. [The faiss library](#). *IEEE Transactions on Big Data*, 12(2):346–361.
- Umberto Eco. 1979. *A Theory of Semiotics*. Indiana University Press, Bloomington, IN.
- Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitansky, Robert Osazuwa Ness, and Jonathan Larson. 2025. [From local to global: A graph rag approach to query-focused summarization](#). *Preprint*, arXiv:2404.16130.
- Nicola Fanelli, Gennaro Vessio, and Giovanna Castellano. 2025. [ArtSeek: Deep artwork understanding via multimodal in-context reasoning and late interaction retrieval](#). *Preprint*, arXiv:2507.21917.
- Zirui Guo, Lianghao Xia, Yanhua Yu, Tu Ao, and Chao Huang. 2025. [LightRAG: Simple and fast retrieval-augmented generation](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 10746–10761, Suzhou, China. Association for Computational Linguistics.
- Yujin Han. 2015. [A study on the shamanistic aspects of folk painting : Focusing on driving out evil spirits and bringing in auspicious signs](#). Ph.D. thesis, Sungkyunkwan University.
- Doobin Im. 2013. [A study on spiritual beauty through making an aesthetic analysis of korean folk painting](#). *The Korean Journal of Art and Media*, 12(4):35–50.
- Yanbei Jiang, Krista A. Ehinger, and Jey Han Lau. 2024. [KALE: an artwork image captioning system augmented with heterogeneous graph](#). In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, Jeju, South Korea. ijcai.org.
- Soohyeong Kim, Seok Jun Hwang, JungHyouon Kim, Jeonghyeon Park, and Yong Suk Choi. 2025. [Re-GraphRAG: Reorganizing fragmented knowledge graphs for multi-perspective retrieval-augmented generation](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 5426–5443, Suzhou, China. Association for Computational Linguistics.
- Tae-Eun Kim. 2021. [The characteristics and intellectual background of jang hanjong’s \(1768-1815\) realistic fish and crab paintings](#). *Art History and Visual Culture*, (27):128–159.
- Taeho Lee. 2013. [A study on the popularity and realistic description of flowers, birds, insects and fishes paintings in the late joseon dynasty](#). *Art History & Cultural Heritages*, 2:115–148.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. [Retrieval-augmented generation for knowledge-intensive nlp tasks](#). In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474, Vancouver, Canada. Curran Associates, Inc.
- Costas Mavromatis and George Karypis. 2025. [GNN-RAG: Graph neural retrieval for efficient large language model reasoning on knowledge graphs](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 16682–16699, Vienna, Austria. Association for Computational Linguistics.
- National Folk Museum of Korea. 2026. Encyclopedia of korean folk culture. <https://folkency.nfm.go.kr>. Accessed: 2026-05-25.
- National Museum of Korea. 2026. emuseum. <https://www.emuseum.go.kr>. Accessed: 2026-05-25.
- Neo4j, Inc. 2024. Neo4j Aura: Fully Automated Cloud Graph Platform. <https://neo4j.com/docs/aura/>. Accessed: 2026-05-25.
- OpenAI. 2024a. GPT-4o mini: Advancing cost-efficient intelligence. <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/>. Accessed: 2026-05-25.
- OpenAI. 2024b. New embedding models and API updates. <https://openai.com/index/new-embedding-models-and-api-updates/>. Accessed: 2026-05-25.
- Charles Sanders Peirce. 1931–1958. *Collected Papers of Charles Sanders Peirce*, volume 1–8. Harvard University Press, Cambridge, MA.
- Leonardo Ranaldi, Barry Haddow, and Alexandra Birch. 2026. [Multilingual retrieval-augmented generation for knowledge-intensive question answering task](#). In *Findings of the Association for Computational Linguistics: EACL 2026*, pages 697–716, Rabat, Morocco. Association for Computational Linguistics.

- Leonardo Ranaldi, Federico Ranaldi, Fabio Massimo Zanzotto, Barry Haddow, and Alexandra Birch. 2025. [Improving multilingual retrieval-augmented language models through dialectic reasoning argumentations](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 9064–9085, Suzhou, China. Association for Computational Linguistics.
- Diego Sanmartin. 2024. [KG-RAG: Bridging the gap between knowledge and creativity](#). *Preprint*, arXiv:2405.12035.
- Ferdinand de Saussure. 2011. *Course in General Linguistics*. Columbia University Press, New York, NY.
- Yunjie Tian, Qixiang Ye, and David Doermann. 2025. [YOLOv12: Attention-centric real-time object detectors](#). In *Advances in Neural Information Processing Systems*, volume 38, pages 78433–78457, San Diego, CA. Curran Associates, Inc.
- Shuai Wang, Ivona Najdenkoska, Hongyi Zhu, Stevan Rudinac, Monika Kackovic, Nachoem Wijnberg, and Marcel Worring. 2025. [ArtRAG: Retrieval-augmented generation with structured context for visual art understanding](#). In *Proceedings of the 33rd ACM International Conference on Multimedia*, page 6700–6709, New York, NY, USA. Association for Computing Machinery.
- Xiaozhou You, Yahui Luo, and Lihong Gu. 2025. [MS-RAG: Simple and effective multi-semantic retrieval-augmented generation](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 22609–22625, Suzhou, China. Association for Computational Linguistics.
- Xiangrong Zhu, Yuexiang Xie, Yi Liu, Yaliang Li, and Wei Hu. 2025. [Knowledge graph-guided retrieval augmented generation](#). In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 8912–8924, Albuquerque, New Mexico. Association for Computational Linguistics.

A Prompt Details

This appendix provides the prompts used for KG construction, visual entity extraction, and symbolic interpretation generation.

A.1 KG Construction Prompt

The construction prompt was used to extract symbolic nodes and relations from corpus passages according to the predefined semiotic schema. The abridged prompt is provided in Figure 5.

A.2 Visual Entity Extraction Prompt

The detection prompt was used to extract visually salient entities from input artwork images for retrieval.

Given an image of a Korean Minhwa painting, identify the visible entities depicted in the scene and output them as a comma-separated list in Korean. Use specific names when clearly identifiable (e.g., "peony" instead of "flower"), otherwise use broader categories. Include only directly observable entities and avoid interpretations or descriptions.

A.3 Interpretation Generation Prompt

The generation prompt was used to produce symbolic interpretations conditioned on the retrieved symbolic context and the original image.

Given an image of a painting, explain the symbolism created by the visible elements together, focusing on how their combined presence and relationships produce meaning. Do not simply list separate symbolic meanings of each element. Write your response in Korean.

B Corpus Collection

The textual corpus used for semiotic KG construction was curated from Korean cultural heritage references and academic literature focusing on symbolic interpretation in Korean *minhwa*. Unlike conventional artwork-description corpora centered on artist metadata or historical provenance, the collected materials emphasize recurring symbolic motifs, vernacular beliefs, folk symbolism, and culturally situated interpretive traditions.

The primary reference sources include 145 entries from the *Encyclopedia of Korean Culture* (EncyKorea), maintained by the [Academy of Korean Studies](#) (2026), and 50 entries from the *Encyclopedia of Korean Folk Culture* (EKFC), maintained by the [National Folk Museum of Korea](#) (2026). EncyKorea article text was retrieved through the

official public API. Under the Academy of Korean Studies' content-use policy, article body text for works for which the Academy holds the relevant economic rights may be used without separate permission, with source attribution in the form "[항목명], 『한국민족문화대백과사전』." We used textual entry content only, not encyclopedia media, and retained entry-level source provenance during corpus processing.

EKFC encyclopedia text is provided under Korea Open Government License (KOGL) Type 2, which requires attribution and permits non-commercial use. The National Folk Museum of Korea specifies the attribution format "[출처: 표제어명-국립민속박물관 한국민속대백과사전]." This entry-level provenance was retained for EKFC-derived corpus chunks. Because no official API was available for EKFC at the time of collection, entries were collected from the encyclopedia webpages.

To supplement these cultural heritage references with domain-specific symbolic analyses, we additionally incorporated three prior academic studies on symbolic representation in late Joseon *minhwa* traditions (Lee, 2013; Cho, 2014a; Kim, 2021). Documents from all sources were segmented into semantic chunks prior to LM-based extraction of symbolic entities and relations according to the semiotic graph schema described in Section 4.

The corpus and derived semiotic KG were used solely for the experiments reported in this study and are not redistributed. Third-party source materials remain subject to their respective terms of use.

C Semiotic KG Statistics

The constructed semiotic KG contains heterogeneous symbolic nodes corresponding to Form, Concept, JointConcept, and Myth. Table 2 summarizes the overall graph statistics, and Table 3 summarizes the distribution of relation types.

Statistic	Count
Total nodes	3322
Total edges	3636
Form nodes	1354
Concept nodes	949
JointConcept nodes	478
Myth nodes	541

Table 2: Summary of the constructed semiotic KG.

Relation Type	Count
CONNOTES	1910
IS_PART_OF	1133
GENERATES_MYTH	593

Table 3: Relation type distribution in the semiotic KG.

D Coverage-Aware Myth Pruning

The traversal stage may produce a large number of overlapping symbolic paths and candidate Myth nodes. To reduce noise while preserving interpretive diversity, SemioticRAG performs coverage-aware bottleneck ranking over all retrieved paths.

For each candidate Myth, evidence is aggregated across all supporting traversal paths. Intermediary JointConcept nodes (or fallback Concept nodes when no intermediary exists) are treated as symbolic bottlenecks and scored using three signals:

- **support count** (c_s): the number of distinct supporting Concept nodes
- **support ratio** (r_s): the proportion of supporting concepts among neighboring concepts
- **confidence sum** (q_s): the summed retrieval confidence of originating source nodes

These signals are combined into a normalized bottleneck ranking score

$$s = \frac{0.5c_s + 0.3r_s + 0.2q_s}{\log(n + 2)}$$

where n denotes the number of neighboring concept nodes connected to the bottleneck node. The logarithmic normalization term penalizes overly connected bottleneck nodes that may otherwise dominate retrieval through graph degree alone.

To prevent over-pruning toward a small subset of highly connected myths, the algorithm additionally enforces a coverage constraint in which each retrieved source Form contributes at least one candidate Myth. This is implemented by identifying the highest-scoring Myth associated with each source node and assigning it a strong selection bonus during final ranking. Remaining Myth candidates then compete globally according to their bottleneck scores.

The final retrieved symbolic context consists of the highest-ranked Myth-connected subgraphs after path merging and deduplication.

E Object Detection Dataset

To support the SemioticRAG-YOLO variant, we constructed a custom object detection dataset for *minhwa* motifs and symbolic visual elements. The dataset consists of 540 manually annotated *minhwa* images with 178 object categories covering recurring symbolic forms such as trees, flowers, birds, fish, animals, mythical creatures, insects, fruits, landscape elements, and man-made objects.

All source images were obtained from the eMuseum digital collection operated by the [National Museum of Korea \(2026\)](#) through its official public API. For each image, the item-level Korea Open Government License (KOGL) status and owning institution were recorded. The downloaded source image files were retained unchanged, while bounding-box annotations were stored separately.

Bounding boxes were annotated manually based on visually identifiable symbolic entities relevant to folk-art interpretation. Because *minhwa* imagery contains long-tail symbolic motifs and highly uneven motif frequencies, the dataset exhibits substantial class imbalance across categories.

The final dataset contains 540 images and was divided into training, validation, and test splits using an 80/10/10 ratio with stratified splitting to maximize class coverage across splits, particularly for rarer symbolic categories.

The dataset was used solely for the non-commercial academic experiments reported in this study. Source images are not redistributed, and the original materials remain subject to the licensing and usage conditions specified by their respective source institutions.

Third-party artwork images reproduced in this paper retain their original licenses and are not covered by the license applied to the authors’ original material.

F Implementation Details

All LM-based stages, including KG construction, visual entity extraction, and symbolic interpretation generation, use gpt-4o-mini-2024-07-18 ([OpenAI, 2024a](#)). All dense retrieval operations use text-embedding-3-large embeddings ([OpenAI, 2024b](#)) with L2-normalized FAISS-based vector retrieval ([Douze et al., 2026](#)), while the symbolic KG is stored and queried using Neo4j Aura ([Neo4j, Inc., 2024](#)).

For retrieval-based systems, the top- K retrieval limit was set to 1 per detected entity, and the max-

imum number of retained Myth candidates after pruning was limited to 5. Generation was performed with temperature 0, top- p 1, seed 42, and a maximum output length of 1024 tokens.

The SemioticRAG-YOLO variant replaces LM-based entity extraction with a custom YOLOv12-M (Tian et al., 2025) detector trained on the object detection dataset described in Appendix E. The detector was trained using a Google Colab G4 instance equipped with an NVIDIA L4 GPU. Training was performed for 100 epochs using image size 1024. The optimizer was configured using the auto setting, which selected AdamW with learning rate 4.4×10^{-5} and momentum 0.9.

G Human Evaluation Protocol

Human evaluation was conducted under a blind ranking protocol in which all generated outputs were anonymized prior to annotation. For each image, annotators ranked outputs from all five systems (Direct, DenseRAG, SemioticRAG, SemioticRAG-YOLO, and ArtRAG) across four evaluation criteria: visual fidelity, coverage, interpretive soundness, and cultural authenticity. Ties were permitted.

Annotators were provided with written evaluation guidelines describing each evaluation criterion and were instructed to comparatively rank outputs rather than assign absolute quality scores. Evaluation was conducted through a Google Forms interface presenting the input image together with anonymized model outputs. The full English translation of the annotator instructions is provided in Figure 6.

The evaluation involved 7 annotators consisting of 2 domain experts and 5 non-experts. Domain experts were recruited through direct email requests for voluntary participation. Participants were not financially compensated. Each annotator evaluated 11 images sampled from the 55-image test split of the object detection dataset described in Appendix E. Because this split was held out during YOLO training, all systems—including SemioticRAG-YOLO—were evaluated exclusively on unseen images.

We report mean ranks averaged across annotations, criterion-specific average ranks for each evaluation criterion (see Table 1), and pairwise win-rates between systems (see Figure 4).

H Additional Evaluation Results

In addition to the merged result reported in Section 5.3, we also report expert and non-expert results separately in Tables 4 and 5 respectively.

The overall ranking is consistent across expert and non-expert evaluators. DenseRAG ranks first in both groups (2.0455 experts; 2.1568 non-experts), followed by SemioticRAG (2.5341; 2.7000), SemioticRAG-YOLO (3.1023; 2.8023), Direct (3.1364; 3.1023), and ArtRAG (4.1818; 4.2386). The same ordering is preserved for coverage and cultural authenticity. Criterion-level differences appear for visual fidelity and interpretive soundness. Non-experts rank SemioticRAG-YOLO above Direct for visual fidelity (2.7455 vs. 2.8818), whereas experts prefer Direct (3.0000 vs. 3.1364). For interpretive soundness, experts rank Direct second behind DenseRAG (2.7727 vs. 2.2273), while non-experts place SemioticRAG and SemioticRAG-YOLO ahead of Direct.

I Full Qualitative Examples

Table 6 summarizes the ground-truth motifs and detected entities for the qualitative examples discussed in Section 5.4. Each figure contains the input artwork image together with the generated outputs from all compared systems. For readability, the original Korean outputs are translated into English and lightly truncated while preserving the core symbolic interpretations and retrieval behaviors discussed in the analysis.

J API Cost Accounting

Table 7 summarizes the API usage and cost accounting for each stage of the pipeline. KG construction incurs a one-time cost C_{KG} over N_{chunk} chunks, while visual entity extraction, embedding generation, and interpretation generation are applied to each of the N_{image} images. Thus, the total inference cost over the dataset is $C_{\text{infer}} = C_{\text{VE}} + C_{\text{Emb}} + C_{\text{Gen}}$, corresponding to $3N_{\text{image}}$ API calls when each stage requires one call per image. The full pipeline cost is $C_{\text{total}} = C_{KG} + C_{\text{infer}}$. We additionally report the average per-image inference cost as $C_{\text{infer}}/N_{\text{image}}$ and the KG-amortized per-image cost as $(C_{KG} + C_{\text{infer}})/N_{\text{image}}$.

System	Fidelity	Coverage	Soundness	Authenticity	Overall	Std Dev
Direct	3.0000	3.0909	2.7727	3.6818	3.1364	1.1664
DenseRAG	2.0909	1.8182	2.2273	2.0455	2.0455	1.1337
SemioticRAG	<u>2.4091</u>	<u>2.4545</u>	2.8636	<u>2.4091</u>	<u>2.5341</u>	1.2950
SemioticRAG-YOLO	3.1364	3.0000	3.4545	2.8182	3.1023	1.2414
ArtRAG	4.3636	4.6364	3.6818	4.0455	4.1818	1.2914

Table 4: Average ranking results by domain experts. The table reports average ranks for visual fidelity, coverage, interpretive soundness, and cultural authenticity, together with overall average rank and standard deviation for each system. Lower rank indicates better performance.

System	Fidelity	Coverage	Soundness	Authenticity	Overall	Std Dev
Direct	2.8818	3.0091	3.1727	3.3455	3.1023	1.2386
DenseRAG	2.3909	1.9727	2.0091	2.2545	2.1568	1.1808
SemioticRAG	2.8909	<u>2.5182</u>	<u>2.7091</u>	<u>2.6818</u>	<u>2.7000</u>	1.3577
SemioticRAG-YOLO	<u>2.7455</u>	2.8272	2.7273	2.9091	2.8023	1.3018
ArtRAG	4.0909	4.6727	4.3818	3.8091	4.2386	1.0231

Table 5: Average ranking results by non-experts. The table reports average ranks for visual fidelity, coverage, interpretive soundness, and cultural authenticity, together with overall average rank and standard deviation for each system. Lower rank indicates better performance.

Image	Ground Truth	GPT Detection	YOLO Detection
Figure 7	Moon, cloud, paulownia, phoenix, peony, lingzhi	Crane, pine tree, bamboo, cloud, sun, rock	Moon, lingzhi, cloud, paulownia, phoenix, peony
Figure 8	Crane, pine	Crane, tree	Pine, crane
Figure 9	Peony, bird, rabbit	Peony, bird, rabbit, rock	Peony, bird, carp
Figure 10	Peony, bud	Peony, leaf, bud, mountain	Peony

Table 6: Ground-truth motifs and detected entities for qualitative examples.

Stage	API Calls	Cost Calculation
KG construction	N_{chunk}	$C_{\text{KG}} = (T_{\text{in}}^{\text{KG}} r_{\text{in}} + T_{\text{out}}^{\text{KG}} r_{\text{out}}) / 10^6$
Visual entity extraction	N_{image}	$C_{\text{VE}} = (T_{\text{in}}^{\text{VE}} r_{\text{in}} + T_{\text{out}}^{\text{VE}} r_{\text{out}}) / 10^6$
Embedding generation	N_{image}	$C_{\text{Emb}} = T_{\text{emb}} r_{\text{emb}} / 10^6$
Interpretation generation	N_{image}	$C_{\text{Gen}} = (T_{\text{in}}^{\text{Gen}} r_{\text{in}} + T_{\text{out}}^{\text{Gen}} r_{\text{out}}) / 10^6$
Recurring inference total	$3N_{\text{image}}$	$C_{\text{infer}} = C_{\text{VE}} + C_{\text{Emb}} + C_{\text{Gen}}$
Full pipeline total	$N_{\text{chunk}} + 3N_{\text{image}}$	$C_{\text{total}} = C_{\text{KG}} + C_{\text{infer}}$

Table 7: Stage-wise API accounting. T denotes billed tokens and r the applicable price per million tokens. Per-image inference cost is $C_{\text{infer}}/N_{\text{image}}$, while KG-amortized cost is $(C_{\text{KG}} + C_{\text{infer}})/N_{\text{image}}$.

You are an expert semiotician using Roland Barthes' model (Connotation → Myth).

From the passage, extract only:

- Forms: things that could be depicted in an artwork (characters, objects, scenes, settings, poses, motifs, etc.).
- Concepts: meanings directly connoted by those Forms.
- Myths: higher-level ideological meanings generated by combinations of Concepts.

Produce a JSON object with English-only keys and schema, containing two arrays: "entities" and "relations".

1. Canonicalization & Deduplication

- Use Korean canonical names, singular where applicable.
- One concept per idea: merge near-duplicates into a single canonical Korean name.
- Each "name" must be unique across all entities.

2. Scope (What to Include/Exclude)

- Include:
 - Forms: only generic, non-specific visual elements that could be staged or depicted (e.g., animal figure, human figure, fire, crowd, weapon, ritual space).
 - Concepts/Myths: only meanings directly attached to those Forms (e.g., Power, Corruption, Resistance, Survival Of The Weak).
- Exclude completely (no entities, no relations):
 - Artworks of any kind, including:
 - artwork titles
 - named works
 - mediums (painting, sculpture, photograph, installation, etc.)
 - styles, movements, schools
 - artists or creators
 - Artwork metadata, including:
 - dates
 - historical periods or dynasties
 - provenance, collections, museums
 - exhibition or publication history
 - Authors, scholars, critics
 - Locations and institutions (cities, museums, collections)

3. Hard Constraints

- For every Concept entity:
 - There must be at least one "Connotes" relation where:
 - "source" is a "Form"
 - "target" is this Concept's "name"
- For every Myth entity:
 - There must be at least one "Generates_Myth" relation where:
 - "target" is this Myth's "name"
 - "source_concepts" contains at least one existing "Concept" name
- All "source", "target", and "source_concepts" values must exactly match "name" values in "entities".
- "source_concepts" must be sorted alphabetically by Korean string order.

4. Expected JSON Structure

{...}

5. Output Language Constraint

- The JSON keys and schema must remain exactly as specified in English.
- All string values, including "name", "description", "source", "target", and "source_concepts", must be written in Korean only.
- Do not use English, romanization, or mixed-language strings anywhere in the output.

Figure 5: KG construction prompt.

Purpose

This evaluation is part of a research study aimed at developing an AI framework capable of understanding the cultural context of Korean minhwa, producing symbolically valid interpretations, and generating visually accurate descriptions.

Participants will review multiple AI-generated descriptions created for the same artwork and evaluate their quality according to the criteria below. The evaluation results will be used to improve AI systems that interpret traditional art in a more culturally appropriate and logically coherent manner.

Evaluation Guidelines

For each artwork, the original minhwa image, the collection information available through eMuseum, and five AI-generated descriptions will be presented together. The collection description is provided as reference material to aid understanding of the artwork, and some material or format-related information may be abbreviated for research purposes.

Participants should comparatively rank the five descriptions for each evaluation criterion. If multiple descriptions are judged to be of the same quality, identical rankings (ties) may be assigned. Each criterion should be judged independently. Please focus primarily on the quality of the content—such as interpretive accuracy, logical coherence, comprehensiveness, and cultural appropriateness—rather than sentence style or literary expression.

Evaluation Criteria

1. Visual Fidelity:

The degree to which the description accurately reflects what is actually depicted in the painting. Mentioned objects, figures, spatial relationships, actions, and contextual information should all be based on evidence visible in the image. Adding nonexistent elements, misidentifying subjects, or making unsupported inferences are factors that lower the evaluation.

2. Coverage:

The degree to which the description sufficiently addresses the important visual and symbolic components of the artwork. Major figures, objects, motifs, compositional relationships, and key elements contributing to the construction of meaning should be appropriately included.

3. Interpretive Soundness:

The degree to which the description logically and coherently explains how the symbolic meanings of individual elements depicted in the painting combine to form an integrated overall meaning, and how that meaning reflects the beliefs, values, aspirations, and ideals of the society of the time.

4. Cultural Authenticity:

A criterion evaluating how well the description aligns with the symbolic conventions, aesthetic traditions, and historical-cultural context of Korean minhwa. Interpretations should reflect meaning systems appropriate to the minhwa tradition and should not rely excessively on external cultural standards or irrelevant interpretive frameworks.

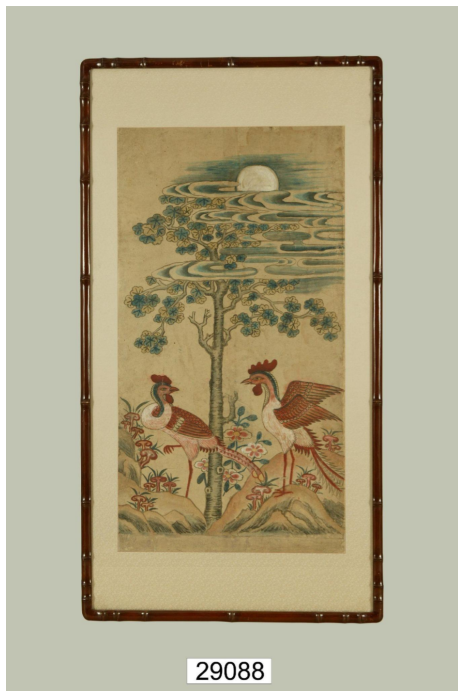
Evaluation Example

The following is a simple example intended to help explain the evaluation criteria. For example, suppose a minhwa painting contains both a tiger and a magpie.

- If the description mentions a dragon or mountain spirit that does not appear in the painting → visual fidelity is low.
- If the description mentions only the tiger and does not explain the magpie or the relationship between the two → coverage is low.
- If the tiger is interpreted as a symbol of authority and warding off evil, the magpie as a bearer of good news, and the two elements are logically connected to express the common people's wishes for peace and fortune at the time → interpretive soundness is high.
- If the interpretation aligns with symbolic meanings commonly recognized in Korean minhwa tradition → cultural authenticity is high.
- Conversely, if the interpretation applies Western mythology or modern popular culture frameworks without justification → cultural authenticity is low.

This evaluation aims to assess the accuracy, faithfulness, logical coherence, and cultural appropriateness of the interpretations rather than the literary quality or writing style of the descriptions themselves.

Figure 6: English translation of annotator instruction.



Direct: “This painting conveys harmony through flowers, birds, and sunlight. The tree symbolizes life and growth, while the birds represent love and peaceful co-existence. The sun adds vitality and hope, emphasizing the interconnectedness of nature and humanity.”

DenseRAG: “The painting centers on pine trees and cranes, both traditional symbols of nobility and longevity. The cranes reinforce ideals of harmony and happiness, while clouds and the moon evoke natural cycles and tranquility. Together, the motifs express an idealized vision of long life and moral virtue.”

SemioticRAG: “The image combines clouds, flowers, trees, the moon, and phoenixes into a symbolic vision of transcendence and natural beauty. Clouds evoke a supernatural realm, while the trees suggest reclusion and harmony with nature. The phoenix contributes auspicious meaning, producing an idealized mythic landscape grounded in traditional aesthetics.”

SemioticRAG-YOLO: “The paulownia tree, lingzhi mushrooms, phoenixes, and clouds form a symbolic composition centered on longevity and prosperity. Lingzhi signifies immortality, while the phoenix represents harmony and abundance. Together, the motifs construct an ideal world of peace, divine protection, and flourishing life.”

ArtRAG: “The painting presents cranes, trees, flowers, and the sun as symbols of longevity, prosperity, and harmony with nature. Their combination reflects traditional Korean ideals of blessing, vitality, and peaceful coexistence.”

Figure 7: Qualitative comparison for the phoenix composition. Image source: 장생도(長生圖), National Folk Museum of Korea (민속 29088), eMuseum, KOGL Type 1.



Direct: “The crane symbolizes longevity and wisdom, conveying elegance and peace. The tree represents harmony with nature, and its curved form symbolizes the journey of life. Together, the crane and tree suggest a harmonious life in nature and emphasize peace and wisdom.”

DenseRAG: “The old pine symbolizes integrity, longevity, and noble character. The crane, traditionally associated with longevity and happiness, gains stronger meaning when paired with the pine. Together, the pine and crane symbolize harmony between humanity and nature and express an ideal state of life grounded in dignity and permanence.”

SemioticRAG: “The tree functions as a background for reclusive life and symbolizes an ideal existence in harmony with nature. The crane conveys auspiciousness and longevity. Their combination reflects how longevity symbolism appears in late Joseon court painting and connects to wishes for prosperity, social order, and flourishing descendants. Together, these motifs construct a mythic ideal of harmonious life.”

SemioticRAG-YOLO: “The pine tree symbolizes Taoist immortality, permanence, and moral virtue, while the crane represents longevity and peace. Together, the two motifs form a symbolic mythology of immortality and protection, expressing an idealized noble life removed from worldly concerns.”

ArtRAG: “The crane and pine tree jointly symbolize longevity and continuity. The pine represents enduring vitality, while the crane signifies happiness and long life. Together they express the traditional Korean ideal of harmony between humanity and nature.”

Figure 8: Qualitative comparison for the crane and pine composition. Image source: 장생도(長生圖), National Folk Museum of Korea (민속 29080), eMuseum, KOGL Type 1.



Direct: “The flowers symbolize life and prosperity, while the pair of birds represent happiness and harmony. The rabbits below symbolize purity and fertility, reinforcing themes of vitality and hope. Together, these motifs express harmony between humanity and nature.”

DenseRAG: “The painting combines peonies, birds, and scholar’s rocks into a harmonious symbolic composition. Peonies symbolize wealth and royal dignity, the birds convey hope and continuity of life, and the rocks symbolize stability and longevity. Together, these motifs express abundance and natural harmony.”

SemioticRAG: “The peonies symbolize wealth and splendor, while the birds function as auspicious and protective symbols. The rabbits introduce themes of instability and survival, and their relationship with the birds reflects social hierarchy and the complexities of human society. Together, these motifs construct a symbolic mythology of prosperity, protection, and the instability of life.”

SemioticRAG-YOLO: “The peonies symbolize prosperity, while the birds convey hope, fortune, and protection from misfortune. The carp below symbolizes ambition, success, and upward advancement, reinforcing a broader symbolic vision of prosperity and achievement.”

ArtRAG: “The painting combines flowers, birds, and rabbits to symbolize abundance and happiness. The flowers represent beauty, the birds express harmony, and the rabbits symbolize prosperity. Together, these motifs emphasize natural harmony and familial well-being.”

Figure 9: Qualitative comparison for the peonies, birds, and rabbits composition. Image source: 8곡 민화(화조도), Jeonju Museum of History (기증 10067), eMuseum, KOGL Type 4.



Direct: “The painting emphasizes flowers and leaves as symbols of vitality, harmony, and renewal. The large blossoms suggest abundance and prosperity, while the buds imply hope and new beginnings. Together, the composition conveys the beauty and emotional calm of nature.”

DenseRAG: “The image centers on peonies, traditional symbols of wealth, prosperity, and aristocratic status in Joseon art. The flowers and surrounding leaves express vitality and flourishing life, while the rocky landscape reinforces stability and harmony between nature and humanity. The composition reflects broader ideals of cultural refinement and national prosperity.”

SemioticRAG: “The combination of peonies, mountains, and leaves constructs a mythic image of prosperity and Korean beauty. The peonies symbolize wealth and splendor, while the mountain backdrop evokes grandeur and royal authority. The dynamic brushwork of the leaves further suggests vitality and cultural identity, producing a more expansive symbolic interpretation.”

SemioticRAG-YOLO: “The peonies symbolize wealth, prosperity, and familial flourishing. Their lush petals and surrounding leaves express abundance and the ideal of a prosperous life. The composition frames the peony not only as a decorative flower, but as a symbolic image of enduring fortune and happiness.”

ArtRAG: “The painting centers on peonies as symbols of wealth and honor, while the surrounding leaves represent continuity and prosperity. Together, these motifs convey domestic happiness and traditional values through a harmonious natural composition.”

Figure 10: Qualitative comparison for the peony-centered composition. Image source: 牡丹圖, Tongyeong City Museum (통영시립 160), eMuseum, KOGL Type 4.