PaintEcho: Bringing Paintings to Photos through Cross-domain Visual Exploration

Yihan Gao, Tan Tang, Jianing Yin, Haobo Zheng, Tianyi Chen, Lu Ying, Yanhong Wu, Yingcai Wu

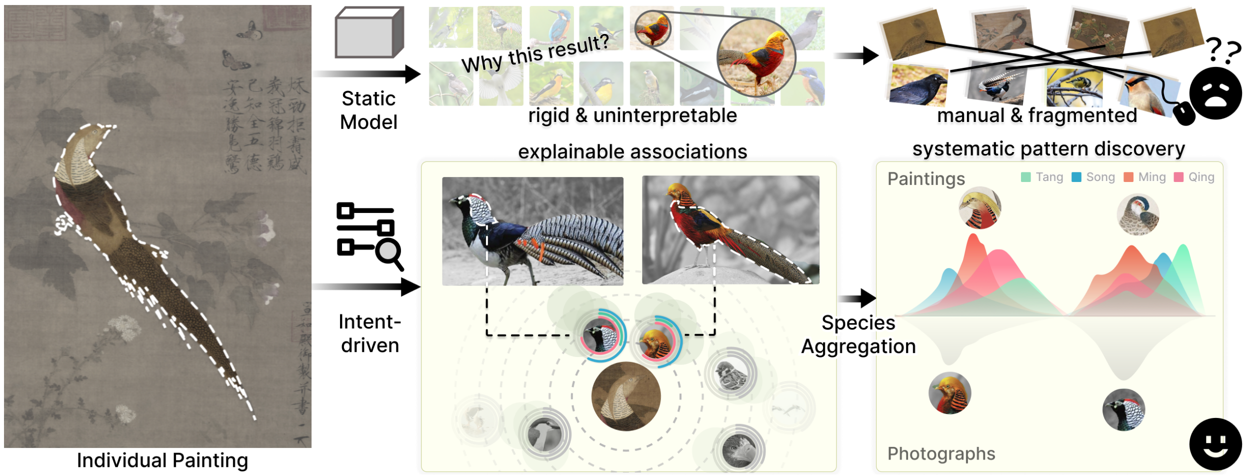

Figure 1: PaintEcho facilitates cross-domain exploration between paintings and photographs. Contrasted with the rigid and manual process of the traditional approach (top), PaintEcho provides an intent-driven workflow for multi-level exploration (bottom). The system leverages a radial layout for local species identification and a "mountain-reflection" metaphor for global pattern discovery.

Abstract— Exploring associations between paintings and photographs has become a vital component of art history research, as it helps reveal the evolution of visual patterns and drive interdisciplinary discovery. However, establishing these cross-domain associations remains difficult due to the stylistic gap between artistic expression and photographic realism, coupled with the multifaceted nature of visual attributes. Existing automated approaches often lack accuracy and interpretability, while their rigid outputs fail to accommodate the flexible, multi-dimensional exploration required by domain experts. Addressing these limitations poses three primary challenges: (a) the mismatch between unified representations and diverse analytical needs, (b) the difficulty of identifying relevant species from large cross-domain candidates, and (c) the synthesis of large-scale artistic insights from fragmented matches. To address these challenges, we first introduce a cross-domain visual association model that disentangles multifaceted attributes, enabling intent-driven feature steering. Building on this, we develop PaintEcho, a visual analytics system that enables multi-level exploration to facilitate the systematic discovery and interpretive analysis of cross-domain associations. Specifically, it employs a query-centric radial layout for local taxonomic organization and a "mountain-reflection" metaphor for global pattern discovery. We validate the effectiveness and usability of our approach through an algorithmic performance analysis, two use cases, and a task-based user study. Notably, PaintEcho enables experts to identify elusive species and uncover systematic stylistic shifts across historical masterpieces.

Index Terms—Visual analytics, cross-domain image association, intent-driven analytics, digital humanities, art history

1 Introduction

Exploring associations between paintings and photographs is a recurring task in art-historical research, as it helps reveal artistic transformation, trace the evolution of visual patterns, and drive interdisciplinary discovery [2,24,26,34,39,41]. For instance, by linking the 12th-century naturalistic masterpiece *Hibiscus and Golden Pheasant* with modern bird photography, researchers discovered a long-hidden scientific fact that the depicted bird is a hybrid [29]. The digitization of artistic heritage and modern photography has enabled systematic, large-scale exploration of these image domains, thereby facilitating such discoveries. However, establishing such cross-domain associations remains difficult because the stylistic gap between artistic expression and photographic realism makes it hard to identify meaningful matches. This difficulty is further compounded by the diversity of visual attributes, such as structure, color, and texture, which creates a vast search space.

Traditionally, art historians identify these associations through image-by-image comparison, leveraging their personal domain knowledge [26]. However, this manual process is time-consuming and labor-intensive for analyzing large-scale image collections. With the rapid progress in computer vision, Cross-domain Image Retrieval (CDIR) algorithms [25,38,43] now enable the automated discovery of such associations. Nevertheless, existing models often yield inaccurate results due to the domain gap and operate as "black boxes", failing to explain why specific paintings and photographs are linked. As a result, domain experts with limited machine learning expertise may struggle to trust these models and effectively validate the associations they generate. Furthermore, existing models typically provide rigid outputs that do not

• *Y. Gao, J. Yin, H. Zheng, L. Ying and Yingcai Wu are with State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China. E-mail: {yihangao, yinjianing, haobozheng, yingluu, ycwu}@zju.edu.cn.*
• *Y. Gao, T. Tang and T. Chen are with Laboratory of Art and Archaeology Image, Zhejiang University, Hangzhou, China. E-mail: {tangtan, chentianyi}@zju.edu.cn.*
• *Yanhong Wu is with Hithink RoyalFlush Information Network Co., Hangzhou, Zhejiang, China. E-mail: wuyanhong@myhexin.com.*
• *Tan Tang is the corresponding author.*

Manuscript received xx xxx. 202x; accepted xx xxx. 202x. Date of Publication xx xxx. 202x; date of current version xx xxx. 202x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.202x.xxxxxxx

support the flexible, multi-dimensional exploration required by domain experts, thereby limiting deeper discovery.

To address these limitations, we collaborate closely with art historians to develop a visual analytics system for the systematic discovery and interpretive analysis of associations across distinct image domains. Our approach integrates multifaceted feature disentanglement with a painting appreciation canon, enabling experts to interactively explore visual associations across large-scale collections. To develop this visualization approach, we need to address three challenges:

C1. Mismatch between a unified representation and diverse analytical needs. Current cross-domain retrieval frameworks typically rely on fixed embedding spaces learned by general-purpose models. However, experts' analytical needs in painting-to-photo analysis are often diverse and evolving rather than fixed. At different stages of exploration, they may attend to different visual attributes, such as structural composition, brushwork, or color usage. A unified representation cannot flexibly support these shifting analytical perspectives, making it difficult to reveal subtle artistic associations.

C2. Identifying relevant species from large cross-domain candidates. For a stylized painting object, experts often examine retrieved real-world photographs to identify relevant species. Here, species identification refers to the selection of plausible visual references at the species-level, rather than the strict biological classification. However, automated models typically return large candidate sets containing both spurious matches caused by the domain gap and redundant images of the same species. As a result, experts must repeatedly compare and cross-reference candidates to identify the most plausible species, which imposes a cognitive burden. This comparison process is poorly supported by conventional layouts, such as linear lists or scatter plots. A better solution should organize candidates by species and visual similarity, enabling effective species identification.

C3. Deriving global artistic insights from fragmented matches. Beyond item-level associations, the analytical focus shifts to uncovering global artistic patterns at scale. Nevertheless, deriving global insights from such fragmented, disjointed matches poses a substantial challenge. This requires aligning photographic collections of identified species with painting sets to reveal distribution laws across multiple facets, such as period and authorship. For example, while Song-dynasty plum blossoms mirror photographic realism, later "ink plums" prioritize calligraphic style over biological accuracy [19]. Consequently, a novel visual tool is needed to synthesize these fragmented records into a coherent narrative of stylistic evolution.

To address these challenges, we propose PaintEcho, a visual analytics system that supports cross-domain analysis between paintings and real-world photographs. For the first challenge, we introduce a cross-domain visual association model that disentangles multifaceted attributes, allowing users to interactively steer feature weights based on dynamic analytical intents. Regarding the second challenge, we develop a query-centric radial layout that clusters photograph candidates by species and encodes cross-domain similarity via radial distance, facilitating efficient species identification. To overcome the third challenge, we propose a "mountain-reflection" metaphor that transforms scattered data into density-based distributions, revealing a coherent stylistic narrative. We evaluate PaintEcho through model experiments and a user study, demonstrating its utility via two use cases.

Our contributions are summarized as follows:

- We collaborate with art historians to understand their workflows and requirements for associating paintings with photographs, then develop a cross-domain visual association model that disentangles visual attributes and enables intent-driven feature steering.
- We propose a novel "mountain-reflection" visual metaphor and develop a visual analytics system for multi-level cross-domain image exploration, namely PaintEcho, to facilitate associative analysis of paintings and photographs.
- We validate the effectiveness of PaintEcho through model experiments, two use cases, and an in-lab comparative user study.

2 Related Work

This section reviews prior research across two primary areas: cross-domain image retrieval and analysis, and visual analytics for image collections.

2.1 Cross-domain Image Retrieval and Analysis

Establishing associations between paintings and photographs requires both CDIR for domain alignment and multi-faceted analysis for diverse analytical needs. Cross-Domain Image Retrieval (CDIR) is a fundamental task that matches similar images across different domains, such as between paintings and realistic photographs. Traditional methods [22, 32, 36] rely heavily on annotated data, limiting their scalability for real-world applications. To bypass manual annotations, recent Unsupervised Cross-Domain Image Retrieval (UCDIR) methods have progressed from basic feature alignment toward refined structural and distributional modeling. Specifically, some methods employ correspondence-free domain alignment [43] and distance-of-distance loss [20] to bridge domain gaps. ProtoOT [25] uses optimal transport to unify intra-domain consistency. To enhance practical utility, recent models further incorporate style generalization for unseen domains [21] and privacy protection for secure retrieval [38]. However, these models typically rely on a singular, rigid similarity metric, falling short of the dynamic and diverse analytical needs of domain experts. In addition, multi-faceted image analysis, often realized through alternative image clustering, aims to extract varied structures from the same dataset to provide diverse analytical perspectives. The advancement of vision-language models (VLMs) has paved the way for text-guided mechanisms, enabling semantic-driven control over the clustering process. For instance, ITGC [51] employs an iterative search to achieve interpretable clustering results via semantic alignment. Similarly, TGAICC [37] utilizes user-specified prompts and consensus clustering to discover diverse, customizable groupings. Nevertheless, these text-driven methods often rely on discrete labels, which struggle to exhaust the multifaceted dimensions of art and capture its complex visual features. In contrast, our method supports flexible and interpretable cross-domain visual association, enabling experts to steer the analysis according to diverse analytical needs.

2.2 Visual Analytics for Image Collections

Visual analytics approaches have become important tools for exploring large-scale image collections, helping users interpret complex patterns in massive image datasets. Existing approaches can generally be divided into two categories: attribute-based methods that rely on faceted metadata and image-based methods that leverage visual features. The first category has evolved through three distinct stages. Initially, studies focused on establishing the foundational principles of multi-dimensional faceted search [49]. Systems such as MediaTable [12] and ICLIC [40] refined interactive categorization through the use of progressive menus. Subsequent studies tailored the organization of image collections to professional contexts; for instance, VIStory [13] and VISImageNavigator [10] employ extrinsic metadata such as authorship and publication venue to guide academic exploration. In the third stage, recent advancements [33, 46] incorporate vision-language models to transform visual styles into filterable semantic attributes, uncovering deeper data relationships. However, attribute-based methods still struggle with cross-domain analysis due to inherently incompatible metadata schemas. In the second category, deep learning has revolutionized feature extraction by encoding images into high-dimensional embeddings, establishing the essential semantic foundation for modern image-based visual analytics. EmbeddingProjector [35] and CDR [44] leverage dimensionality reduction to project images into 2D scatterplots, facilitating interactive exploration and the refinement of visual clustering. To address scalability, DendroMap [3] utilizes hierarchical treemaps for multi-level abstraction, while graph-based frameworks [1, 9, 31] leverage similarity-linked networks to enable non-linear navigation. Furthermore, to bridge the gap between visual features and high-level concepts, several systems [14, 17, 45, 48] employ cross-modal embeddings, caption-derived keywords, and taxonomic representations to support concept-driven navigation. Despite substantial progress in both categories, existing methods are confined to single-domain exploration. Consequently, they lack robust support for cross-domain relational alignment and comparative analysis. PaintEcho addresses this shortfall by enabling domain experts to visually contrast distributions and identify latent associations between distinct image collections.

3 BACKGROUND

This section presents the problem formulation for associating paintings with photos, the design requirements, and the data characteristics.

3.1 Problem Formulation

To better understand the problem and its domain context, we closely collaborated with four domain experts (E1–E4). E1, E2, and E3 are senior art historians with expertise in painting authentication, appreciation, and historical research, respectively. E4 is an expert in information visualization with an interdisciplinary background in computer science and ancient painting research. We held regular meetings and in-depth interviews with these experts over four months, supplemented by on-site visits to observe their traditional workflows. From these interviews and observations, we formalized the associative analysis between paintings and real-world photographs as a three-stage workflow:

Stage 1. Picking objects and clarifying analysis focuses. In this stage, experts first identify the specific object of interest among the multiple elements depicted in a painting. Owing to color degradation [23] and blurred boundaries, interactive segmentation is required to ensure precise extraction where automation fails. The subsequent analysis is guided by the experts' diverse intents, which diverge according to their research focus. For example, E1 prioritizes anatomical features, such as beak curvature, to verify a bird's biological identity. Meanwhile, E2 focuses on brushwork textures, such as the thickness and pressure of ink lines, to analyze the artist's unique painting technique. Hence, this stage aims to translate these high-level analytical preferences into a structured search configuration, enabling intent-based image retrieval.

Stage 2. Associating photos and identifying related species. Building upon the extracted instances and specified intents, experts navigate photographic datasets to establish cross-domain associations. They search for matching photos and pinpoint relevant species by comparing visual features with the painting instance. However, as E3 noted, art historians often lack the biological expertise required for accurate species recognition, making this process slow and labor-intensive. Accordingly, this stage focuses on enabling experts to efficiently filter candidates and pinpoint target species through a visual organization.

Stage 3. Aggregating collections and uncovering global patterns. Once the target species has been identified, experts shift their focus from individual cases to broader artistic themes. As E1 and E2 emphasized, art history research often extends beyond single works to trace how recurring motifs evolve across painting collections, which is essential for uncovering global patterns across different periods and regions. Traditionally, experts rely on their domain expertise to manually collect and group paintings. Nevertheless, as E4 highlighted, this piecemeal approach is error-prone and often masks broader patterns of visual evolution. Therefore, this stage aims to facilitate the systematic aggregation of painting data around confirmed species, enabling the discovery of global patterns.

3.2 Requirement Analysis

To identify the specific requirements for associating paintings with photographs, we synthesized findings from our literature review and expert feedback. Guided by the proposed three-stage workflow, we derived six primary requirements to guide the system design.

R1. Extract instances from paintings (S1). Experts require the unit of analysis to be individual instances (e.g., a bird or a flower) rather than the entire painting to ensure object-level comparative precision. Therefore, they need to extract clean instances from paintings to avoid background noise during cross-domain matching.

R2. Specify analytical intents for targeted discovery (S1). Experts need to specify analytical intents (e.g., a principle of "Six Laws of Xie He" [6]) to align system logic with their research focus. This ensures that the results are grounded in the expert's research context and can be directly interpreted in relation to their research goals.

R3. Support hierarchical exploration across analytical scales (S2, S3). To navigate extensive datasets, experts require multi-level exploration of cross-domain associations. The system should support a seamless transition from instance-level species identification (S2) to collection-level pattern discovery (S3), enabling experts to move from individual cases to broader comparative insights.

R4. Organize photographic candidates by species and similarity (S2). Experts require retrieved photographs to be organized by taxonomic categories and visual similarity so that they can efficiently navigate large-scale candidate sets. Compared with flat-list browsing, this structured organization reduces cognitive overhead.

R5. Provide rationale for image similarity (S2). Experts require interpretable evidence to understand why the system considers a photograph similar to the painting instance. Revealing relevant feature correspondences enables them to assess whether model-generated suggestions are consistent with their professional judgment.

R6. Align cross-domain distributions in a unified overview (S3). Experts seek a unified overview to align and compare the distributions of painting motifs and real-world photographs. Such distributional alignment allows experts to identify global patterns that are often obscured at the individual case level.

3.3 Data and Approach Overview

To facilitate cross-domain association, we use two primary data sources: naturalistic paintings and real-world photographs. **The painting dataset** comprises 5054 high-resolution paintings, sourced from Zhonghua Zhenbao Guan [52]. Each instance is enriched with detailed metadata, such as artist attribution and creation periods. **The photographic dataset** consists of 11406 real-world photographs curated from the Global Biodiversity Information Facility (GBIF) [15] and Bing [27]. It encompasses 101 distinct biological species, with each photograph accompanied by ground-truth species-level labels provided by the original data sources. Based on those datasets, we first develop computational models and then propose novel visualization designs to support painting queries, image association analysis, and image collection exploration (see Fig. 2).

4 DATA PROCESSING AND MODELS

This section introduces how we transform raw paintings into structured queries, align cross-domain features in calibrated subspaces, and map semantic associations through multi-scale visualizations.

4.1 Interactive Query Formulation

We adopt an interactive approach for query formulation, which transforms raw painting images into structured queries. This process consists of two steps: extracting the target instance from the painting and quantifying the user's analytical intents into measurable parameters.

4.1.1 SAM-based Instance Extraction

Extracting clear instances from paintings (**R1**) is challenging, as the visual features are often compromised by centuries of fading and wear [23]. To address this, we employ Segment Anything with Concepts (SAM 3) [7] to isolate target objects from complex backgrounds. SAM 3 enables concept-driven localization, allowing users to identify specific objects via text prompts (e.g., "yellow flower") even in degraded paintings. For higher precision, users can iteratively refine the segmentation through positive and negative clicks. Specifically, positive clicks act as foreground prompts to capture missing details, while negative clicks serve as background constraints to exclude irrelevant clutter. This process generates high-fidelity query instances, serving for subsequent cross-domain image retrieval.

4.1.2 Analytical Intent Parameterization

Beyond instance extraction, users can provide textual descriptions of their analytical intents (**R2**). Then we bridge the semantic gap by quantifying abstract user intents into three measurable parameters: **Structure**, **Color**, and **Texture**. These dimensions are grounded in

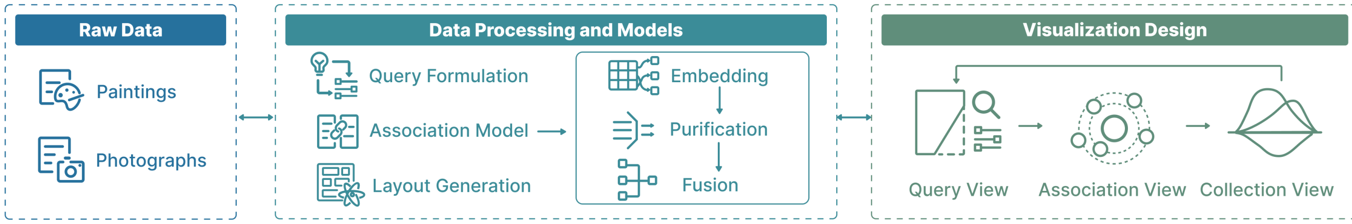

Figure 2: PaintEcho Overview: Data comprises collections of paintings and photographs. The backend performs query formulation, cross-domain association, and layout generation. The interface integrates the query view, association view, and collection view for multi-level exploration.

the "Six Laws of Xie He" [6], a seminal canon for painting evaluation. Specifically, we map the *Conformity with the object* (depiction of forms) to **Structure**, *Suitability to type* (application of color) to **Color**, and *Bone Method* (use of the brush) to **Texture**. Other principles, such as those regarding metaphysical spirit or global layout, are excluded as they are extrinsic to the visual characteristics of specific instances. To implement this parameterization, we utilize the Llama-3.3-70b-versatile model [16] to translate user-provided descriptive text into an initial weight configuration $\mathbf{w} = \{w_s, w_c, w_t\}$. This automatic initialization reduces the effort of configuring the weights while still allowing manual refinement, ensuring a systematic transition from abstract intent to quantifiable parameters. During retrieval, the three features serve as weighted dimensions: **Structure** mainly captures object contours and morphology, **Color** captures color distributions, and **Texture** captures brushwork or local surface patterns.

4.2 Cross-Domain Visual Association Model

We develop a cross-domain visual association model for the core retrieval task. This model operates in two stages: we first calibrate structure-, texture-, and color-specific representations into domain-invariant subspaces to reduce the gap between paintings and photographs; we then fuse these calibrated similarities using the intent weights to produce the final retrieval score.

4.2.1 Domain-Invariant Subspace Calibration

Traditional cross-domain methods often rely on a unified global embedding, which typically overlooks fine-grained details. To address this, we leverage the hierarchical architecture of DINOv2 (ViT-L/14) [28] to capture multi-level visual semantics by mapping intent-related dimensions to specific network layers. We extract structure (v_s), texture (v_t), and color (v_c) features from the deepest (layer 24), intermediate (layer 13), and shallowest (layer 1) blocks, respectively. This layer-wise strategy preserves the unique traits of each dimension, effectively capturing geometric topology, brushwork patterns, and color distributions.

However, these raw feature vectors v_d ($d \in \{s,t,c\}$) remain confounded by irrelevant factors, requiring a transformation to isolate domain-invariant semantics. Inspired by the calibrated projection framework originally proposed for mitigating social biases in Vision-Language Models [11], we generalize this methodology to cross-domain image association. For each visual dimension d, we formulate a cross-domain projection operator P_d^* to filter out the media-induced biases between paintings and photos. This process is executed in two stages:

Subspace-based Feature Purification. For the target dimension d, we use a bias matrix $X_{\text{bias},d} \in \mathbb{R}^{D \times N}$ to represent features that are irrelevant to d, with these features extracted from media-specific noise (e.g., blank aged silk) and exemplars of non-d dimensions. To identify the primary axes of these irrelevant features, we perform truncated SVD [18] on $X_{\text{bias},d} = U_d \Sigma_d V_d^\top$ and retain the top-k left singular vectors, denoted as $U_{d,k} = [u_1, u_2, \ldots, u_k] \in \mathbb{R}^{D \times k}$. These vectors span the bias subspace. We then define an initial projection operator P_d onto the orthogonal complement of this bias subspace:

$$P_d = I - U_{d,k} U_{d,k}^\top \tag{1}$$

Applying this operator to the raw feature v_d yields a purified vector $v'_d = P_d v_d$, which suppresses interference from irrelevant factors and preserves information relevant to the target dimension.

Exemplar-driven Domain Alignment. While purification removes noise, the domain gap remains unaddressed. To achieve domain invariance, we refine the initial projector P_d using a calibration set $\mathscr{E} = \{(v_{p,i}^{(d)}, v_{r,i}^{(d)})\}_{i=1}^n$ with $n = 15$ cross-domain pairs. Here, $v_{p,i}^{(d)}$ and $v_{r,i}^{(d)}$ serve as positive pairs representing the painting and real-world photo instances of the same visual entity, guiding the model to bridge the stylistic gap. To enforce this alignment, we define an optimal operator P_d^* that minimizes a joint loss function:

$$\min_{P_d^*} \frac{\lambda}{n} \sum_{i=1}^{n} \|P_d^* v_{p,i}^{(d)} - P_d^* v_{r,i}^{(d)}\|^2 + \|P_d^* - P_d\|_F^2 \tag{2}$$

The first term of Equation 2 enforces semantic consistency by minimizing the mean squared distance between projected exemplar pairs. The second term acts as a proximity constraint that ensures P_d^* retains the purification effects established in the previous stage. Since the objective function is quadratic, it admits a highly efficient closed-form solution, enabling near-instantaneous computation of the optimal operator [11]:

$$P_d^* = P_d \left(I + \frac{\lambda}{n} \sum_{i=1}^{n} (v_{p,i}^{(d)} - v_{r,i}^{(d)})(v_{p,i}^{(d)} - v_{r,i}^{(d)})^\top \right)^{-1} \tag{3}$$

This analytical solution enables efficient few-shot calibration, making the method suitable for interactive use. We then apply the calibrated projector P_d^* to obtain $v''_d = P_d^* v_d$, which aligns painting and photo features within a shared domain-invariant subspace.

4.2.2 Multi-dimensional Feature Fusion

After obtaining the calibrated representations v''_d for structure (s), texture (t), and color (c), we synthesize these independent signals into a unified similarity metric. Applying the modeled weights w, the final cross-domain similarity S between a painting x_p and a photograph x_r is defined as:

$$S(x_p, x_r) = \sum_{d \in \{s,t,c\}} w_d \cdot \text{sim}(v''_{p,d}, v''_{r,d}) \tag{4}$$

where $\text{sim}(\cdot,\cdot)$ denotes the cosine similarity. This formulation ensures that the association is strictly grounded in the users' exploration goals.

4.3 Multi-Scale Visual Mapping

To facilitate cross-domain analysis, we propose two specialized visual mapping strategies: a query-centric radial layout for instance association and a mirrored distribution alignment for global comparison.

4.3.1 Query-Centric Radial Layout

This layout maps calibrated feature vectors into a polar coordinate system (R, θ) centered on the query painting (Fig. 4-c). For each photo instance $x_{r,i}$, its radial distance is determined by its similarity S_i to the query. To mitigate visual overcrowding and emphasize high-relevance results, we apply an exponential scaling to the radius: $R_i = \beta(\exp(\alpha(1 - S_i)) - 1)$. The angular coordinate θ_i encodes the categorical topology of the gallery. We derive angles from centered t-SNE coordinates: $\theta_i = \text{atan2}(ty_i - \bar{ty}, tx_i - \bar{tx})$. Because t-SNE preserves the local manifold structure, this mapping translates latent semantic

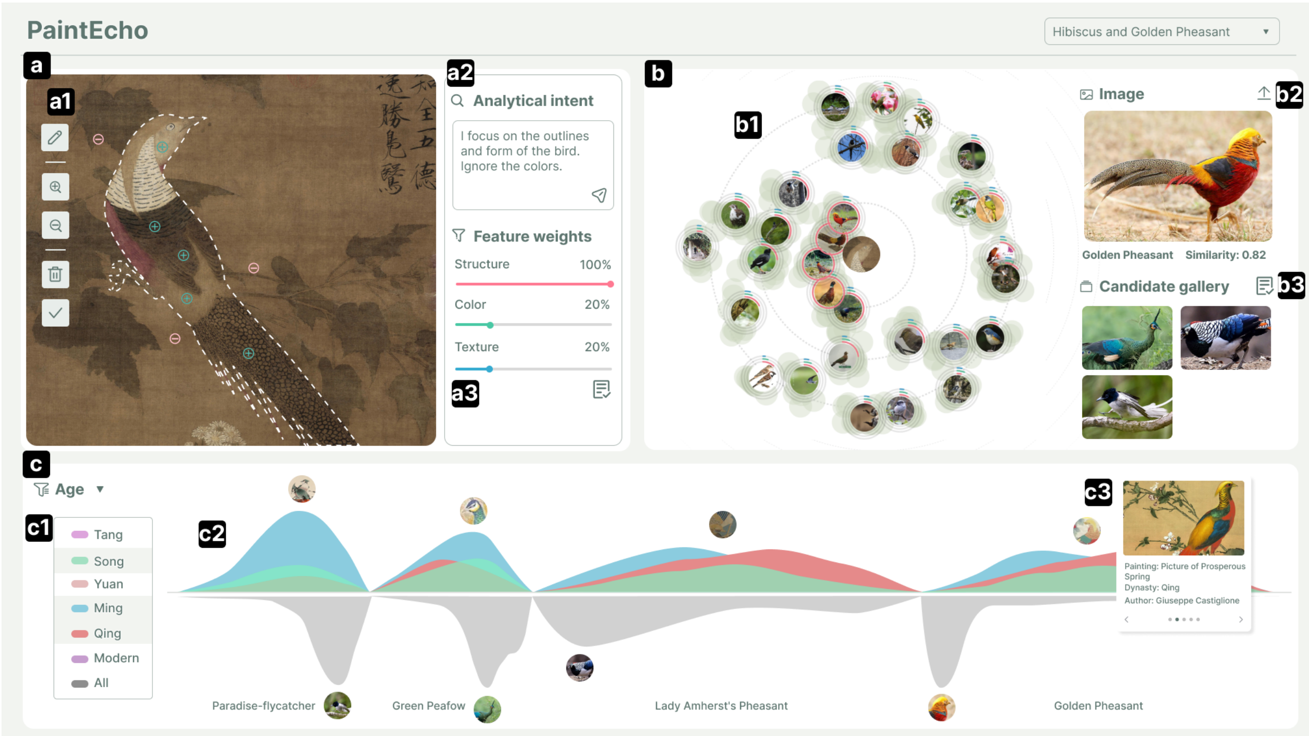

Figure 3: PaintEcho interface: (a) The query view consists of a painting canvas (a1) for target instance extraction and an intent panel (a2, a3) for analytical intent specification. (b) The association view employs a radial plot (b1) for species identification, with a detail view (b2) and a candidate gallery (b3) for inspection and management. (c) The collection view utilizes the ridgeline plot (c2) for distribution alignment via a "mountain-reflection" metaphor, supported by a control panel (c1) and detailed tooltip (c3).

affinities into coherent angular sectors, thereby spatially clustering similar instances. To improve scalability while preserving species identity, each species is represented by its manifold centroid as a primary node, with other instances rendered as subtle shadows. Since biological categories exhibit semantic divergence, they naturally occupy discrete angular sectors, inherently minimizing overlap among primary nodes. We further apply a force-directed refinement to resolve local overlaps while maintaining angular consistency.

4.3.2 Mirrored Distribution Alignment

To facilitate a global comparison, we implement a cross-domain alignment that projects the distributions of paintings and photographs onto a shared semantic axis. This approach enables users to inspect categorical overlaps and visual shifts between the two domains. The alignment is established by projecting high-dimensional features onto the first principal component (PC_1), which serves as a unified reference axis for both domains. To model the distribution along this axis, we compute the density $\hat{f}_h(x)$ using Kernel Density Estimation (KDE), defined as $\hat{f}_h(x) = \frac{1}{nh}\sum K(\frac{x-x_i}{h})$, where h is the bandwidth. The densities are rendered in a mirrored configuration, with the painting and photographic distributions plotted in opposite directions. This symmetrical mapping highlights distributional gaps, effectively characterizing the semantic divergence between artistic representations and photographic reality. Furthermore, distributions can be grouped by metadata, revealing how cross-domain alignment shifts across different contexts.

5 Visualization Designs

PaintEcho comprises three coordinated views aligned with the three-stage workflow (Sec. 3.1) and six requirements (Sec. 3.2) for systematic discovery and interpretive analysis of cross-domain associations. Specifically, the query view (Stage 1) enables users to extract target instances from degraded paintings (**R1**) and specify their analytical intents (**R2**). The association view (Stage 2) organizes photographs in a radial layout for species identification (**R4**) and provides interpretable evidence for potential matches (**R5**). The collection view (Stage 3) employs a "mountain-reflection" metaphor to align cross-domain distributions in a unified overview (**R6**). This design ensures a seamless transition from individual instances to global patterns, thereby supporting hierarchical exploration (**R3**).

5.1 Query View

Serving as a structured starting point, the query view (Fig. 3-a) comprises a painting canvas for target instance extraction (R1) and an intent panel for analytical intent specification (R2).

The painting canvas (Fig. 3-a1) displays the full painting and supports SAM-based [7] instance extraction. Users first locate target instances using textual prompts via the pen tool, then refine boundaries in degraded areas with positive and negative points, visually encoded as colored dots. Once confirmed, the extracted instance is automatically centered in the radial plot(Fig. 3-b1), anchoring subsequent cross-domain comparisons. The intent panel combines a text prompt and interactive sliders. Within the analytical intent field (Fig. 3-a2), users can describe their research focus in plain text, which the system parses to suggest initial parameter configurations. Users can then refine the weights for texture, structure, and color via the sliders (Fig. 3-a3), grounded in the "Six Laws of Xie He" [6]. In addition, users can adjust the feature weights to the painting condition. When degradation makes the original color information unreliable, they can reduce the color weight. Weight confirmation then triggers a cross-domain search to retrieve relevant photographs. Together, the text prompt and sliders ensure semantically steered and numerically precise retrieval.

5.2 Association View

Driven by the query inputs, the association view (Fig. 3-b) supports painting–photo comparison through a radial plot (Fig. 3-b1) for candidate organization and an information panel for inspection and selection.

Visual encoding. To organize photographic candidates by species and similarity (**R4**), the radial plot (Fig. 4-c) adapts SolarView [8], which places a focal entity at the center and embeds its neighbors along concentric circles for organizing query-centered similarity relationships. We employ a polar coordinate system centered on the painting instance, where radial distance encodes the intent-weighted similarity and angular position preserves high-dimensional semantic proximity. Thus, more relevant candidates appear closer to the center, while visually related species form coherent local neighborhoods. SolarView is designed to support exploration of relatively small local neighborhoods, whereas our setting involves numerous retrieved photographs that users compare at the species level rather than inspect individually. We therefore use representative thumbnails to reduce clutter. For each species, we select the candidate nearest the centroid as the representative thumbnail, while others remain as simple circles. This approach highlights prototypical candidates while preserving the overall data distribution. Furthermore, to make cross-domain similarity more interpretable (R5), each thumbnail is surrounded by three concentric arcs. The arc colors represent structure (red), color (green), and texture (blue), consistent with the weight sliders (Fig. 3-a3). Starting from the top center, each arc's length encodes its contribution to the overall similarity score, helping users interpret and validate the retrieval results.

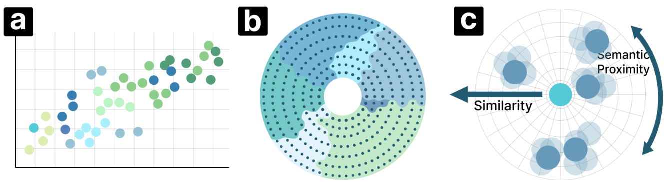

Figure 4: Alternative visual designs for the association view: (a) a scatterplot, (b) a circular grid, and (c) our final radial layout.

Alternative design. During iterative design, two alternative strategies for organizing photographic candidates were considered. A 2D scatterplot (Fig. 4-a) effectively reveals global semantic clusters, but it is less suitable for query-centered comparison. The painting query is not visually distinguished, and spatial distance in a 2D layout is often a poor proxy for similarity. We also evaluated a circular grid layout (Fig. 4-b) inspired by VADIS's document map [30]. While ensuring a tidy, overlap-free arrangement, it has two major limitations for image association. First, color coding does not scale well as the number of species grows, because limited palettes reduce category discriminability. Second, representing candidates as identical dots precludes at-a-glance recognition and requires point-by-point inspection.

Interaction. The radial plot and information panel are coordinated to support the exploration of candidates. Selecting any candidate in the radial plot (Fig. 3-b1) highlights all circles of the same species and updates the photo detail view (Fig. 3-b2) with its high-resolution version. Users can then add validated photographs to the candidate gallery (Fig. 3-b3) and confirm the curated set to proceed to collection-level exploration of paintings associated with the identified species.

5.3 Collection View

To support the discovery of broader artistic patterns, the collection view (Fig. 3-c) provides an overview of the entire dataset. While the association view focuses on instance-level analysis, this view facilitates the synthesis of global patterns around identified species, supporting hierarchical exploration across analytical scales (R3).

Visual encoding. To align cross-domain distributions (R6), we introduce a "mountain-reflection" metaphor that translates statistical density into a landscape-inspired symmetric layout. Paintings are visualized as upper mountains, while photographs appear as their mirrored reflections in the water. This metaphor bridges historical art with modern photography, establishing a seamless intersection where ancient paintings find a "visual echo" in the contemporary world. Specifically, for each identified species, we employ ridgeline plots (Fig. 3-c2) to visualize its prevalence. The horizontal axis encodes a semantic space dynamically shaped by the user's exploration intent in the query view, while the vertical amplitude encodes the instance frequency. Within this context, higher peaks intuitively signify prominent clusters, marking where species were most frequently documented in both art and nature. In addition, the ridges are color-coded by user-selected attributes (e.g., age or author) and overlaid to reveal the cumulative composition and evolution over the selected dimension. The vertical alignment facilitates direct cross-domain comparison, enabling users to discern whether artistic interpretations align with or deviate from biological reality.

Alternative design. We evaluated two primary layouts for the ridgeline plots: stacked and overlaid. We initially considered a stacked design (Fig. 5-a) where categories are layered vertically. While this layout creates a clean "mountain" silhouette and avoids occlusion, it introduces a baseline shift for each category. This displacement makes it challenging to precisely compare the absolute density peaks and semantic shifts across different eras or regions. Ultimately, we adopted the overlaid design (Fig. 5-b) to prioritize analytical precision. By anchoring all ridges to a shared baseline, this layout enables users to directly perceive relative shifts in distribution peaks and widths. To mitigate potential visual occlusion, we employ semi-transparent gradient fills that preserve the legibility of overlapping patterns. This design was further validated by a user study (Sec. 6.2), where the overlaid layout outperformed the stacked alternative.

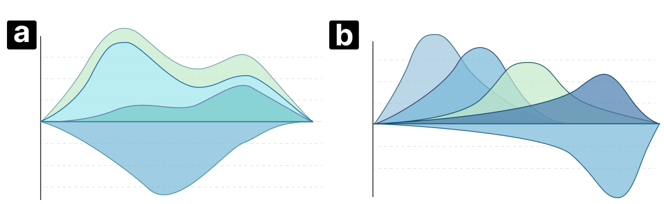

Figure 5: Alternative designs for the collection view. Comparison between the baseline stacked ridgeline (a) and our final overlaid design (b).

Interaction. The control panel (Fig. 3-c1) supports attribute switching and legend-item toggling to refine comparisons. To bridge the gap between statistical trends and individual artworks, each prominent peak is annotated with a representative thumbnail. Hovering over a circular painting thumbnail triggers a detailed tooltip (Fig. 3-c3) containing an enlarged painting instance and metadata. Users can browse additional related paintings and photographs to qualitatively verify global trends and investigate the underlying factors driving these patterns.

5.4 Implementation

PaintEcho is implemented as a web-based visual analytics system with data storage, a Python backend, and a frontend developed with Vue.js [50] and D3.js [5]. Data storage manages curated collections of paintings and photographs as structured input for the backend. The backend supports interactive query formulation, a cross-domain visual association model, and multi-scale layout generation, while the frontend supports interactive exploration through three coordinated views.

6 Evaluation

We evaluate PaintEcho comprehensively: we first assess our cross-domain visual association model, demonstrating its capacity to bridge the domain gap and support diverse analytical needs; we next conduct a user study to evaluate the system's usability and effectiveness, focusing on species identification and a baseline comparison for global pattern discovery; finally, we present two case studies with domain experts to demonstrate the practical utility of PaintEcho in exploring cross-domain associations between paintings and photographs. IRB review was not required for this study according to our institutional policies. All participants provided informed consent prior to participation.

6.1 Model Experiments

To evaluate our cross-domain visual association model, we conducted a quantitative experiment on both a public benchmark and our custom dataset. Specifically, we utilized the Office-Home dataset [42] (Art → Real task) as a general benchmark for cross-domain association. To address the insufficient taxonomic granularity of the general benchmark, we developed a custom P2P dataset, comprising expert-verified, fine-grained associations between paintings and real-world photographs. This dataset was developed using the image collections detailed in

Sec. 3.3. To ensure a high-fidelity ground truth, we randomly sampled painting instances (n=30) and established candidate biological correspondences based on morphological evidence. These associations were then independently reviewed and approved by experts (E1–E4). To ensure the consistency of our findings, extended evaluations with more sample settings are provided in the supplementary material.

We compared our method with two representative baselines: DINOv2-L [28], a foundation model that serves as our feature extraction backbone, and DUDE [47], the current state-of-the-art (SOTA) in unsupervised cross-domain retrieval. Due to the lack of a publicly available implementation, DUDE was omitted from the custom P2P dataset evaluation. As the baselines are limited to static retrieval, we configured our model with a default weight setting ($w_s = 1.0, w_c = 0.3$, and $w_t = 0.2$) to ensure a fair comparison. The baselines were evaluated using their default configurations. Performance was measured using Precision@k ($P@k$ for $k = 1,5,15$) [47], representing the accuracy within the top-k results.

Table 1: Comparison of retrieval performance on Office-Home and custom P2P datasets. The best results are highlighted in **bold**.

Dataset	Method	P@1	P@5	P@15
Office-Home (Art→Real)	DUDE	59.41	57.59	55.45
	DINOv2-L	76.51	75.73	73.66
	Ours (CDVAM)	**78.62**	**78.19**	**76.33**
Custom P2P	DUDE	-	-	-
	DINOv2-L	**54.17**	48.33	48.33
	Ours (CDVAM)	**54.17**	**50.83**	**52.50**

As shown in Table 1, our method matches or outperforms the baselines across both datasets. On the Office-Home benchmark, PaintEcho achieves a $P@15$ of 76.33%, outperforming DINOv2-L by 2.67% and DUDE by 20.88%. On our custom P2P dataset, CDVAM shows larger gains at broader retrieval scales (e.g., $k = 5,15$). The improved top-k precision helps surface the target species despite significant artistic distortion, supporting subsequent cross-domain analysis. While the improvement over DINOv2-L baseline is moderate, CDVAM is not introduced solely for higher $P@k$. Beyond retrieval performance, it supports the visual analytics workflow by enabling intent-driven feature steering and providing interpretable similarity evidence for cross-domain association. Overall, these quantitative results demonstrate that our model provides a competitive retrieval basis for cross-domain association analysis.

6.2 User Study

We conducted a task-based user study to evaluate PaintEcho's cross-domain association capabilities, employing both quantitative and qualitative methods to assess effectiveness and usability.

6.2.1 Participants and Setup

We recruited 12 art-history researchers (1 male, 11 females; all with at least a bachelor's degree in relevant fields), including ten junior and two senior researchers with over eight years of experience. None had prior experience with PaintEcho. A pre-study questionnaire showed that all participants were familiar with basic visualizations; six were well-versed in traditional painting theory, and six had prior exposure to computer vision. Notably, 11 participants reported frequent needs for biological references in their research, underscoring the practical utility of PaintEcho. Each participant received $15 for the one-hour study.

Because no existing system, to our knowledge, provides a comparable workflow for this complex and multifaceted cross-domain association task, direct comparison with external tools was infeasible. Consequently, we aligned our evaluation with the analytical workflow and design requirements. The first stage, instance extraction and intent specification, provides the functional foundation for the analysis workflow. For species identification, we evaluated the association view by measuring user accuracy in identifying real-world species from large candidate sets. Given their task-specific nature, baseline comparisons for these two stages were deemed unnecessary. In contrast, for global artistic insight discovery, we developed a baseline version of PaintEcho to isolate the effect of the collection view design. It replaces the proposed overlaid layout with a stacked arrangement (see Fig. 5), while others remain identical for a fair comparison.

The study comprised two tasks: a targeted fundamental task (Task 1) and an open-ended exploration task (Task 2). For Task 1, we employed a within-subjects design to compare PaintEcho and the baseline system using two bird-themed paintings (P1, P2). To minimize order and learning effects, the 12 participants were randomly assigned to four groups that counterbalanced both system order and painting order across the two trials. In contrast, Task 2 was an exploratory session where all participants used PaintEcho with floral painting (P3). Regarding the experimental materials, P1 and P2 were supported by a photograph library containing 68 bird species and 1,495 painting instances, while P3 was paired with a database of 33 flower species and 3,559 instances.

6.2.2 Tasks and Procedure

The one-hour study was organized into three phases: orientation and practice (15 min), formal evaluation (30 min), and a concluding debriefing (15 min). We first introduced the research objectives and tasks, demonstrated the system, and allowed participants to practice with a sample painting and ask questions until they felt confident using the system. The formal evaluation consisted of two tasks. In Task 1, participants followed the three-step workflow in Sec. 3.1 under one of four group conditions: (1) segment the target instance from the painting and specify a shape-centric analytical intent, proceeding when satisfied; (2) identify related species in the photographic library and select the single best-matching species using the system's similarity evidence; and (3) answer four predefined questions about cross-dynasty patterns of morphological realism deviation. We recorded response accuracy against the established ground truth, with completion time specifically measured for the final synthesis step. In Task 2, participants engaged in a 10-minute open-ended think-aloud exploration, freely investigating cross-domain associations related to their research interests. Finally, they filled out the System Usability Scale (SUS) and additional subjective questionnaires, followed by a semi-structured interview.

6.2.3 Results and Analysis

In this section, we report the evaluation results from three perspectives: subjective questionnaires, objective metrics, and qualitative feedback. For all subjective and objective comparisons between PaintEcho and the baseline, we used the Wilcoxon Signed-Rank Test (N=12).

Subjective Questionnaires. We used SUS to evaluate the system's usability and custom 7-point Likert scales to assess its effectiveness against specific functional requirements. As shown in Fig. 6-b, PaintEcho achieved an average SUS score of 80.42 ($SD = 10.60$), indicating "Excellent" usability and confirming its suitability for professional art-historical analysis. The custom questionnaire comprised seven 7-point Likert-scale items ($1 =$ strongly disagree, $7 =$ strongly agree), specifically mapping Q1-Q6 to our design requirements (R1-R6) and Q7 to the fostering of novel artistic insights. While we recorded comparative ratings for both systems on Q6 and Q7, Q1–Q5 were assessed exclusively for PaintEcho, as these items evaluate functional components shared by both systems. As shown in Fig. 6-a, participants reported positive feedback across all functional requirements of PaintEcho. The high scores for Q1 ($M = 6.33, SD = 0.65$) and Q2 ($M = 6.17, SD = 0.94$) demonstrate the efficacy of interactive extraction and intent modeling. Users noted that the query view enabled the isolation of target instances from degraded backgrounds and aligned system logic with their research focuses. The hierarchical exploration logic (Q3: $M = 5.83, SD = 1.03$) was well-received, as it mirrors participants' natural process of moving from individual instances to broader motifs. Furthermore, Q4 ($M = 6.08, SD = 0.79$) confirms the radial plot's utility for species identification. Q5 ($M = 5.42, SD = 1.08$) indicates that the similarity evidence provided interpretable justifications for cross-domain validation. In comparative assessments, PaintEcho significantly outperformed the baseline in revealing global cross-domain distributions, with median

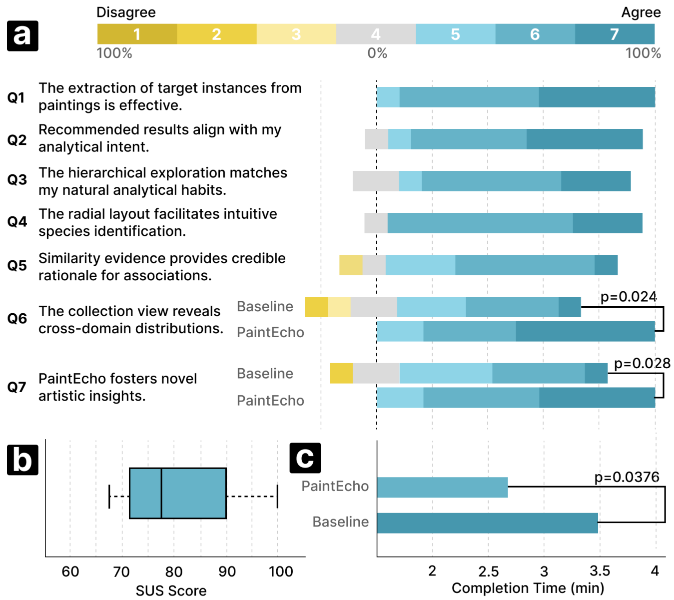

Figure 6: User study results: (a) Participant responses to the custom questionnaire. (b) SUS scores for system usability. (c) Completion time for global pattern discovery between PaintEcho and the baseline.

scores of 6.5 versus 5.0 (Q6: $W = 5.00, Z = -2.26, p < 0.05, r = 0.65$). Finally, as a measure of holistic impact, PaintEcho was rated significantly higher for fostering novel artistic insights (Q7: $Md_{\text{PaintEcho}} = 6.0$ vs. $Md_{\text{baseline}} = 5.0; W = 2.00, Z = -2.19, p < 0.05, r = 0.63$).

Objective Metrics. We further quantified participant performance, focusing on identification accuracy and the efficiency of global pattern discovery. For species identification, participants achieved a high average accuracy of 83.33%, with all errors localized within the same biological family as the ground truth. For global pattern discovery, no significant difference in accuracy was found between PaintEcho and the baseline ($p > 0.05$), as both reached near-perfect performance (approaching 100%). This ceiling effect indicates that both systems provided sufficient information to derive the correct patterns. However, as shown in Fig. 6-c, a significant difference emerged in task completion time ($Md_{\text{PaintEcho}} = 2.68, Md_{\text{baseline}} = 3.49; W = 12.00, Z = -2.08, p < 0.05, r = 0.60$). This gain is attributed to the overlaid layout, which facilitates direct visual juxtaposition and reduces the cognitive overhead of context switching required by the stacked baseline.

Qualitative Feedback. Post-experiment interviews and think-aloud sessions provided qualitative insights into PaintEcho's support for cross-domain art-historical research and suggestions for refinement.

The radial plot facilitates intent-driven species identification. The radial plot simplifies species identification by grouping photographic candidates into species and mapping similarity to radial distance. We observed that participants typically focused their analysis on the innermost rings; as P1 noted, "*I prioritized the center circle. If I found a satisfactory match there, I wouldn't bother checking the less relevant species in the outer rings.*" Additionally, the rings encoded by color foster trust in the model by providing similarity evidence. P12 stated that "*These visual cues provided immediate feedback on whether the results truly align with my analytical intent, which enhanced my confidence in the system's recommendations.*" Furthermore, the angular layout groups similar species together, helping users locate related species through semantic proximity. P3 described this process as "*finding visual anchors*", noting that "*finding a close match allowed them to explore the surrounding neighborhood to pinpoint the exact species.*"

The mountain-reflection design fosters artistic insights by visually bridging paintings and photographs. The collection view unifies cross-domain datasets in a single visual space to reveal global patterns. The vertical alignment of "mountains" and "reflections" facilitates direct cross-domain comparison, making the differences between paintings and photographs intuitive. P11 noted that "*The shift between corresponding peaks allowed me to judge the degree of artistic stylization.*" P9 added that "*This layout reveals where the artist departed from reality, highlighting the specific modifications to a species' shape or color.*" In addition, the layout reveals the artistic evolution across dynasties. P4 traced a clear evolution from the varied species depicted in the Song dynasty to the highly symbolic "Four Gentlemen" motifs of later periods. As P7 remarked, "*The collection view visually captures the historical signatures.*" Moreover, the "mountain-reflection" metaphor fosters cognitive resonance, making cross-domain associations more intuitive. Participants viewed the metaphor to be a conceptual bridge rather than a mere tool. P10 said, "*Ancient paintings and modern photographs mirrored each other across the river of time, transforming technical similarity into a narrative of cultural continuity.*" To better support the inspection of subtle stylistic nuances, P3 suggested adding a local magnification feature to the ridgeline plots.

6.3 Use Cases

We present two use cases to demonstrate the utility of PaintEcho in art-historical research, focusing on avian morphology (Case I) and floral textures (Case II).

6.3.1 Case I: Hibiscus and Golden Pheasant

We first selected Emperor Huizong's (Zhao Ji) masterpiece, *Hibiscus and Golden Pheasant*, focusing on identifying its biological prototype and tracing the evolution of related motifs across artistic periods. Noting that the masterpiece depicts a bird perched among floral branches, we utilized the interactive segmentation tool (Fig. 3-a1) to isolate the bird from the composition. We then specified the analytical intent by issuing the prompt (Fig. 3-a2): "I focus on the outlines and form of the bird. Ignore the colors." Following the feature weights (structure: 1.0, color: 0.2, texture: 0.2; Fig. 3-a3), the association view was updated to display photographic candidates with high similarity. In the radial plot (Fig. 3-b1), Golden Pheasant and Lady Amherst's Pheasant appeared as the closest matches, with dominant red arcs indicating that the retrieval was steered by our structural focus. Through the photo detail view (Fig. 3-b2), we verified the painted bird as a hybrid specimen, noting the combination of a Lady Amherst's Pheasant neck ruff with a Golden Pheasant body. This observation matches the published finding that the depicted bird is a hybrid [29], showing that PaintEcho can help surface evidence that previously required extensive manual investigation.

We then added both pheasants and Green Peafowl to the candidate gallery (Fig. 3-b3) for their structural similarities. Following the same angular direction, we spotted Amur Paradise Flycatcher. Noting the visual rhythm shared between their long tails, we added it to the gallery to contrast their respective painting styles. Moving to the collection view (Fig. 3-c), the "mountain-reflection" layout (Fig. 3-c2) showed that the painting distributions of the two pheasant species were nearly identical, despite their divergent photographic reflections. This finding suggested that historical artists often conflated these two species into a single morphological archetype, likely due to their frequent natural hybridization. The two pheasant species also showed pronounced peak offsets from biological reality, challenging our initial impression of the masterpiece's strict naturalism. By inspecting representative paintings (Fig. 3-c3), we traced this discrepancy to a "looking-back" pose, which is rare in modern photography but frequently used by historical artists to maximize the display of the birds' ornate plumage. Furthermore, we noticed that Amur Paradise Flycatcher paintings exhibited sharp, unified peaks that were slightly offset from biological reality. Further inspection revealed that artists systematically elongated the tail feathers to emphasize the bird's symbolic association with longevity.

Finally, we grouped the painting collection by dynasty to track the evolution of morphological realism over time. Given our expectation that Song paintings represented a peak of realism, we were surprised to find that Ming pheasant-related works showed higher structural fidelity than those from the Song and Qing dynasties. Further categorization by author (Fig. 7-a) showed that high morphological realism was characteristic of many Ming works, notably in the precise golden pheasant depictions by Lü Ji. However, when we shifted the analytical intent to color, the ranking inverted to Qing, Ming, and then Song (Fig. 7-b).

This shift may be caused by the profound impact of European missionary painters like Giuseppe Castiglione in the Qing dynasty, alongside the long-term advancement of pigment technologies. Taken together, this case opens up further art-historical questions about species conflation and shifts in visual fidelity across historical periods.

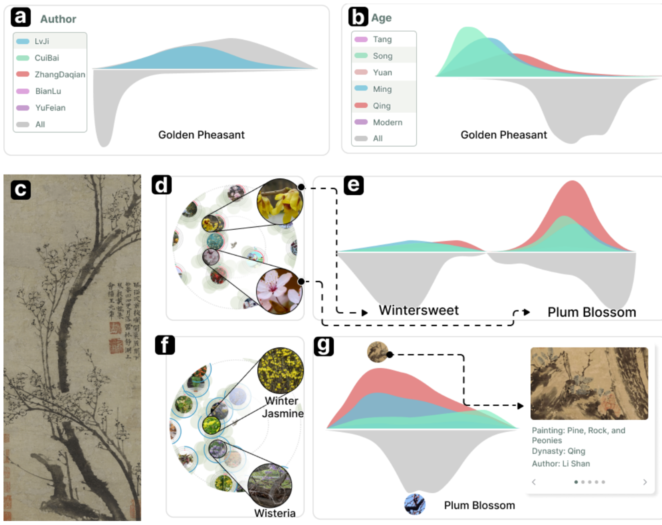

Figure 7: Use case results. (a-b) Case I: Global patterns of morphological and color fidelity across authors and dynasties. (c-g) Exploration workflow for Case II: (c) Wang Mian's Ink Plum Blossoms; (d-e) structure-driven species identification and distribution alignment; (f-g) texture-driven exploration, uncovering stylistic parallels across disparate motifs.

6.3.2 Case II: Ink Plum Blossoms

We selected Wang Mian's masterpiece, *Ink Plum Blossoms* (Fig. 7-c), to investigate how PaintEcho interprets minimalist, monochromatic ink-wash paintings. With a shape-oriented intent, the radial plot (Fig. 7-d) placed Plum Blossom, Wintersweet, Apricot Blossom, and Chinese Flowering Apple in the innermost rings. Comparing these candidates, we confirmed the subject as Plum Blossom. Interestingly, while Plum Blossom and Wintersweet are often conflated in Chinese nomenclature, the system separated them into discrete angular sectors. We could easily notice their biological differences, observing that the painting's rounded petal contours closely align with the morphology of Plum Blossom, distinct from the waxy, pointed petals of Wintersweet. Moving to the collection view (Fig. 7-e), we noted a sharp contrast: while Plum Blossom paintings were far more abundant, they exhibited lower structural fidelity than the rarer, more biologically accurate Wintersweet paintings. We further grouped the collection chronologically and discovered a sudden surge in Qing-dynasty Plum Blossom paintings, characterized by a sharp, uniform peak in the mountain-reflection layout. This quantitative signature may serve as evidence of symbolization. Since its canonization as one of the "Four Gentlemen" in the late Ming Dynasty, Plum Blossom has become a vessel for fortitude and noble integrity. This cultural shift led Qing artists to favor standardized formulas over life study, narrowing the motif from naturalism to symbolism. This observation aligns with prior studies of ink plum painting, which discuss the plum blossom as a symbolic motif in Chinese literati art [4].

We then investigated the textural feature. In the association view (Fig. 7-f), the system retrieved Winter Jasmine and Wisteria as the most similar candidates. Rather than algorithmic errors, closer inspection revealed striking textural parallels: dry-brush strokes captured the weathered vitality of plum trunks, echoing the gnarled twists of wisteria vines, while dense, elongated branches and "dotting" techniques conveyed flourishing luxuriance—mirroring winter jasmine morphology. As shown in Fig. 7-g, Plum Blossom photographs retrieved the painting *Pine, Rock, and Peonies* based on texture similarity. Despite the thematic divergence, the *Cun* (texture strokes) used for the pine bark and rocks shared a consistent tactile essence of weathered ruggedness with Wang Mian's plum trunks. These findings suggest that PaintEcho can help analysts trace the transition from naturalism to symbolism, while also raising further questions about motif standardization and shared brushwork conventions across different painting subjects.

7 Discussion

We discuss the implications, generalizability, design lessons learned, and limitations for future work.

Implications. PaintEcho establishes a model-assisted paradigm for cross-domain exploration between paintings and real-world photographs. It streamlines manual comparison into an AI-assisted workflow, enabling experts to identify relevant species and derive global artistic insights from fragmented matches. For art history research, PaintEcho reveals previously inaccessible image-based visual connections. As shown in the use cases, these connections can be organized as visual evidence for expert interpretation, linking system findings with prior scholarship and supporting follow-up art-historical studies. For visual analytics research, our model bridges the domain gap via disentangled feature representations, enabling interpretable cross-domain visual associations. Our "mountain-reflection" metaphor provides a scalable representation of directed associations, applicable to diverse tasks in digital humanities and scientific visualization.

Generalizability. PaintEcho's current implementation and evaluation focus on biological motifs, particularly birds and flowers. Its generalizability spans two levels. At the object-category level, PaintEcho can be extended to other recurring painting objects, such as human figures, architecture, artifacts, and landscapes, given corresponding photographic datasets and expert-validated category labels. At the painting-type level, the framework may apply to other painting datasets, such as oil paintings and murals. In these settings, analytical dimensions may need domain-specific adaptation to accommodate different art-historical theories, but the core framework remains unchanged.

Lessons learned. We report three transferable design lessons from PaintEcho for visual analytics systems that support cross-domain visual associations between distinct image collections. First, such systems should support multi-level exploration rather than isolated retrieval by connecting instance-level association with collection-level pattern discovery, thereby helping users synthesize fragmented matches into global insights. Second, algorithm designs should align with domain expertise. Deep-learning features integrated with domain theories and disentangled intuitive controls turn the "black box" into an interpretable tool for art research, not just computational outputs. Third, visual metaphors like "mountain-reflection" act as a conceptual bridge, translating abstract similarity scores into a natural visual language. Such metaphors help domain experts synthesize fragmented matches into coherent narratives and lower barriers to complex analytical systems.

Limitations and future work. Despite these benefits, PaintEcho suffers from three limitations: First, PaintEcho currently focuses on visual associations only and does not incorporate cultural or symbolic meanings, making the integration of deep cultural features a promising future work. Second, retrieval may be less effective when painting instances are severely degraded or when the corresponding objects are absent from the photo collection, resulting in incomplete or less relevant candidate sets. Third, the collection view depends on sufficient data coverage; limited data may produce unstable density-based distributions that should be interpreted cautiously. As a proof-of-concept prototype, PaintEcho currently uses offline image datasets; integrating online image retrieval engines would be a valuable improvement.

8 Conclusion

This study introduces PaintEcho, a visual analytics system for discovering cross-domain associations between paintings and photographs. By disentangling visual attributes, the system enables intent-driven exploration ranging from species identification to global pattern analysis. The radial layout and "mountain-reflection" metaphor effectively reduce cognitive load and reveal artistic insights. Validated through a user study and use cases, PaintEcho demonstrates its utility in supporting art-historical research. Future work will integrate historical contexts to explore the deeper cultural symbolism embedded in artistic motifs.

Acknowledgments

The work was supported by NSFC (62532011), Zhejiang Provincial Natural Science Foundation of China under Grant No. LD25F020003, and NSFC (62421003, 62302440).

References

[1] K. U. Barthel, N. Hezel, K. Schall, and K. Jung. Combining semantic and visual image graphs for efficient search and exploration of large dynamic image collections. In *Proceedings of the 2nd International Workshop on Interactive Multimedia Retrieval*, pp. 1–8. ACM, Lisboa, 2022. doi: 10.1145/3552467.3554796 2

[2] W. Benjamin. The work of art in the age of mechanical reproduction. In H. Arendt, ed., *Illuminations*, pp. 217–251. Schocken Books, New York, 1969. 1

[3] D. Bertucci, M. M. Hamid, Y. Anand, A. Ruangrotsakun, D. Tabatabai, M. Perez et al. DendroMap: Visual exploration of large-scale image datasets for machine learning with treemaps. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):320–330, 2023. doi: 10.1109/TVCG.2022.3209425 2

[4] M. Bickford. *Ink Plum: The making of a Chinese scholar-painting genre*. Cambridge University Press, 1996. 9

[5] M. Bostock, V. Ogievetsky, and J. Heer. D3: Data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011. doi: 10.1109/TVCG.2011.185 6

[6] S. Bush and H.-y. Shih. *Early Chinese texts on painting*. Hong Kong University Press, 2012. doi: 10.1515/9789882208865 3, 4, 5

[7] N. Carion, L. Gustafson, Y.-T. Hu, S. Debnath, R. Hu, D. Suris et al. SAM 3: Segment Anything with concepts. arXiv preprint, 2025. doi: 10.48550/arXiv.2511.16719 3, 5

[8] T. Castermans, K. Verbeek, B. Speckmann, M. A. Westenberg, R. Koopman, S. Wang et al. SolarView: Low distortion radial embedding with a focus. *IEEE Transactions on Visualization and Computer Graphics*, 25(10):2969–2982, 2019. doi: 10.1109/TVCG.2018.2865361 5

[9] C. Chen, Z. Wang, J. Wu, X. Wang, L.-Z. Guo, Y.-F. Li et al. Interactive graph construction for graph-based semi-supervised learning. *IEEE Transactions on Visualization and Computer Graphics*, 27(9):3701–3716, 2021. doi: 10.1109/TVCG.2021.3084694 2

[10] J. Chen, M. Ling, R. Li, P. Isenberg, T. Isenberg, M. Sedlmair et al. VIS30K: A collection of figures and tables from IEEE visualization conference publications. *IEEE Transactions on Visualization and Computer Graphics*, 27(9):3826–3833, 2021. doi: 10.1109/TVCG.2021.3054916 2

[11] C.-Y. Chuang, V. Jampani, Y. Li, A. Torralba, and S. Jegelka. Debiasing vision-language models via biased prompts. arXiv preprint, 2023. doi: 10.48550/arXiv.2302.00070 4

[12] O. de Rooij, J. J. van Wijk, and M. Worring. MediaTable: Interactive categorization of multimedia collections. *IEEE Computer Graphics and Applications*, 30(5):42–51, 2010. doi: 10.1109/MCG.2010.66 2

[13] A. Dong, W. Zeng, X. Chen, and Z. Cheng. VIStory: Interactive storyboard for exploring visual information in scientific publications. In *Proceedings of the 12th International Symposium on Visual Information Communication and Interaction*, pp. 12:1–12:8. Association for Computing Machinery, New York, 2019. doi: 10.1145/3356422.3356430 2

[14] Y. Feng, X. Wang, K. K. Wong, S. Wang, Y. Lu, M. Zhu et al. PromptMagician: Interactive prompt engineering for text-to-image creation. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):295–305, 2024. doi: 10.1109/TVCG.2023.3327168 2

[15] Global Biodiversity Information Facility. What is GBIF? Website, 2026. `https://www.gbif.org/what-is-gbif`. 3

[16] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle et al. The Llama 3 herd of models. arXiv preprint, 2024. doi: 10.48550/arXiv.2407.21783 4

[17] F. Grötschla, L. A. Lanzendörfer, M. Calzavara, and R. Wattenhofer. AEye: A visualization tool for image datasets. In *Proceedings of IEEE Visualization and Visual Analytics (VIS)*, pp. 281–285. IEEE, Tampa, 2024. doi: 10.1109/VIS55277.2024.00064 2

[18] P. C. Hansen. The truncatedSVD as a method for regularization. *BIT Numerical Mathematics*, 27(4):534–553, 1987. doi: 10.1007/BF01937276 4

[19] R. E. Harrist. Review of Ink Plum: The making of a Chinese scholar-painting genre. *Artibus Asiae*, 59(1/2):143–146, 1999. doi: 10.2307/3249866 2

[20] C. Hu and G. H. Lee. Feature representation learning for unsupervised cross-domain image retrieval. In *Computer Vision – ECCV 2022*, pp. 529–544. Springer Nature Switzerland, Cham, 2022. doi: 10.1007/978-3-031-19836-6_30 2

[21] C. Hu, C. Zhang, and G. H. Lee. Unsupervised feature representation learning for domain-generalized cross-domain image retrieval. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10982–10991. IEEE, Paris, 2023. doi: 10.1109/ICCV51070.2023.01011 2

[22] X. Ji, W. Wang, M. Zhang, and Y. Yang. Cross-domain image retrieval with attention modeling. In *Proceedings of the 25th ACM international conference on Multimedia*, pp. 1654–1662. Association for Computing Machinery, New York, 2017. doi: 10.1145/3123266.3123429 2

[23] M. Kotsidi, G. Gorgolis, M. G. Pastore Carbone, G. Anagnostopoulos, G. Paterakis, G. Poggi et al. Preventing colour fading in artworks with graphene veils. *Nature Nanotechnology*, 16(9):1004–1010, 2021. doi: 10.1038/s41565-021-00934-z 3

[24] S. Lang and B. Ommer. Transforming information into knowledge: How computational methods reshape art history. *Digital Humanities Quarterly*, 15(3):000560:1–000560:31, 2021. doi: 10.63744/4vkc72ax2hj6 1

[25] B. Li, Y. Shi, Q. Yu, and J. Wang. Unsupervised cross-domain image retrieval via prototypical optimal transport. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(4):3009–3017, 2024. doi: 10.1609/aaai.v38i4.28083 1, 2

[26] L. Manovich. *Cultural analytics*. MIT Press, 2020. doi: 10.7551/mitpress/11214.001.0001 1

[27] Microsoft. Microsoft Bing Search. Website, 2026. `https://www.bing.com`. 3

[28] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov et al. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. `https://openreview.net/forum?id=a68SUt6zFt`. 4, 7

[29] M.-S. Peng, F. Wu, R. W. Murphy, X.-J. Yang, and Y.-P. Zhang. An ancient record of an avian hybrid and the potential uses of art in ecology and conservation. *Ibis*, 158(2):444–445, 2016. doi: 10.1111/ibi.12354 1, 8

[30] R. Qiu, Y. Tu, P.-Y. Yen, and H.-W. Shen. VADIS: A visual analytics pipeline for dynamic document representation and information-seeking. *IEEE Transactions on Visualization and Computer Graphics*, 31(1):1312–1321, 2025. doi: 10.1109/TVCG.2024.3456339 6

[31] U. Rashid, M. Saddal, G. Farooq, M. A. Khan, and N. Ahmad. An SUI-based approach to explore visual search results cluster-graphs. *PLOS ONE*, 18(1):e0280400, 2023. doi: 10.1371/journal.pone.0280400 2

[32] A. Sain, A. K. Bhunia, Y. Yang, T. Xiang, and Y.-Z. Song. StyleMeUp: Towards style-agnostic sketch-based image retrieval. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8500–8509. IEEE, Nashville, 2021. doi: 10.1109/CVPR46437.2021.00840 2

[33] S. Schneider, M. Springstein, J. Stalter, D. Ritter, R. Ewerth, and E. Müller-Budack. Uncovering biases through faceted image retrieval: Wikidata as a data provider for art-historical research. *Journal of Open Humanities Data*, 12(1):8:1–8:14, 2026. doi: 10.5334/johd.436 2

[34] X. Shen, A. A. Efros, and M. Aubry. Discovering visual patterns in art collections with spatially-consistent feature learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9270–9279. IEEE, Long Beach, 2019. doi: 10.1109/CVPR.2019.00950 1

[35] D. Smilkov, N. Thorat, C. Nicholson, E. Reif, F. B. Viégas, and M. Wattenberg. Embedding Projector: Interactive visualization and interpretation of embeddings. arXiv preprint, 2016. doi: 10.48550/arXiv.1611.05469 2

[36] J. Song, Q. Yu, Y.-Z. Song, T. Xiang, and T. M. Hospedales. Deep spatial-semantic attention for fine-grained sketch-based image retrieval. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 5552–5561. IEEE, Venice, 2017. doi: 10.1109/ICCV.2017.592 2

[37] A. Stephan, L. Miklautz, C. Leiber, P. H. Luz De Araujo, D. Répás, C. Plant et al. Text-guided alternative image clustering. In *Proceedings of the 9th Workshop on Representation Learning for NLP (RepL4NLP-2024)*, pp. 177–190. Association for Computational Linguistics, Bangkok, 2024. doi: 10.18653/v1/2024.repl4nlp-1.13 2

[38] Z. Tang, H. Fan, J. Zhang, H. Ma, X. Gu, B. Li et al. ShieldIR: Privacy-preserving unsupervised cross-domain image retrieval via dual protection transformation. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 6383–6392. Association for Computing Machinery, New York, 2025. doi: 10.1145/3746027.3754804 1, 2

[39] N. Ufer, M. Simon, S. Lang, and B. Ommer. Large-scale interactive

retrieval in art collections using multi-style feature aggregation. *PLOS ONE*, 16(11):e0259718, 2021. doi: 10.1371/journal.pone.0259718 1

[40] P. van der Corput and J. J. van Wijk. ICLIC: Interactive categorization of large image collections. In *2016 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 152–159. IEEE Computer Society, Taipei, 2016. doi: 10.1109/PACIFICVIS.2016.7465263 2

[41] M. J. P. Van Zuijlen, H. Lin, K. Bala, S. C. Pont, and M. W. A. Wijntjes. Materials in Paintings (MIP): An interdisciplinary dataset for perception, art history, and computer vision. *PLOS ONE*, 16(8):e0255109, 2021. doi: 10.1371/journal.pone.0255109 1

[42] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5385–5394. IEEE, Honolulu, 2017. doi: 10.1109/CVPR.2017.572 6

[43] X. Wang, D. Peng, M. Yan, and P. Hu. Correspondence-free domain alignment for unsupervised cross-domain image retrieval. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(8):10200–10208, 2023. doi: 10.1609/aaai.v37i8.26215 1, 2

[44] J. Xia, L. Huang, W. Lin, X. Zhao, J. Wu, Y. Chen et al. Interactive visual cluster analysis by contrastive dimensionality reduction. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):734–744, 2023. doi: 10.1109/TVCG.2022.3209423 2

[45] X. Xie, X. Cai, J. Zhou, N. Cao, and Y. Wu. A semantic-based method for visualizing large image collections. *IEEE Transactions on Visualization and Computer Graphics*, 25(7):2362–2377, 2019. doi: 10.1109/TVCG.2018.2835485 2

[46] X. Xuan, X. Wang, W. He, J. P. Ono, L. Gou, K.-L. Ma et al. VISTA: A visual analytics framework to enhance foundation model-generated data labels. *IEEE Transactions on Visualization and Computer Graphics*, 31(10):6991–7003, 2025. doi: 10.1109/TVCG.2025.3535896 2

[47] R. Yang, P. Hu, Y. Li, and X. Peng. DUDE: Diffusion-based unsupervised cross-domain image retrieval. arXiv preprint, 2025. doi: 10.48550/arXiv.2509.04193 7

[48] Y. Ye, R. Huang, and W. Zeng. VISAtlas: An image-based exploration and query system for large visualization collections via neural image embedding. *IEEE Transactions on Visualization and Computer Graphics*, 30(7):3224–3240, 2024. doi: 10.1109/TVCG.2022.3229023 2

[49] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 401–408. Association for Computing Machinery, New York, 2003. doi: 10.1145/642611.642681 2

[50] E. You. Vue.js: The progressive JavaScript framework. Website, 2026. `https://vuejs.org/`. 6

[51] B. Zhao and O. Mac Aodha. Interpretable text-guided image clustering via iterative search. In *Proceedings of the 36th British Machine Vision Conference*. BMVA, Sheffield, 2025. `https://bmvc2025.bmva.org/proceedings/768/`. 2

[52] Zhonghua Zhenbaoguan. China National Treasure Museum. Website, 2026. `http://www.ltfc.net`. 3