Classification of Abstract Painting

Raphaël Legrand, Enguerrand Le Niliot, Eliott Dionnet, Arthur Evain

June 2026

Supervisors: Henri Maître, Loïc le Folgoc

Abstract

This work proposes an unsupervised classification of a corpus of abstract paintings based exclusively on objective visual properties, without resorting to any data regarding the artists or their schools. We mobilize low-level descriptors (color, texture, composition) combined with the latent spaces of a Convolutional Neural Network (CNN) and a Variational Autoencoder (VAE), fused via weighted *alpha-blending*. These multimodal representations were reduced by Principal Component Analysis (PCA) and partitioned using K-Means. The classification is refined by a quality filter based on the Silhouette score. The project's goal is to evaluate to what extent the emerging clusters correspond to the categories produced by art history and art criticism.

1 Statement on the Use of Artificial Intelligence

In accordance with the project guidelines, this section details exhaustively the use of generative artificial intelligences throughout our study and during the drafting of this report. The models used for this project are Claude and Gemini.

Generally speaking, artificial intelligence assisted in code development, exclusively for visualization with Matplotlib. AI also helped with the formatting of this report in LaTeX.

For the sake of clarity, we detail our specific queries below, separating methodological assistance from pure code generation.

1. Methodological questions and avenues for reflection

- **CNN Extraction:** Request for information and theoretical explanations on the extraction process of feature maps from the first layers of a convolutional neural network.
- **Siamese Space:** Request for theoretical information on the use of a Siamese space for a VAE, and its practical application in our project.
- **Fusion Strategy:** Search for avenues and methodological advice on the most relevant way to fuse our different descriptor matrices (CNN, geometric, etc.).
- **2D Visualization:** Guidance on which method to use for 2D projection in order to integrate it into the software (t-SNE/UMAP/PaCMAP).
- **State-of-the-Art Assistance:** Search for articles and help understanding them on topics where we are not experts.

2. Assistance with pure code and visualization

- **Help with documentation and feature implementation:** AI allowed us to quickly access documentation to translate abstract feature ideas into concrete machine operators, and to find the name of the library and Python function that implements it.
- **Generating Summaries (Fusion):** Help displaying the stylistic distribution of our clusters, as well as plotting the Inter-Artist distance matrix.
- **Topological Inspection (Fusion):** Help creating a script to visually display 5 images located at the center of a cluster, 5 at its border, and 5 drawn randomly.
- **Visualization Software:** Debugging assistance as well as help putting the software online. AI was also useful for generating usable CSVs at the end of the codes for the software.

2 Introduction

Abstract art constitutes one of the most singular and resistant corpora that an automatic classification system can encounter. Born at the turn of the 20th century from the deliberate rejection of the codes of figurative representation, it is precisely defined by the absence of visual references (objects, characters, perspectives) upon which computer vision algorithms usually rely. Yet, for a century, historians, critics, philosophers, psychologists, sociologists, and machine learning engineers have all attempted to order this corpus. This project is part of this interdisciplinary tradition, with an additional and deliberate constraint: to ignore all metadata about the painters, their schools, or their biographies, in order to propose an **unsupervised classification** based exclusively on the objective visual properties of the artworks themselves.

2.1 Philosophical issues

Any classification of abstract art immediately encounters a philosophical objection. Weitz [Weitz, 1956] argues that art is an *open concept* whose boundaries no necessary and sufficient condition can delimit; Danto radicalizes this point by showing that two visually identical paintings can belong to entirely distinct aesthetic categories depending on their context [Danto, 1964]. Our project fully assumes this limitation: what we are classifying are *visual signals*, not artworks in Danto's sense. Goodman nevertheless offers a favorable foundation [Goodman, 1968]: the abstract artwork *exemplifies* properties that it highlights (the heaviness of a Rothko, the regularity of a Mondrian)—properties accessible to quantitative descriptors which justify our approach.

2.2 Psychological and neuroaesthetic issues

The psychology of perception directly legitimizes our technical choices. Berlyne's collative variables [Berlyne, 1971] (complexity, novelty, ambiguity) translate into quantitative measures like texture entropy or chromatic variance, which will be computed in the geometric approach. The model by Leder et al. [Leder et al., 2004] shows that, when faced with abstraction, the absence of object recognition forces the cognitive system towards high-level stylistic processing, motivating the use of deep representations. We will attempt to model these using CNNs and VAEs. Finally, Zeki provides a neurobiological basis for the relevance of differentiated descriptors, as color, form, and motion are processed by distinct cortical areas [Zeki, 1999].

2.3 Human perception and categorization of abstract artworks

Faced with an abstract artwork, the observer performs a rapid visual categorization based on prototype detection [Rosch, 1978]: geometric regularity, chromatic palette, and gestural dynamism are extracted almost immediately, prior to any conscious identification of the movement or style. This extraction relies on Gestalt principles of global synthesis (tension, balance, dynamism) that the eye processes preferentially over local details [Arnheim, 1974], which justifies the inclusion in our pipeline of compositional descriptors capturing the overall structure. These perceptions are then encoded in memory in the form of implicit *stylistic schemas* [Martindale, 1990]: an observer intuitively groups two Rothkos not because they verbalize the concept of *Color Field*, but because the similarity of their chromatic structure and luminous texture activates the same memory representations. It is precisely this type of similarity (intersubjective, pre-verbal, based solely on the visual signal) that our multimodal fusion (CNN, VAE, classical descriptors) seeks to capture and formalize.

2.4 Motivation by example: discriminative visual features

Before describing our pipeline, we illustrate with a few contrasted examples why the objective visual properties we measure constitute relevant descriptors to organize this corpus.

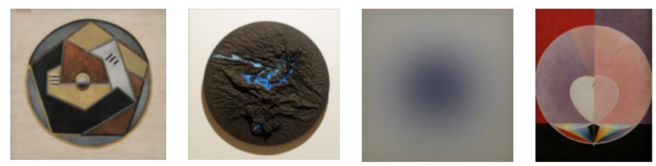


Figure 1: Example of artworks by different painters all featuring a centered circle: intuitively, we would like to classify them all together.

These contrasts illustrate why a multimodal representation is necessary: no single descriptor alone is sufficient to simultaneously discriminate all these dimensions. The JPEG file size distinguishes Soulages from Pollock but confuses Mondrian and Rothko; Hough lines separate Mondrian from Kandinsky but ignore Hartung's gesturality; the autocorrelation score identifies Vasarely and Hantaï but does not differentiate their respective palettes. It is precisely the combination of these complementary descriptors (classical, CNN, VAE) that gives our pipeline its discriminative capacity.

2.5 Computational state of the art

2.5.1 *Hand-crafted descriptors*

Early approaches to computational art classification relied on manually designed descriptors. The exploited descriptors fall into three families [Rathi et al., 2025]: *color* (RGB/HSV histograms, dominant palettes via k-means), *texture* (Gabor filters, GLCM, LBP), and *composition* (Canny/Sobel edges, Fourier transform, symmetry) [Saleh and Elgammal, 2015]. These approaches succeed on the most highly contrasted styles but fail on finely defined categories, whose distinctions rely on contextual criteria inaccessible to the raw visual signal.

2.5.2 Convolutional Neural Networks

The introduction of deep CNNs has significantly advanced the automatic extraction of visual representations for painting classification. Lecoutre et al. [2017] show that a ResNet50 pre-trained on ImageNet and partially re-trained (about 20% of the top layers) on the WikiArt corpus produces features that are much more discriminative for pictorial style than generic ImageNet features: the higher layers, too specialized in naturalistic object recognition, must be adapted to the artistic domain to capture relevant stylistic variations. This result directly motivates our choice to use the intermediate activations of a pre-trained CNN as a source of features: they encode visual structures of increasing complexity, from local textures in the early layers to overall compositional relationships in the deep layers.

Lelièvre and Neri [2021] provide complementary insight specific to abstract art: in a painting orientation task, the deep layers of the network are heavily solicited for abstract artworks much more so than for figurative ones, indicating that global composition (and not local cues) constitutes the dominant stylistic signal in this corpus. This observation justifies our choice not to restrict ourselves to the activations of the final layer, but to extract them at multiple depths of the network.

2.5.3 Variational Autoencoders as feature extractors

Alongside discriminative approaches, variational autoencoders (VAEs) have emerged as extractors of latent representations for art. Unlike CNNs trained on labels, a VAE learns a continuous probabilistic distribution in a compact latent space, capturing global stylistic variations without supervision. Castellano and Vessio [2020] show that such latent spaces spontaneously encode low-granularity stylistic attributes, with the style/content boundary emerging from the model rather than from an imposed bias. Our pipeline directly exploits these latent vectors, fused with CNN features and classical descriptors, to obtain a multimodal representation richer than any single isolated source.

2.5.4 Unsupervised classification

The unsupervised classification of paintings remains significantly less explored than supervised classification. The absence of an indisputable *ground truth* requires combining internal consistency metrics (Silhouette score, Davies-Bouldin index) and human validation, two sources of truth of different and sometimes contradictory natures.

The most performant approaches combine dimensionality reduction and clustering algorithms: projecting into a low-dimensional space prior to clustering significantly improves partition quality, eliminating noise inherent to high-dimensional spaces and revealing the global geometric structure of the corpus [Allaoui et al., 2023]. In our pipeline, this reduction is performed via PCA separately for each feature family, before weighted fusion and K-means clustering.

A recent result [Castellano and Vessio, 2020] shows that latent spaces learned without supervision spontaneously encode recognizable stylistic attributes, confirming that the visual structure of art contains sufficient information to allow non-trivial organization without annotation.

2.5.5 Limitations specific to abstract art and implications for our approach

The survey by Rathi et al. [2025] identifies four challenges specific to this corpus, which constitute concrete constraints on our pipeline.

High intra-class variance. Two artworks by the same artist can be visually more dissimilar than two artworks of distinct styles. For us, this means that intra-painter consistency cannot serve as the primary validation criterion: a visually relevant clustering could separate two periods of the same artist, and this would be a legitimate result. We nevertheless use intra-painter consistency as a secondary indicator, keeping this limitation in mind.

Ambiguity of historical labels. Art history categories have fuzzy boundaries, recognized as such by the experts themselves. This implies that a clear partition produced by our algorithm cannot be deemed incorrect solely because it does not coincide with historiographical divisions: our classification aims for visually homogeneous groups, not necessarily equivalents to historical movements.

The problem of intention. Models classify based on exhibited visual properties, ignoring the context which largely determines expert human classification. We fully assume this limitation: our goal is not to replicate the judgment of the art historian, but to produce coherent groups in terms of the visual signal. This distinction is at the heart of our approach and its evaluation.

Dataset imbalance. Abstract art is underrepresented in large reference datasets, meaning our CNN features, pre-trained on ImageNet or WikiArt, were optimized on distributions very different from our corpus. This is one of the reasons we complement CNN features with a VAE trained directly on our corpus: it learns a representation specific to our data, without external distribution bias.

2.6 Positioning and outline

This project sits at the convergence of all these traditions. It operates in Goodman's philosophical space (exhibited properties) rather than Danto's (institutional context); it recognizes with Bourdieu and Becker that reference labels are situated social constructs; and it proposes an engineering framework of *multimodal fusion* (low-level descriptors, CNN features, and probabilistic VAE representations, combined by weighted *alpha-blending*, reduced by PCA and partitioned by K-Means, complemented by agglomerative hierarchical clustering and a quality filter via Silhouette score) to propose an emergent partition independent of these constructs. The central question is as follows: *do the emerging clusters correspond to the categories produced by art history and criticism, and if not, what do these divergences teach us?* The remainder of the report describes the pipeline (Section 2), presents the results (Section 3), confronts them with historical categories and human observers (Section 4), before concluding (Section 5).

3 Dataset and Preprocessing

3.1 Description of the study corpus

The database used consists of 1,700 abstract artworks. While this corpus offers valuable diversity for an unsupervised classification approach, preliminary quantitative analysis reveals several structural limitations and significant methodological biases. It is necessary to formalize these in order to rigorously interpret the topology of the resulting clusters.

3.2 Distribution bias and labeling anomalies

First, a significant imbalance in the distribution of artworks per artist is observed. While some painters are heavily overrepresented in the dataset, others are documented by only a single canvas. In the context of unsupervised *clustering*, this numerical heterogeneity constitutes a major bias: the model risks grouping artworks based on the statistical recurrence of features specific to an overrepresented artist (low intra-artist variance), rather than on universal and transversal stylistic criteria.

Furthermore, the *dataset* presents critical labeling anomalies: several strictly identical artworks appear in duplicate under different artist attributions. Although our algorithmic approach explicitly discards nominal metadata during the learning phase, these label collisions complicate the validation phase and the evaluation of the semantic relevance of the obtained classes.

3.3 Heterogeneity of formats and geometric impact of preprocessing

Another major technical constraint lies in the great variability of the original geometric formats, as the artworks present highly heterogeneous aspect ratios (height/width). To be encoded by standardized computational vision architectures (such as CNNs), these images must mandatorily be resized to uniform matrix dimensions.

To compensate for this divergence without squashing or anamorphosing the spatial structures of the artworks, reflection *padding* had to be applied. From an aesthetic point of view, these manipulations alter the integrity of the original artwork. From a computational point of view, they introduce non-negligible artificial artifacts:

- **Symmetry artifacts:** The duplication or mirroring of borders creates patterns of axial symmetry entirely non-existent in the initial composition. This risks distorting texture and spatial organization descriptors (such as edge analysis or local entropy).
- **Chromatic distribution bias:** The replication of peripheral zones artificially overrepresents certain color palettes at the expense of the chromatic balance initially intended by the painter.

Consequently, the impact of these technical preprocessing choices must be systematically considered during the analysis of the classifier's discriminative criteria and the confrontation of the results with our own human observation.

4 Feature Extraction

In order to extract relevant descriptors, we implemented three approaches inspired by the state of the art. For each set of features, a Principal Component Analysis (PCA) was applied to reduce dimensionality.

4.1 Geometric Approach

Color Features

The chromatic dimension is essential in abstract art. The following features aim to quantify the use of color, brightness, and overall chromatic composition:

- **HSV Histograms**: Analysis of color distribution in the (Hue, Saturation, Value) space, capturing the dominant tonality and color intensity.
- **Lab Statistics**: Calculation of statistical moments (mean, standard deviation, skewness, kurtosis) in the perceptual

L*a*b* space, modeling human perception of color contrasts.

- **Dominant Palette**: Extraction of main colors and their proportions via the K-Means algorithm, reflecting the artist's chromatic signature.
- **Spatial Brightness**: Distribution of luminance over a spatial grid, capturing the balance of chiaroscuro in the composition.
- **Chromatic Richness**: Measurement of the diversity and variety of colors used in the artwork.
- **Physical Metadata**: Integration of file size and aspect ratio (portrait, landscape, or square format), providing information on the artwork's format and graphic complexity.

Texture Features

Texture translates the materiality, brushstroke, and quality of the pictorial surface. The following descriptors were used:

- **Co-occurrence Matrices**: Calculation of contrast, correlation, energy, and homogeneity to quantify the spatial relationships between neighboring gray levels.
- **Local Binary Patterns (LBP)**: Multi-scale analysis of local surface roughness, effective for detecting micro-textures.
- **Gabor Filters**: Use of a filter bank simulating the human visual system's response to specific spatial frequencies and orientations.
- **Autocorrelation**: Measurement of the repetition and periodicity of spatial patterns in the image.
- **Spectral Analysis**: Identification of dominant spatial frequencies and peaks in the frequency domain, revealing the rhythm of the texture.
- **Repetition**: Explicit quantification of the regularity and repetitiveness of pictorial patterns.

Composition and Structure

This pillar analyzes spatial composition, shapes, structural complexity, and the organization of the artwork:

- **Histogram of Oriented Gradients (HOG)**: Captures the local distribution of gradient intensities.
- **Edge and Shape Analysis**: Edge detection (Deriche filter), contour extraction, geometric shape analysis, and straight line detection (Hough transform).
- **Grid Analysis**: Study of information distribution (symmetry, centrality, gradients) across a spatial grid dividing the image into sub-regions.
- **Global Symmetry**: Measurement of the degree of axial or central symmetry of the composition.
- **Directional Coherence**: Use of the structure tensor to evaluate the alignment and strength of local structures (e.g., a strong, directional brushstroke).
- **Fractal Complexity**: Calculation of the fractal dimension to estimate the complexity, irregularity, and self-similarity of patterns and contours.
- **Visual Disorder**: Measurement of Shannon entropy to quantify uncertainty and information density, complemented by statistical analysis of gradient magnitude.
- **High-Level Frequency Features**: Analysis of the power spectrum slope (indicating high-frequency energy decay) and energy in high frequencies, reflecting the level of fine detail in the composition.

4.2 VAE Approach

The VAE (Variational Auto-Encoder) approach quickly appeared to us as an ideal solution for abstract art classification. Two main reasons motivated this choice:

- Without wanting to generate new images, a VAE can offer us a direct visualization of the images in the latent space.
- A VAE is a non-linear extension of PCA, which can therefore allow us to capture features (other than geometric features) in our images.
- It relies on unsupervised learning: without labels on our images (other than the artist's name), a VAE allows us to group certain paintings together automatically.

Since we want the rendering of our latent space to give us something compact and continuous (i.e., with no gaps between our latent space vectors), we chose to use a β-VAE to further penalize reconstruction errors so that our latent space resembles a standard normal distribution as much as possible.

Likewise, in the case of VAEs, training can take a significant amount of computation time. This is why we chose to resize our images to a 256×256 pixel format.

4.2.1 *Construction of the β-VAE*

For the initial construction of our VAE, we chose to sequentially apply 5 convolutional layers, each halving the dimension of our images. The output of the last convolution gives us: $8 \times 8 \times 512$, which once flattened results in vectors of size 32768.

Because we want to obtain vectors of size LATENTDIM = 128, we pass these vectors one final time through linear layers: one layer that outputs the **mean** μ as well as the **log of the variance** $\log(\sigma^2)$.

We use the reparameterization trick to generate a point sampled from the normal distribution with parameters (μ, σ^2).

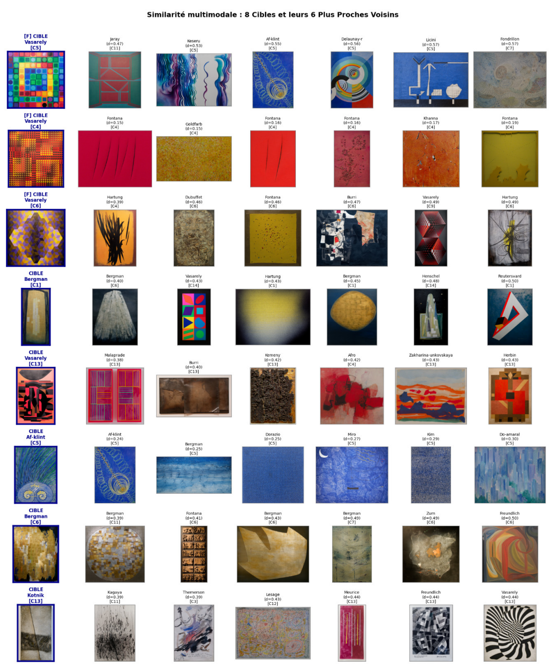


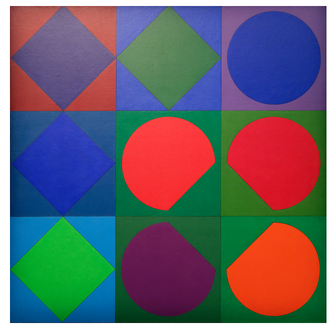
Figure 2: Nearest neighbors of each image. Similarities in shapes and colors are visible.

Here we can easily identify the first problem with our VAE approach: geometric structures are not preserved within the different clusters. Instead, our implementation tends to group images from a colorimetric point of view.

After analyzing our code and how a VAE works, we can attribute this result to a single factor: the loss function. Indeed, for a classic VAE, the loss function is Pixel-MSE, meaning it will primarily penalize color discrepancies on the image. If, for example, a spot of color is shifted by a few pixels on another image, the loss will be large.

However, as stated in the introduction, when classifying abstract art, one must primarily account for a lyrical abstraction (focused on color) and a geometric abstraction (focused on shapes). Thus, we understand that a whole set of features was not taken into account by the VAE.

(a) (Pollock) - Focused on colorimetry



(b) (Vasarely) - Focused on shapes

4.2.2 *Better Feature Extraction*

In our current VAE, during reconstruction, our model tends to generate blurry images. Indeed, because the penalization is primarily based on color differences and not on the global structure, our model will often generate blurry images with solid colors in order to minimize this loss.

Since we want to avoid this, we introduced a **new loss function**, relying on a convolutional network. Here we take the VGG model, pre-trained on ImageNet, which has thus learned to detect structures present in an image (lines, contours, contrasts, textures, ...). The loss function operates as follows:

- we pass the original image through the VGG network;
- we pass the image generated by the VAE through the VGG network;
- we calculate the difference between the two outputs of the VGG network.

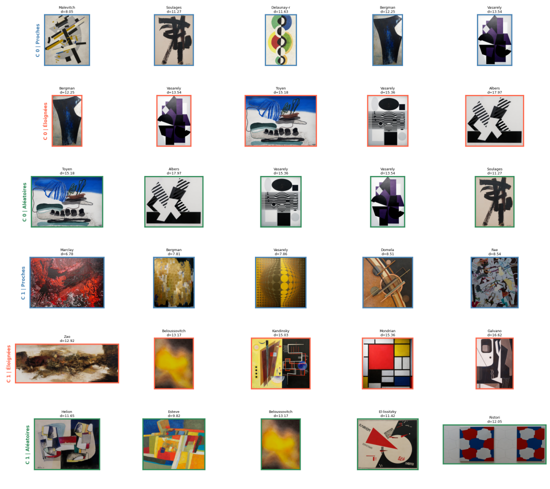


Figure 4: VAE results using a VGG loss.

Error at last epoch	Classic VAE	VAE + VGG
Train	12561.7	6823.2
Validation	12529.2	6551.1
Reconstruction	11540.4	6064.7
KL	255.3	189.6

Table 1: Error comparison between the classic VAE and the VAE with VGG loss.

4.2.3 Outlook: Use of a Siamese Space

Although fusing the spaces of the β-VAE (focused on colors and global style) provides a very rich feature base, this combined space suffers from a major limitation: it has no notion of artistic meaning or authorial identity. Vectors are grouped purely statistically and mathematically.

This is where the Siamese network comes in, aiming to metamorphose this raw space into a discriminative metric space. Its usefulness rests on three fundamental pillars:

- **Semantic merging (pairwise refinement):** By training with a contrastive loss (*Contrastive Loss*), the Siamese network explicitly learns to draw artworks by the same artist closer and push away those by different artists. It forces the latent space to organize itself not only by raw visual appearance, but by stylistic authorial consistency.
- **Creation of ultra-clear clusters:** As highlighted by the final t-SNE visualization of the project, the Siamese space transforms an initially diffuse and mixed point cloud (the fused space) into extremely distinct and separated artist clusters.
- **Few-shot learning (rare artist classification):** This is the greatest strength of this approach. In an abstract art dataset where many artists only have two or three available artworks, a standard classifier (like a large neural network) would fail due to lack of data. The Siamese space solves this problem by enabling prototype classification (nearest neighbor method): it is sufficient to calculate the average of an artist's Siamese vectors to create their signature, allowing the identification of a new artwork with remarkable precision, even if the artist is very rare.

In summary, the Siamese network acts as the expert translator for our model. It takes raw visual descriptors to create a space of meaning, essential for authentication and artistic style recognition. We were not, however, able to fully implement it before submitting the report, as it is a semi-supervised method and thus does not strictly fit the project criteria. That is why we do not present results for it here.

4.3 CNN Approach

The use of pre-trained convolutional neural networks (CNNs), although standardized in computer vision, poses a major challenge in the context of abstract art. Classic architectures trained on ImageNet optimized their filters to detect semantic objects. However, abstract art is precisely defined by the absence of identifiable objects. Directly extracting vectors from the final *Fully Connected* layers would introduce a strong semantic bias: the network would attempt to group paintings by seeking illusions of objects (machine pareidolia) rather than analyzing pictorial technique. To circumvent this bias, we implemented an approach inspired by Neural Style Transfer works, focusing exclusively on texture correlations.

Architectural choice: VGG16 over ResNet

Although more recent and deeper architectures like ResNet dominate classification tasks today, we deliberately chose to base our extractor on **VGG16**. The ResNet architecture relies on residual connections (*skip connections*) that force the addition of activations from shallow layers with those from deep layers. This fusion mixes texture information (low level) and semantic information (high level), making the isolation of pure style mathematically unstable. Conversely, the strict sequential architecture of VGG16 allows for clear filter hierarchization: each convolutional block encodes a precise texture scale, free of residual semantic pollution.

Stylistic Signature Extraction via Gram Matrices

Rather than using raw activation maps, we extract the style of each painting via the **Gram Matrix** of feature maps at different depths in the network. For a given layer producing a tensor F with dimensions $C \times H \times W$ (C channels, $H \times W$ spatial dimensions), the Gram matrix G of size $C \times C$ is defined by:

$$G_{ij} = \frac{1}{C \cdot H \cdot W} \sum_{k=1}^{H \cdot W} F_{ik} F_{jk}$$

This dot product between channels cancels out spatial information (the localization of the brushstroke) to retain only correlation information: the nature of the texture and color interaction.

In practice, our `VGGStyleExtractor` class intercepts activations at the output of four successive VGG16 convolutional blocks, obtained by slicing the network's `features` at indices `[:4]`, `[4:9]`, `[9:16]`, and `[16:23]` in the PyTorch implementation. These extraction points correspond respectively to the outputs of `conv1_2`, `conv2_2`, `conv3_3`, and `conv4_3`. This choice covers a complete spectrum ranging from micro-textures (canvas grain, fine brushstrokes) to macro-structures (color flats, overall chromatic organization). To optimize memory space, only the upper triangle of each symmetric matrix is retained via `triu_indices` and flattened into a single vector prior to concatenation.

Robustness via Spatial Sampling

Abstract art presents high intra-image variability: a Pollock painting is not stylistically uniform from one corner to another. To obtain a robust signature, we adopted a strategy of extraction via random *patches*. For each artwork, **10 random regions** (*random crops*) of size 256×256 pixels are extracted, each normalized according to ImageNet statistics. The style vector (the concatenation of the Gram matrices from the 4 layers) is computed independently for each patch via `torch.no_grad()`, ensuring inference without weight updates. The final signature of the artwork corresponds to the **average of these 10 vectors**, producing a representation invariant to scale and decentralized compositions.

Dimensionality Reduction and Normalization Pipeline

The raw stylistic vectors resulting from the concatenation of the four Gram matrices exhibit very high dimensionality. We applied the following reduction pipeline, in this order:

1. **Missing value correction**: replacement of any potential `NaN` values by the median of each dimension.
2. L_2 **Normalization**: harmonizing the energy of the vectors (`Normalizer(norm='l2')`) to prevent extreme local contrasts from dominating the space.
3. **Principal Component Analysis (PCA)**: projection onto the **50 principal components**, retaining the core of the stylistic variance.
4. **Standardization**: centering and scaling (`StandardScaler`) to guarantee equivalent weight for each latent dimension.

On this 50-dimensional latent space, we attempted to jointly study the elbow method (intra-cluster inertia) and the Silhouette score to select an optimal number of clusters, but this was modified during the fusion of approaches.

5 Fusion of Approaches

Before any weighting, it is imperative to harmonize the scales of the three subspaces to prevent one from dominating the distance calculations simply due to its raw variance. Each matrix first undergoes an L_2 normalization (projecting vectors onto a unit hypersphere), followed by standardization (centering-scaling to impose a mean of zero and unit variance). Only on these harmonized spaces do the semantic weights hold true meaning.

The three reduced spaces are then concatenated following weighting (see Equation 1):

$$\mathbf{X}_{\text{final}} = \left[\tilde{w}_{\text{cl}} \cdot \mathbf{X}_{\text{cl}} \;\middle\|\; \tilde{w}_{\text{CNN}} \cdot \mathbf{X}_{\text{CNN}} \;\middle\|\; \tilde{w}_{\text{VAE}} \cdot \mathbf{X}_{\text{VAE}} \right] \in \mathbb{R}^{n \times 3d} \tag{1}$$

where the normalized weights are defined by:

$$\tilde{w}_\star = \frac{w_\star}{w_{\text{CNN}} + w_{\text{VAE}} + w_{\text{cl}}} \tag{2}$$

Multiplication by the weight prior to concatenation is equivalent to scaling each block of columns: a source with a weight twice as high contributes twice as much to the Euclidean norm of $\mathbf{X}_{\text{final}}$, and thus twice as much to the K-means inertia.

This formulation presents two practical advantages. On the one hand, the weights are relative integers $(w_{\text{CNN}}, w_{\text{VAE}}, w_{\text{cl}}) \in \mathbb{N}^3$, making the search grid finite and interpretable. On the other hand, the dimension of $\mathbf{X}_{\text{final}}$ remains constant ($3d$) regardless of the weight configuration, ensuring the comparability of Silhouette scores across configurations.

5.1 Topological validation of the unified space

Before applying a discrete partitioning algorithm such as K-means, it is essential to verify whether the continuous space $\mathbf{X}_{\text{final}}$

convincingly encodes a coherent stylistic semantics. Echoing questions regarding the machine's decision criteria, we calculated their inter-artist cosine distance matrix.

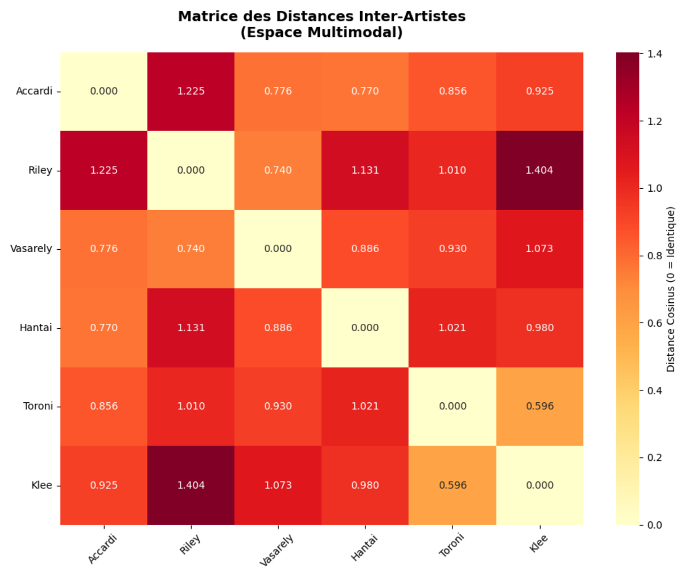


Figure 5: Matrix of cosine distances between the centroids of reference artists in the fused space $\mathbf{X}_{\text{final}}$. A distance close to 0 indicates strong stylistic similarity.

Analysis of this matrix (Figure 5) validates our multimodal fusion pipeline, as the mathematical alignments closely match art history:

- **The optical and geometric art pole:** Bridget Riley and Victor Vasarely exhibit the lowest cosine distance in the matrix (0.776). Our space perfectly captured the essence of Op Art (geometric rigor, edge contrast, recurring patterns) and naturally associates them.
- **The break with lyrical abstraction:** Conversely, these same geometric artists (Riley, Vasarely) show extreme distances (reaching up to 1.404) with Klee.

This step confirms that our super-vector does not simply stack features, but has organized abstract art into a valid topological *continuum*. The role of K-means, detailed in the following section, will be to discretize this continuous space into distinct movements.

5.2 K-means Classification

The K-means algorithm minimizes intra-cluster inertia:

$$\mathcal{J}(C_1, \ldots, C_K) = \sum_{k=1}^{K} \sum_{\mathbf{x}_i \in C_k} \|\mathbf{x}_i - \boldsymbol{\mu}_k\|_2^2 \tag{3}$$

where $\boldsymbol{\mu}_k = \frac{1}{|C_k|} \sum_{\mathbf{x}_i \in C_k} \mathbf{x}_i$ is the centroid of cluster k. We use a `k-means++` initialization with $n_{\text{init}} = 10$ repetitions to limit sensitivity to random initialization.

Distances to the centroid provided by `sklearn`'s `transform` method are used to order the images within each cluster: a low distance indicates an image close to the cluster's stylistic prototype, a high distance indicates a boundary case.

5.3 Hyperparameter search

The pipeline involves two families of hyperparameters: the combination weights and the number of clusters K. Their search follows a sequential two-step protocol, followed by a final visual validation.

Step 1 — Weight calibration at fixed K

The weights (w_{CNN}, w_{VAE}, w_{cl}) are explored empirically. Lacking absolute ground truth, hyperparameter optimization was driven by an iterative heuristic evaluation. Inspection of the nearest neighbors space for complex queries allowed us to adjust the weightings. The optimal balance was achieved with the configuration ($w_{\text{CNN}} = 0.6$, $w_{\text{VAE}} = 0.2$, $w_{\text{cl}} = 0.2$).

Step 2 — Selection of K with fixed weights

With the optimal weights fixed, the number of clusters is explored over the range $K \in \{3, \ldots, 50\}$. For each value of K, the same method as before was employed (displaying the classes in addition). We tried to supplement it with **inertia** which allows for the application of the elbow method.

Ultimately, the result was not convincing (highly variable depending on the previous weight combination). We therefore opted for a more manual method where we observed, for each cluster, the images closest to the centroid, the images furthest away, and a few random images. We found that the result was convincing starting from a cluster count of $K = 20$, and that it didn't significantly improve by increasing K (it created superfluous and/or heterogeneous classes instead of separating classes that clearly combined two styles separable by the naked eye).

To validate this K, we also inspected the nearest neighbors by cosine distance for a set of reference images corresponding to artists designated by the jury: Accardi, Riley, Vasarely, Hantaï, Toroni, and Klee.

$$d_{\cos}(\mathbf{u}, \mathbf{v}) = 1 - \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\|_2 \, \|\mathbf{v}\|_2} \tag{4}$$

Cosine distance is preferred over Euclidean distance because it is insensitive to the absolute amplitude of vectors: only the direction in the feature space matters, which is consistent with comparing stylistic profiles rather than raw intensities. The choice of cosine distance for evaluation is also mathematically consistent with our K-means partitioning: since our vectors previously underwent L_2 normalization, minimizing the Euclidean distance (the core of K-means) is formally equivalent to maximizing cosine similarity.

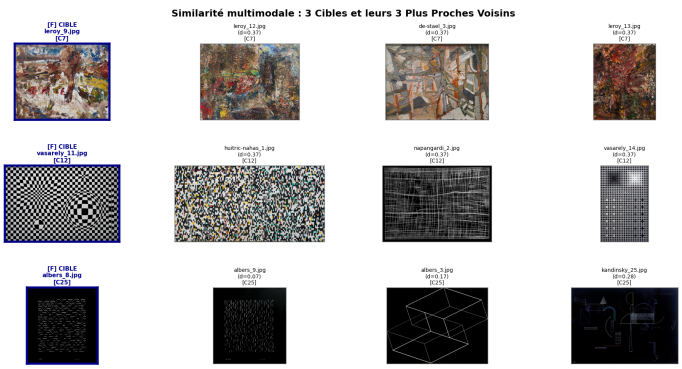


Figure 6: Validation of the fusion space via nearest neighbor extraction. The queries deliberately target stylistic extremes (extreme materiality, optical instability, geometric minimalism).

As illustrated in Figure 6, the final representation successfully captures complex stylistic nuances without being fooled by visual noise. This panel of artists was specifically designed to test the algorithm against various challenges: Leroy's chaotic texture, Vasarely's high spatial frequencies, and Albers's strict color flats. Finally, the nearly systematic alignment between the nearest neighbors and cluster labels $[C_k]$ proves that the continuity of our similarity space translates faithfully into our discrete classification.

5.4 Diagnosis of cluster heterogeneity

Despite the overall relevance of the obtained partition, an expert visual audit of the clusters revealed a fundamental limitation of the K-means algorithm applied to abstract art. The algorithm implicitly assumes that clusters are spherical, isotropic, and of comparable densities. However, art history reveals a very different topology: certain movements (like geometric abstraction or Op Art) are highly standardized and form dense subspaces, while others (like Action Painting or Art Informel) present considerable intra-class variance.

This topological asymmetry results in the creation of catch-all classes. Artworks located at the boundaries of these high-variance movements are mathematically absorbed by clusters whose stylistic essence they do not share, thus degrading the class's visual coherence. To objectify this phenomenon, we relied on the individual Silhouette score $s(\mathbf{x}_i) \in [-1, 1]$, which evaluates an artwork's anchorage in its own cluster compared to the closest neighboring cluster.

5.5 Adaptive skimming via core analysis

To purify heterogeneous clusters without amputating naturally dense ones, filtering by a global threshold proved ineffective. We therefore implemented an adaptive skimming method based on the analysis of the *hard core* (Core Silhouette) of each cluster.

For a given cluster C_k, the goal is to define an exclusion threshold dictated solely by its most characteristic representatives, thus ignoring peripheral noise that would drag the average down. Let $S_k = \{s(\mathbf{x}_i) \mid \mathbf{x}_i \in C_k\}$ be the set of silhouette scores for cluster k. We define the core $\mathcal{N}_k \subset C_k$ as the upper half of the distribution:

$$\mathcal{N}_k = \{\mathbf{x}_i \in C_k \mid s(\mathbf{x}_i) \geq \mathrm{median}(S_k)\} \tag{5}$$

We then compute the mean $\mu_{\mathcal{N}_k}$ and the standard deviation $\sigma_{\mathcal{N}_k}$ of the scores strictly on this core. The tolerance threshold τ_k for cluster k is then defined by a statistical cut-off, secured by a parameterizable floor $\tau_{\min}$:

$$\tau_k = \max\left(\tau_{\min},\ \mu_{\mathcal{N}_k} - \lambda \cdot \sigma_{\mathcal{N}_k}\right) \tag{6}$$

where λ is a severity hyperparameter (set empirically to 2.5) and $\tau_{\min}$ is an adjustable parameter defining the minimum tolerated silhouette score. Any artwork $\mathbf{x}_i \in C_k$ satisfying $s(\mathbf{x}_i) < \tau_k$ is rejected from its initial class and reassigned to a global pool of outliers, denoted C_{-1}. This method allowed us to restore exceptional visual purity to the core of our K main clusters.

5.6 Recursive classification of outliers

Set C_{-1} groups the artworks rejected during skimming. However, assuming that this set consists only of absolute noise would be an analytical error. From the perspective of art sociology, these outliers can represent minority sub-movements, transitional schools, or style fusions that were statistically crushed by major movements during the initial K-means pass.

To verify this hypothesis, we isolated the sub-matrix $\mathbf{X}_{\text{marginaux}} \subset \mathbf{X}_{\text{final}}$ corresponding to the artworks in C_{-1}, and applied an independent sub-clustering to look for new topologies. This recursive approach revealed K' new subclasses (denoted $A.0, A.1, \ldots$). The emergence of logical groupings among these outliers (for example, the gathering of canvases by Dubuffet and Alechinsky initially scattered) confirms that the initial rejection was indeed symptomatic of unexplored stylistic bridges rather than a simple model error.

5.7 Assessment: Unified stylistic distribution matrix

The integration of the skimmed clusters and the new subclasses results in a final partition. In order to evaluate the extent to which this taxonomy succeeds in identifying artists' signatures, we generated a stylistic distribution matrix (Figure 7).

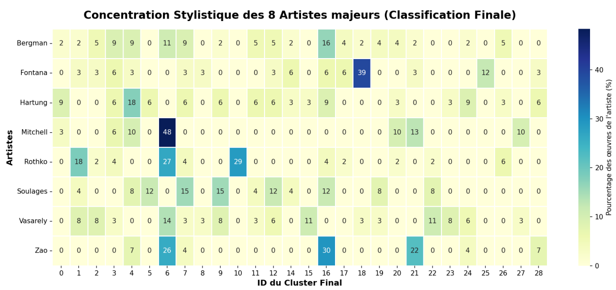


Figure 7: Stylistic concentration of the 8 major artists. The strong concentrations validate the model's ability to isolate the visual identity of the authors.

Analysis of this tool validates our pipeline: cells showing strong concentrations indicate that the model successfully isolated the visual identity of an artist blindly, thus freeing itself from any metadata.

6 Visualization of Results

6.1 2D projection via PaCMAP

Once the final partition is obtained, visualizing the K_{total} clusters in a low-dimensional space constitutes an essential step to qualitatively evaluate our pipeline's coherence. The direct projection of $\mathbf{X}_{\text{final}} \in \mathbb{R}^{n\times 3d}$ being impossible to interpret visually, we resort to a dimensionality reduction to $\mathbb{R}^2$.

We opted for **PaCMAP** (*Pairwise Controlled Manifold Approximation* [Wang et al., 2021]) rather than classical t-SNE. t-SNE exclusively optimizes local structure—close neighbors in the original space remain close in the projection—but systematically distorts inter-cluster distances: two groups far apart on the t-SNE map are not necessarily far apart in the real space. PaCMAP corrects this limitation by explicitly balancing three types of pairs during optimization:

- **Close pairs**: maintenance of local coherence within each cluster.
- **Mid-distance pairs**: preservation of mesoscopic structure between neighboring clusters.
- **Far pairs**: conservation of the global topology of the corpus.

This property is particularly valuable in our context: the relative position of the clusters on the map faithfully reflects the cosine distances calculated in $\mathbf{X}_{\text{final}}$, validated in Section 5.1. An observer can thus interpret the spatial proximity of two clusters not only as local similarity, but as true stylistic continuity in the feature space.

The 2D coordinates produced by PaCMAP are then linearly normalized to the interval $[-60, 60]$ to ensure compatibility with the interactive rendering engine (see Section 6.2).

6.2 AbstractViz: Interactive visualization platform

The classification results are showcased via **AbstractViz**[1], an interactive web application developed specifically for this project, relying on **deck.gl** (WebGL rendering via an orthographic view) on the client side and **Flask** on the server side.

Technical architecture

Each artwork is represented by a point on the PaCMAP plane, colored according to its cluster or its artist at the user's choice. The 1,700 thumbnails are base64-encoded on the server side and transmitted to the client upon initial load; they substitute the points beyond a certain zoom threshold, offering a smooth transition between the overview and individual inspection. A *bloom* mechanism locally disperses superimposed points at high magnification, allowing precise selection of artworks in dense areas.

Figure 8: 2D Map of the AbstractViz visualization.

Contribution to the evaluation

AbstractViz plays a dual role in our approach. On the one hand, it serves as the **visual validation** tool described in Section 5.3: inspecting neighbors and cluster composition directly guided the selection of hyperparameters K and λ. On the other hand, it allows for a **human confrontation** of the obtained taxonomy: a non-expert observer can navigate freely within the corpus, verify that artworks in a single cluster share perceptible coherence, and identify edge cases revealing stylistic boundaries.

[1] https://perso.telecom-paristech.fr/edionnet-24/

7 Analysis of the Final Classification

You can find this classification on the site `https://perso.telecom-paristech.fr/edionnet-24/` via AbstractViz, or view the complete classification here: `https://github.com/blastoncrush/projet-im06/tree/main/classification_finale`. You have at the following link: `https://share.rezel.net/s/BP3zjVKp` an image which presents, for each cluster (identified from C0 to C19 for base clusters and A.0 to A.9 for those added for outliers), in a row: the 5 images closest to the centroid, the 5 furthest, and 5 random images from the cluster.

Here is the breakdown cluster by cluster:

- **C0** · *28 artworks*: Dominant blue and yellow palette, horizontal movement. Some Dubuffets are present: although they look out of place, their presence can be understood through their movement and palette.
- **C1** · *31 artworks*: Large central shape, mostly rectangular.
- **C2** · *68 artworks*: Centered circle. The Lee and Gorin pieces on the border are understandable (shapes close to a circle).
- **C3** · *117 artworks*: Blue and yellow palette, present but controlled movement.
- **C4** · *165 artworks*: Movement and texture (embodied by Leroy), dull/dark palette.
- **C5** · *11 artworks*: Few artworks, all predominantly black with some white or orange linear shapes.
- **C6** · *188 artworks*: Rich and saturated color palette, movement (embodied by Mitchell).
- **C7** · *82 artworks*: Vertical movement, blue and yellow palette.
- **C8** · *48 artworks*: Small squares and rectangles (exclusively) of varied colors.
- **C9** · *78 artworks*: Smaller squares (like Klee). We would have preferred to keep only artworks where the squares are well-ordered, but the Leger-f painting present in the center introduces artworks with patterns (like Hartung, for example).
- **C10** · *29 artworks*: Vertical stacking of monochrome rectangles (embodied by Rothko).
- **C11** · *49 artworks*: Central subject composed of lines on a white background.
- **C12** · *35 artworks*: Effect of horizontally aligned elements, often lines. Featuring mainly warm colors.
- **C13** · *14 artworks*: A central vertical line, often the sole object of the painting, immersed in an almost uniform background color.
- **C14** · *35 artworks*: Central subject and vignetting background.
- **C15** · *84 artworks*: High dynamic movement (embodied by Pollock). Paintings heavily loaded with small patterns (like some Vasarelys) are also present, which is less convincing but understandable (they are present at the cluster's border). A form of rough symmetry is also present.
- **C16** · *121 artworks*: Horizontal motifs with a cold palette, notably blue, and clearly visible brushstrokes.
- **C17** · *56 artworks*: 2 or 3 geometric shapes with a very uniform rendering.
- **C18** · *108 artworks*: Lightly pressed strokes featuring some geometric shapes (which attracted the Vasarely to the border).
- **C19** · *29 artworks*: Vertical, thin polygons.
- **A0** · *52 artworks*: Fractal artworks with strong geometric shapes.
- **A1** · *47 artworks*: Cold colors forming erratic yet horizontal strokes with square motifs.
- **A2** · *24 artworks*: Squares and rectangles interwoven across the entire artwork.
- **A3** · *33 artworks*: Warm colors with small repeated and dense geometric motifs.
- **A4** · *44 artworks*: Chaotic motifs with a predominantly dark palette and heavily loaded with small, complex shapes.
- **A5** · *21 artworks*: Vertical motifs with dark colors.
- **A6** · *24 artworks*: Semi-circles and squares with a central subject.
- **A7** · *38 artworks*: Chaotic brushstrokes containing mostly colors and very high-frequency geometric shapes.

- **A8** · *32 artworks*: Lightly pressed brushstrokes on a white background with well-marked lines.
- **A9** · *9 artworks*: Red, black, or gray polygons on a white background.

8 Conclusion

This project aimed to tackle a challenge at the boundary of computer vision and art philosophy: classifying a complex corpus of 1,700 abstract paintings in an entirely unsupervised manner, relying exclusively on the raw visual signal and abstracting away any contextual or biographical metadata Goodman [1968], Danto [1964]. Our experiments highlighted the inadequacy of isolated approaches in capturing the semantic richness of abstraction. While a classic VAE equipped with a Pixel-MSE loss tends to favor purely colorimetric organization at the expense of geometric structures, introducing a loss based on the VGG16 network helped mitigate this bias by forcing the model to encode high-level features, such as textures and micro-structures. By harmonizing and fusing these probabilistic representations with texture Gram matrices derived from a sequential CNN and handcrafted geometric descriptors, our multimodal fusion pipeline via weighted *alpha-blending* succeeded in formalizing a continuous space endowed with genuine stylistic coherence.

Faced with the challenges of this corpus (intra-class variance, numerical imbalance, preprocessing artifacts), the obtained partitions demonstrate the relevance of our approach. While some emerging clusters faithfully match historical categories—thus validating the existence of visual signatures specific to certain movements—the observed divergences are equally instructive. They demonstrate that the machine breaks free from the social constructs of art history to group artworks according to strict formal criteria, whereas human classification often relies on the invisible intention of the artist.

Perspectives

Although the unified fused space offers an excellent statistical foundation, the lack of labels during partitioning limits its ability to finely group artworks based on an authorial signature or subtle stylistic kinship. A major advancement of this work lies in the conception of a methodological extension based on a **Siamese network**. By exploiting a contrastive loss (*Contrastive Loss*), this network makes it possible to metamorphose our raw space into a highly discriminative metric space. Although its full implementation falls outside the scope of the immediate quantitative results of this report, the theoretical foundations laid open three crucial perspectives for the computational analysis of abstract art:

- **Semantic merging by pairs:** Forcing the latent space to reorganize its distances to bring artworks of similar stylistic coherence closer together and repel antagonistic styles.
- **Ultra-clear clustering:** Transforming the continuous and sometimes diffuse point cloud of the fused space into cleanly isolated artist clusters, as suggested by our initial t-SNE projections.
- **Few-Shot Learning via prototypes:** Solving the critical problem of rare artists (those with only one or two canvases in the dataset) by computing "prototype" signature vectors capable of identifying a painter's style from an extremely limited number of examples.

Ultimately, this project demonstrates that while the machine cannot capture the invisible "intention" of the artist in Danto's sense, it proves to be a powerful tool for revealing the formal structure of abstract art, paving the way for authentication and recommendation systems free from traditional historiographical biases.

References

Mebarka Allaoui et al. Considerably improving clustering algorithms using UMAP. *Lecture Notes in Computer Science*, 2023.

Rudolf Arnheim. *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press, 1974.

Daniel E. Berlyne. *Aesthetics and Psychobiology*. Appleton-Century-Crofts, 1971.

Giovanna Castellano and Gennaro Vessio. Deep convolutional embedding for digitized painting clustering. *arXiv:2003.08597*, 2020.

Arthur Danto. The artworld. *Journal of Philosophy*, 61(19), 1964.

Nelson Goodman. *Languages of Art*. Bobbs-Merrill, 1968.

Adrian Lecoutre, Benjamin Negrevergne, and Florian Yger. Recognizing art style automatically in painting with deep learning. In *Asian Conference on Machine Learning*, pages 327–342, 2017.

Helmut Leder, Benno Belke, Axel Oeberst, and Dorothee Augustin. A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95:489–508, 2004.

Pierre Lelièvre and Peter Neri. A deep-learning framework for human perception of abstract art composition. *Journal of Vision*, 21(5):9, 2021.

Colin Martindale. *The Clockwork Muse: The Predictability of Artistic Change*. Basic Books, 1990.

Vaibhav Rathi, Ankit Sharma, et al. A survey of computational techniques for fine art painting classification. *Image and Vision Computing*, 161:105626, 2025.

Eleanor Rosch. Principles of categorization. In Eleanor Rosch and Barbara B. Lloyd, editors, *Cognition and Categorization*, pages 27–48. Lawrence Erlbaum Associates, 1978.

Babak Saleh and Ahmed Elgammal. A unified framework for painting classification. In *IEEE ICDMW*, pages 1254–1261, 2015.

Yingfan Wang, Haiyang Huang, Cynthia Rudin, and Yaron Shaposhnik. Understanding how dimension reduction tools work. *Journal of Machine Learning Research*, 22(201):1–73, 2021.

Morris Weitz. The role of theory in aesthetics. *Journal of Aesthetics and Art Criticism*, 15(1):27–35, 1956.

Semir Zeki. *Inner Vision: An Exploration of Art and the Brain*. Oxford University Press, 1999.