Chromatic Echoes:
a color-based quantitative analysis of art history

Giulia Monteiro Silva Gomes Vieira
Universidade Federal de Minas Gerais
Disciplina de Mineração de dados aplicada 2025.1

1 Introduction

The study of color in art history is a multifaceted discipline, encompassing the physics of light, the chemistry of pigments, and the cultural symbolism of specific hues. As detailed in further sections, color is arguably the most relative medium in art, its perception and meaning shaped by a complex interplay of scientific, technological, and cultural forces. Historical narratives have been constructed primarily through qualitative analysis of a limited canon of works, however, the digital age presents an unprecedented opportunity to augment this approach with large-scale quantitative analysis. By applying data mining techniques to vast collections of digitized artworks, we can move from anecdotal evidence to empirical patterns, testing hypotheses about color usage in a systematic and scalable manner. This projects aims to stress the characterization capability of color in paintings, quantitatively comparing results to art history's theories.

2 Methodologies of color analysis in Art History

The scholarly interpretation of color in art requires a diverse set of analytical frameworks:

1. **Formal Analysis:** Formalism is the foundational method of art historical analysis, focusing on the visual elements of a work—line - for example it's **color palette** -, and their **relationships**, **independent of external context**.

2. **Iconography and Symbolism:** This methodology investigates the **conventional meanings and symbolic content** within a specific cultural and historical context.

3. **Material usage and process reconstruction:** Non-destructive techniques like X-ray fluorescence spectroscopy (XRF) can identify the chemical elements present in a pigment without taking a sample. For a more detailed look, conservators can take a microscopic sample of paint to create a cross-section, which reveals the artist's layering technique. More advanced methods like Macroscopic X-ray powder diffraction (MA-XRPD) can map the precise distribution of different crystalline materials across a painting's entire surface, while High-Performance Liquid Chromatography (HPLC) can identify organic colorants in microscopic samples. This analysis can infer if paintings are seen differently today than when they were created, for example by paint oxidation.

A truly exhaustive analysis of color must move beyond a simple description of palette to encompass the work's formal structure, its symbolic content, its material composition, and the very nature of its perception. The most complete understanding emerges from a multi-layered "archaeological" approach that synthesizes these methodologies. No single method is sufficient because the color we see today is a composite artifact of the artist's original intent (formal and symbolic choices), the material properties of the pigments used, and the chemical transformations that have occurred over time. Nevertheless, to perform analysis of type 2 one must extensively catalog and annotate artworks based on their iconography, symbols and historical context; likewise, to perform analysis of type 3 one must extensive catalog and annotate artworks based on their material and techniques usage. Both cases require intensive human labor for interpretation and experimentation. Hence, despite the clear detailing advantages of utilizing a multi-layer approach, that is not yet scalable, and, therefore, does not utilize computational capabilities to their fullest. Furthermore, both methods rely on art history's preconceptions and interdisciplinary knowledge, which are themselves the subject of challenge in this project. Thus, **this work focuses on formal analysis of colors in paintings**, pushing the borders of what can be learned from context-free color recognition and pattern matching.

3 Color Theory

To comprehend the use of color in artworks, one must first comprehend color.

3.1 The attributes of color

To comprehend color, one must first comprehend its three fundamental attributes (Edwards, 2004):

- **Hue:** the pure color itself, its family name as it appears on the **color wheel**. It is the most basic quality of a color. The most common color-wheel used by artists today is the 12-hue color wheel, a direct descendant of Newton's circle that visualizes the relationships between colors in a subtractive (pigment-based) system (Call, 2016). It is composed of three hierarchical groups of colors: (1) Primary Colors, which cannot be created by mixing other pigments (red, yellow, and blue); (2) Secondary Colors, each created by mixing two primary colors, (orange, green, violet); (3) Tertiary Colors, created by mixing a primary color with an adjacent secondary color (these are given hyphenated names, such as yellow-green, blue-violet, and red-orange, which indicate their parentage).
- **Saturation (or Intensity/Chroma)**: This refers to the brilliance or dullness of a color. A highly saturated color is pure, intense, and **vibrant**. A de-saturated or low-intensity color appears **muted** or dull. Saturation is typically reduced by adding the color's complement or by adding gray, which creates a tone.
- **Value (or Brightness):** This describes the relative **lightness** or **darkness** of a color. A color's value is raised (lightened) by adding white, which creates a tint. Its value is lowered (darkened) by adding black, which creates a shade.

3.2 Color interpretation schemes

Fundamentally, color is not an intrinsic property of an object, but a perceptual phenomenon applied by the brain to the raw data of the world (Koenderink et al., 2020). For that reason, the systematic study of color has a long and often contentious history, marked by a schism between two distinct approaches: one rooted in objective physics and the other in subjective human experience.

The first tradition, the **descriptive-physical interpretation**, seeks to explain the objective mechanics of color. This scheme's origin can be traced to antiquity, with early theories documented in texts attributed to Aristotle, who proposed that all colors arose from a struggle between light (white) and darkness (black). This created a linear scale with yellow near white and blue near black, a model that privileged **value** over **hue** and persisted for nearly two millennia. This worldview was shattered in the late 17th century by Sir Isaac Newton. His experiments with prisms, detailed in his seminal 1704 work Opticks, demonstrated that white light is not pure but a composite of a spectrum of colored rays: red, orange, yellow, green, blue, indigo, and violet. By arranging these **hues** in a closed ring, Newton created the first **color wheel**, a revolutionary model that revealed relational properties, most importantly the concept of "complementary" or opposite colors.

Here, what is perceived as color is the visual sensation caused by specific wavelengths of light being reflected from an object's surface into our eyes, while other wavelengths are absorbed (Call, 2016). These wavelengths are a collective contribution of 15 distinct physical and chemical mechanisms that produce color (Nassau, 1987), which range from the simple excitations of incandescence (the light emitted by a heated object) and gas excitations (as in a neon tube) to the complex transitions within molecular orbitals that give organic dyes their color. Our perception of a red rose, for instance, is the brain's interpretation of long-wavelength light reflecting from its petals; the rose itself is absorbing most other wavelengths and is, in a physical sense, everything but red (St. Clair, 2017, p. 17).

In the second tradition, the **phenomenological-psychological interpretation**, the German poet and scientist Johann Wolfgang von Goethe mounted, in his book Theory of Colours (1810), a famous challenge to Newton. He rejected the purely physical explanation, arguing instead for a theory based on human perception and psychological experience (Edwards, 2004). Goethe focused on phenomena that Newton's model did not easily explain, such as colored shadows, after-images, and the powerful symbolic and emotional character of colors. His work, though often dismissed by the scientific establishment, was profoundly influential on artists, particularly Romantics like J.M.W. Turner, who were more interested in the subjective, emotional experience of color than its optical mechanics. To these authors, the cultural-linguistic framework through which we learn to see, and the environmental context in which a color is presented have direct influence on a color's interpretation.

The theory of linguistic relativity, or the Sapir-Whorf hypothesis, posits that the language we speak shapes our perception of the world, a concept strongly supported by studies in color terminology (Whorf, 1956; Deutscher, 2010). Languages vary in how they partition the color spectrum into named categories, which in turn influences cognitive processing (Regier and Kay, 2009). For example, Russian speakers, who have distinct basic terms for light blue (goluboj) and dark blue (sinij), can discriminate between these shades faster than English speakers, who use a single term, "blue" (Alimpieva, 1982a; Winawer et al., 2007). This "categorical perception" effect is also seen with the Himba people of Namibia, whose language groups blue and green under one term but has multiple terms for shades of green, making them slower to distinguish blue from green but faster at differentiating between greens (Roberson et al., 2005; Gilbert et al., 2006). Beyond perception, culture assigns symbolic meanings to colors that are learned and can be contradictory across societies; for instance, white is for weddings in the West but for funerals in many Eastern cultures, demonstrating that color perception is an act of interpretation embedded in social norms (Gage, 2009).

Furthermore, a color's appearance is profoundly influenced by its immediate visual environment, a phenomenon known as simultaneous contrast, where adjacent colors alter our perception of each other's **hue, value, and saturation** (Edwards, 2004). The artist Josef Albers (1971) dedicated his work to demonstrating this principle, famously stating that "color deceives continually". His core exercises prove color's relativity: one physically identical color can be made to look like two different colors when placed on different backgrounds, as the surrounding color pushes the central color's appearance toward its own complement. Conversely, two different colors can be made to look alike by placing them on a ground that visually "subtracts" their differences in hue and light. This **psycho-physiological** effect is caused by the after-image, where the eye, after staring at one color, projects its complement onto adjacent areas, showing that even a trained eye sees color not as a fixed fact but as a dynamic interaction (Edwards, 2004).

3.3 Color model choice

In the computational analysis of art, the choice of a color model is a foundational decision that profoundly impacts the relevance and interpretability of the results. While digital images are natively stored and displayed using the Red, Green, and Blue **(RGB) color model**, this system is fundamentally **machine-oriented**, designed to instruct electronic displays how to emit light (Color Schemer, n.d.). For the art historian, whose goal is to analyze color as a perceptual and compositional element, the RGB model presents significant limitations. Its three values are not intuitive, as mixing them produces results that defy the logic of traditional pigment-based color theory; for instance, additively mixing red and green light results in yellow, a counter-intuitive outcome for anyone trained in subtractive paint mixing (Color Schemer, n.d.). Furthermore, the RGB values for a single perceived color can vary dramatically under different lighting conditions, making it an unstable basis for analysis (MDPI, 2024). For these reasons, a transition to a more perceptually aligned model is necessary for meaningful art historical data mining.

A suitable color model for art history analysis must both align with the descriptive-physical tradition - analyzing paintings through a measurable, optical property, which can be understood by computers -; and phenomenological-psychological approach - acting as clues to cultural meaning and perceptual context, which are art's history object of interest -.
The **Hue, Saturation, and Value (HSV) color model**, developed by Alvy Ray Smith in 1978, which reorganizes the RGB data into three components that directly correspond to the core tenets of artistic color theory, fits that intersection. The HSV model, often visualized as a cylinder or cone, directly maps to the way artists are taught to think about and manipulate color. Hue is represented as an angle, Saturation as the radius from the center, and Value as the height along the central axis, as seen in Figure 1. **This geometric arrangement means that similar colors have similar HSV values, allowing for more effective and meaningful color clustering and pattern recognition.** By translating the machine-centric RGB data into the perceptually relevant HSV space, we build a critical bridge between the quantitative capabilities of the computer and the qualitative, theory-driven questions of art history.

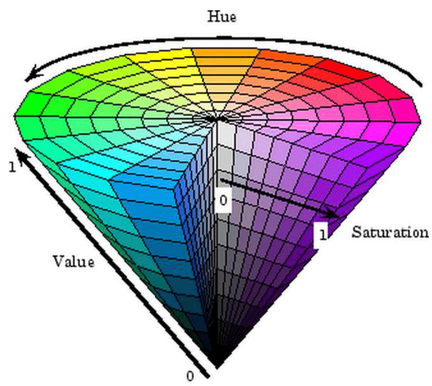

Figure 1: The HSV color model visualized as a cylinder, showing hue as angle, saturation as radius, and value as height.

4 The application of color in artistic practice

To derive meaning from independent color palettes, as proposed in the **formal analysis method** chosen in this work, general guidelines of color usage help filter inference possibilities. Despite not using contextual annotation for each artwork, when looking for specific patterns, assumptions can be made when taking in account: (1) general notions from formal academic artistic studies; (2) material constraints as pigment availability through time and space.

4.1 Artistic intent

Artists often organize their **palettes** according to established **relationships** on the **color wheel**. Regarding **hue**, the most prominent relationships are:

- **Monochromatic:** uses a **single hue with variations in values and saturation**. By limiting the palette, it produces a strong sense of unity and often a quiet, subtle mood. A powerful example is found in Pablo Picasso's "Blue Period" (1901–1904), in works like The Celestina (1903), Figure 2.
- **Analogous:** This scheme uses colors that are **adjacent on the color wheel**. Because the colors are closely related, this palette is inherently harmonious and often creates a serene and comfortable feeling. A notorious example is Vincent van Gogh's The Olive Trees (1889), Figure 3.

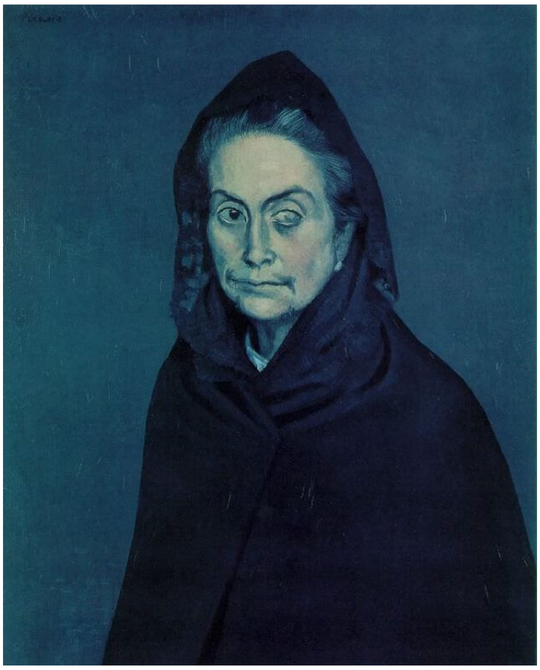

Figure 2: Picasso, Celestina (Monochromatic)

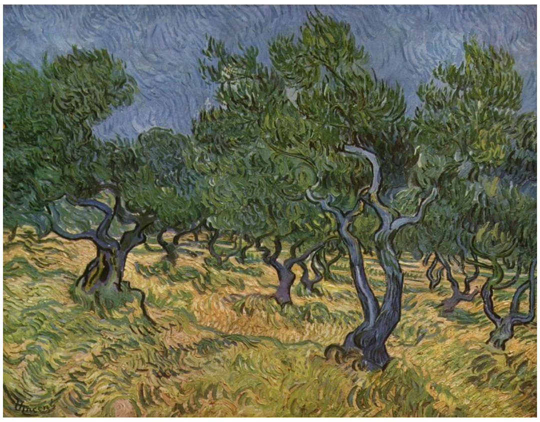

Figure 3: Van Gogh, The Olive Trees (Analogous)

- **Complementary:** This scheme uses colors **directly opposite each other on the wheel**. It produces the strongest possible contrast and visual excitement. When placed side-by-side, complements intensify each other, creating a shimmering or vibrating effect at their boundary, an effect known as simultaneous contrast (Edwards, 2004). Vincent van Gogh's 1888 masterpiece, The Bedroom at Arles 4, is a seminal example. Here, the complementary scheme is used to create a visual and emotional dissonance that is deeply unsettling.

- **Triadic:** This scheme uses three colors that are **equidistant on the color wheel**. A triadic palette is inherently balanced and vibrant. The Dutch artist Piet Mondrian famously used a primary triad in his neoplastic compositions, like Composition with Red, Blue, and Yellow (1930), Figure 5, to achieve what he believed was a universal, dynamic harmony reflecting the underlying structure of reality.

Figure 4: Van Gogh, The Bedroom at Arles (Complementary)

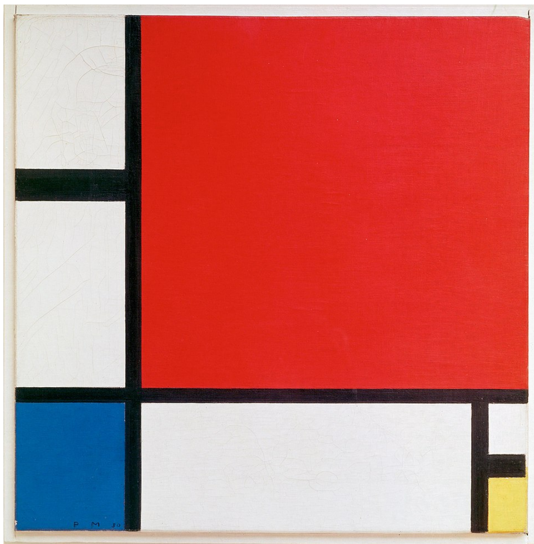

Figure 5: Mondrian, Composition with Red, Blue, and Yellow (Triadic)

Regarding **saturation** and **value**, juxtapositions of color can create **illusions of depth, transparency or luminosity**. The raw example of that color usage is Albers' series of paintings Homage to the Square, Figure 6, which started in 1949 and continued until his death. Nevertheless, key techniques used in previous art movements demonstrate the same property. For example: Renaissance's landscape painting, which relied heavily on atmospheric perspective (that mimics the effect of the Earth's atmosphere on distant objects, as objects recede into the distance, they appear lighter in value, and less saturated), for example The Baptism of Christ (1475), from Andrea del Verrocchio and Leonardo Da Vinci, Figure 9; and Reinaissance's *chiaroscuro* (literally "light-dark") / Baroque's *tenebrism* (from tenebroso, "dark" or "gloomy") - which refer to the use of strong value contrasts to model form and create a sense of drama and vivid three-dimensionality - for example The Baptism of St. Matthew (1599-1600), from Michelangelo Merisi da Caravaggio, Figure 8.

One important turning-point for the color use in art was the invention of photography. As now the machine depicted the real world with precision, avant-garde artists of the 19th and 20th century began to liberate color from its traditional descriptive role. Instead of using color to mimic the natural world, they began to use it for its purely expressive and emotional qualities, as stated by the artists of Fauvism, which can be seen in works like Matisse's Woman with a Hat (1905), Figure 7. The breakage of artistic rules infers the beginning of modernity in art.

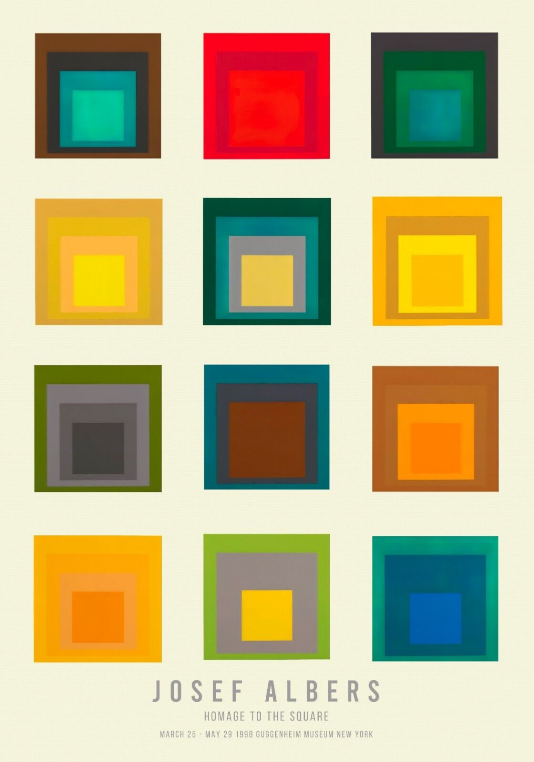

Figure 6: Alber, Homage to the Square collection set of paintings

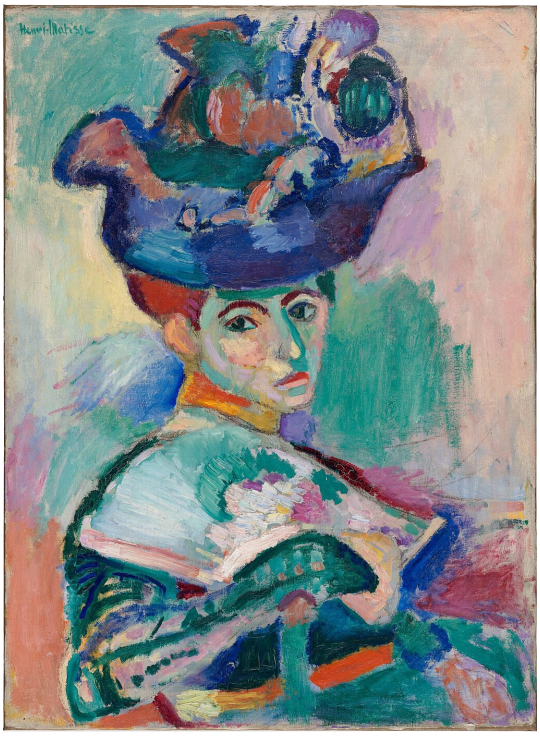

Figure 7: Matisse, Woman with a Hat

Figure 8: Caravaggio, The Calling of Saint Matthew

Figure 9: Verrocchio and DaVinci, The Baptism of Christ

4.2 Pigment availability constraints

More than artists' intents, the history of art is inseparable from the history of its materials. The aesthetic possibilities available to artists in any given era have always been defined by the pigments on their palette (St. Clair, 2017). This history is not a simple story of linear progress, but a dynamic feedback loop where artistic desire, economic forces, and technological innovation constantly influence one another. The demand for certain aesthetic effects drives innovation and trade, and the resulting supply of new materials enables new aesthetic movements, which in turn create new demands (Delamare and Guineau, 2000).

- **The Prehistoric Palette (c. 40,000 BCE):** The first artists worked with a limited palette derived directly from their immediate environment (St. Clair, 2017; Winsor & Newton, n.d.). This foundational palette consisted of five basic colors:**red and yellow** (from ochre clays rich in iron oxides); **brown** (from other earth tones); **black** (from burnt charcoal or the soot of burning animal fats); and **white** (from chalk). These earth pigments were ground into powders and mixed with binders like animal fat or saliva to create the first paints (Call, 2016).
- **The Medieval and Renaissance Palette (c. 500-1600):** The medieval and Renaissance periods saw a significant expansion of the artist's palette, pigments were sourced from a variety of materials: minerals like azurite (**blue**) and malachite (**green**), plants like saffron (**yellow**), and even insects like kermes, which produced a **crimson red** (World History Encyclopedia, 2020; St. Clair, 2017). Nevertheless, high-quality pigments remained rare, laborious to produce, and extremely expensive to source (World History Encyclopedia, 2020).
- **The Age of Discovery and Global Trade (c. 1500-1800):** The establishment of global trade routes by European powers following the Age of Discovery introduced a host of new and exotic colorants to the European palette, sparking new artistic possibilities. New pigments were brought from the New World and Eastern imports.
- **The Chemical Revolution and the Modern Palette (19th Century):** Before the Industrial Revolution, the prohibitive cost of pigments directly influenced artistic practice. Patrons' contracts often specified which expensive colors were to be used and on which figures, reserving ultramarine for the robes of the Virgin Mary to signify both her divine status and the patron's piety and wealth (St. Clair, 2017; World History Encyclopedia, 2020). The choice of pigment was as much a display of economic power as it was an aesthetic decision (World History Encyclopedia, 2020). Afterwards, the rapid advances in chemistry during the 19th century fundamentally transformed the artist's palette with **synthetic pigments**. For the first time, a vast array of stable, vibrant, and affordable colors became widely available. Furthermore, with the invention of the collapsible tin paint tube, the messy and inefficient practice of storing paint in pig's bladders was no longer necessary, and the tube allowed paint to be preserved for longer and, most importantly, to be **easily transported**.

Given this history, the use of color in art history is guided by specific material markers, that can be used as a starting point to comprehend aesthetic changes in paintings through time and space.

5 Related Work

The application of computational methods to the study of art history represents a significant paradigm shift, moving the field from purely qualitative, human-expert-driven analysis towards quantitative, data-driven inquiry at an unprecedented scale. The digitization of vast collections by museums, galleries, and online aggregators has created a wealth of data, enabling researchers to explore patterns in style, iconography, and artistic influence across thousands of works simultaneously. This transition, however, is not without its challenges. The very datasets that enable this research often carry inherent biases, and the computational techniques applied must be chosen and interpreted with a sensitivity to art historical context. This review synthesizes the foundational datasets and methodological approaches that define the field of computational art analysis, **with a specific focus on the role of color** in characterizing artworks by time, style, and geography. We survey the literature surrounding the two most prominent large-scale datasets, **OmniArt** and **WikiArt**, detailing the techniques applied to them and identifying the critical research gaps that motivate the present study.

5.1 OmniArt vs. WikiArt

The OmniArt dataset, developed at the University of Amsterdam, stands as a large-scale, museum-centric benchmark designed specifically for advanced machine learning applications. It represents an effort to aggregate and harmonize data from over 15 distinct online sources, including world-renowned institutional collections like the Rijksmuseum, the Museum of Modern Art (MoMA), and The Metropolitan Museum of Art, alongside user-generated content from platforms such as WikiArt and DeviantArt. In its initial release, it contained nearly half a million samples, with later versions expanding to over two million images, covering a vast historical range from 157 BCE to 2017. The principal strength and defining characteristic of OmniArt is its exceptionally rich and structured metadata. Core attributes include artist, creation period, style, school and artwork type and, most importantly for the present work, **explicit color palettes with weights**.

In parallel, the WikiArt dataset has become one of the most widely used resources in the field, primarily due to its scale and accessibility. Scraped from the public-domain art encyclopedia WikiArt.org, it contains tens of thousands of images, with some versions exceeding 250,000 artworks from over 3,000 artists. In contrast to OmniArt's deep metadata, WikiArt's primary value lies in its provision of simple, standardized labels for fundamental classification tasks: artist, style, and genre. This straightforward structure has made it the de facto choice for researchers training and benchmarking supervised classification models.

Further complicating its use are issues of data provenance and reliability. The metadata is generated through open, crowdsourced editing, which, while enabling its large scale, makes it "subject to mistakes and biases, and non-transparent moderation". This lack of curatorial oversight raises questions about the accuracy of its labels for fine-grained analysis. Finally, the dataset's lack of rich metadata on materials, techniques, or explicit color information severely limits the types of data mining questions that can be explored, a key distinction from the OmniArt benchmark.

5.2 Methodological Approaches to Artwork Characterization

5.3 Supervised Classification and Transfer Learning

The predominant task in computational art analysis has been supervised classification, where a model learns to predict an artwork's attributes, such as its style, genre, or artist, from its image.

Convolutional Neural Networks (CNNs), often paired with transfer learning, have become the standard for this task. In transfer learning, a model pre-trained on a massive, general-purpose image dataset like ImageNet is fine-tuned on the smaller, domain-specific art dataset. This approach leverages the powerful, generalized visual features learned from millions of images.

A significant body of work has used the WikiArt dataset for this purpose. Early work by Saleh and Elgammal (2015) introduced the dataset alongside a metric learning approach on handcrafted features for style and genre classification. More recent studies have employed deep CNNs. Pushing the state of the art further, Liu et al. (2022) applied Big Transfer (BiT) learning, fine-tuning a ResNet-v2 model pre-trained on the 14-million-image ImageNet-21k dataset. This approach achieved state-of-the-art results for artist, genre, and style classification on WikiArt, demonstrating the power of large-scale pre-training for art analysis.

On the OmniArt dataset, classification tasks have also been a focus, but often as part of a broader set of benchmark challenges. Strezoski and Worring (2018), who first introduced the dataset, established baseline scores for artist attribution, type, style, and school prediction using CNNs, providing a reference point for future research. The dataset has also been used to frame creation period estimation as a regression or classification problem, a task for which its broad temporal scope is particularly well-suited.

5.4 Advanced Architectures

While single-task classification has proven effective, researchers have increasingly turned to more advanced architectures to capture the complex, interrelated nature of artistic attributes. The rich, structured metadata of OmniArt has been particularly instrumental in facilitating this shift. The paradigm of Multi-Task Learning (MTL) is central to the OmniArt project's ethos. In their foundational paper, Strezoski and Worring (2017) proposed a deep learning model with a shared representation layer that is trained to predict multiple attributes—artist, style, and period—simultaneously. The underlying principle is that learning these related tasks together allows the model to leverage shared information, leading to improved generalization and performance compared to training separate models for each task. This approach was shown to outperform both single-task CNNs and older methods based on handcrafted features.

To better exploit the explicit relationships within art historical data, some researchers have moved beyond standard CNNs to Graph Neural Networks (GNNs). Worring et al. (2021) proposed a method that constructs a knowledge graph from artwork metadata, where nodes represent entities like artists and styles, and edges represent their relationships. By applying a GNN, the model can learn from both the visual features of the artwork and the contextual information embedded in the graph structure. This hybrid approach was shown to improve classification accuracy, particularly for classes with few examples (a common problem in art datasets), demonstrating the value of explicitly modeling the relational context of art.

More recently, the field has begun to explore the capabilities of the Transformer architecture and Multimodal Large Language Models (MLLMs). While transformers are widely used for style transfer, their ability to capture long-range dependencies in an image makes them a promising alternative to CNNs for representation learning. The newest frontier involves using vision-capable LLMs like GPT-4V for art analysis. Khadangi et al. (2025) propose using MLLMs on a subset of OmniArt not for classification, but to automate a formal, descriptive analysis of artworks, decoding aesthetic elements such as composition, technique, and **color usage**. This represents a significant shift from predictive tasks to generative, interpretive ones, moving closer to emulating the qualitative analysis performed by human art historians.

5.5 The role of color in prior computational art analysis

Despite its elemental importance, **the treatment of color in large-scale computational art analysis has been surprisingly limited**, often relegated to an implicit feature rather than a primary object of study.

5.5.1 Explicit Color Feature Engineering

The most significant work to center color as an explicit, engineered feature is the WikiArtVectors project by Desikan et al. (2022). The authors begin with the premise that standard deep learning models, while powerful, often fail to effectively utilize color information when representing images. To address this gap, they developed a novel pipeline to create explicit vector representations for both the style and color of each painting in the WikiArt dataset. For style, they used a conventional approach: features were extracted from a pre-trained VGG19 CNN and then reduced using Principal Component Analysis (PCA) to a 256-dimension vector. Their key innovation was in the treatment of color. For each artwork, they generated an eight-dimensional color-distributional vector by transforming the image into a perceptually uniform colorspace, which better reflects human color perception, and then binning the color values into a probability distribution. The stated goal of this project was not to build a predictive model but to provide a new dataset resource for social scientists, digital humanists, and cultural analysts. These vectors are intended to be used with information-theoretic and distance metrics to "identify large-scale patterns" and "explore relationships" across art, filling a crucial gap in the digital humanities literature. This work is foundational because it provides a clear methodology for extracting explicit, perceptually meaningful color features from artworks.

5.5.2 Implicit and Metadata-Based Color Usage

Outside of the WikiArtVectors project, the treatment of color in the literature is far less direct. The vast majority of classification studies that use CNNs on either WikiArt or OmniArt utilize color only implicitly. The raw RGB pixel data of the image is fed into the network, and the model learns to extract whatever features are most discriminative for the task at hand. While color information is undoubtedly part of what the network learns, this "black box" approach makes it impossible to isolate the specific contribution of color or to understand how the model is using it. The learned features are entangled, and the model's decisions are not directly interpretable in terms of specific color palettes or distributions. This situation reveals a curious disconnect in the field, particularly concerning the OmniArt dataset. As previously noted, OmniArt's rich metadata explicitly includes color codes and palettes for its artworks This data is pre-processed and ready for analysis. Yet, a review of the published research that uses OmniArt shows that this valuable, explicit color information remains largely unexploited. The primary research thrusts have focused on multi-task learning and classification using attributes like style, artist, and period, but not on large-scale inference driven by the provided color data. This reveals an inefficiency in the research landscape: the dataset that lacks explicit color metadata (WikiArt) has inspired a major effort to engineer those features from scratch, while the dataset that provides them pre-packaged (OmniArt) has not seen them widely used for dedicated color analysis. This strongly suggests that a study focused on leveraging the explicit color data within OmniArt is an overdue and logical next step.

6 Proposal

The body of literature on computational art analysis has matured significantly, establishing robust methodologies and foundational datasets. Despite these advances, this review reveals a significant, underexplored research territory of color, chronology, and geography.

Therefore, the present study is positioned to make a novel contribution by directly addressing this confluence of gaps. We propose to move beyond standard classification tasks by developing a data mining framework that uses color as a primary, explicit feature to predict the geographic origin and temporal period of artworks. This approach is innovative in three key aspects: (1) it elevates color from an implicit, entangled feature within a CNN to the primary driver of analysis; (2) it targets geography, a fundamental dimension of art historical context that has been largely overlooked in large-scale computational studies; (3) it leverages the rich but untapped color metadata of the **OmniArt dataset**, providing a new application for this powerful resource while remaining cognizant of the known biases that affect large art corpora. By bridging these identified gaps, this work aims to demonstrate the predictive power of color and open a new avenue for **spatio-temporal analysis** in computational art history.

The project can be found on https://github.com/vieira-giulia/art_data_mining .

7 Dataset

OmniArt's metadata - publicly available as a 1.23GB csv file [1] It is structured as seen in Table 1. In this project only the portion of Omniart referring to rows where the attribute artwork_type == "painting" (264247) was used, as color analysis is most interesting to paintings. Columns artist_first_name, and artist_last_name were also ignored due to their redundancy with artist_full_name. This resulted in a much smaller search space of 74.5MB. That was obtained by running the following command-line, and afterwards explored using Python:

```
csvgrep -c artwork_type -m painting omniart_v3_datadump.csv | csvcut -c \
artwork_name,artist_full_name,creation_year,century,school,color_pallete,\
dominant_color,palette_count > omniart-paintings-filtered.csv
```

Feature	Type	Description
`title`	Nominal	Title of the artwork.
`artist_full_name`	Nominal	Artist's full name.
`artist_first_name`	Nominal	Artist's first name.
`artist_last_name`	Nominal	Artist's last name.
`artwork_type`	Nominal	Art category the piece belongs to
`year`	Numeric (int)	Year of creation.
`century`	Numeric (int)	Century of creation.
`color_palette`	Nominal list	Extracted top 10 dominant RGB colors (hex).
`dominant_color`	Nominal	Most dominant RGB color (hex).
`color_count`	Numeric list (int)	Number of times the (10) dominant color appears.
`school`	Nominal	Art school the piece belongs to

Table 1: Attributes included in the OmniArt dataset.

[1] https://www.vistory-omniart.com

7.1 Dataset Exploration

To start this work, further dataset exploration was performed. Figure 10 shows that it is **sparsely missing**, which enables more robust and reproducible models, with complete-case analysis and statistical integrity. Nevertheless, as seen in Figure 11, even having a wide range of creation *years* — from -1000 to 2017 — the average *century* registered is the 20th century, which encompasses even the 25th percentile of art pieces. Therefore, **analysis performed on Omniart are inherently skewed toward more recent patterns**.

Similarly, most *schools* are cataloged as *modern*, which defines almost 70% of the dataset, followed by *unknown*, with nearly 18%. Combining both, only around 12% of the dataset is left with diverse *school* attributions that can be explored. Still, these limitations are not final. Due to OmniArt's high volume of data, there are enough art pieces cataloged to perform meaningful analysis, as stated by the project referenced in Section 5.

Moroever, *school* is not an intuitively labeled attribute, as it does not refer to artistic styles, but to regional attributions, as seen in Figure 12, in addition to *modern* and *unkown* denominations.

```
Empty values per column:
artwork_name        2
artist_full_name    0
creation_year       0
century             0
school              4
color_pallete       0
dominant_color      0
palette_count       0
dtype: int64
```

Figure 10: Empty cells

Descriptive statistics for numerical columns:

	creation_year	century
count	264247.000000	264247.000000
mean	1927.266542	20.037310
std	169.213789	1.820886
min	-1000.000000	-9.000000
25%	1913.000000	20.000000
50%	2010.000000	21.000000
75%	2013.000000	21.000000
max	2017.000000	21.000000

Figure 11: Dataset statistics

```
Proportion of paintings per school:
school
modern             0.696961
unknown            0.177042
italian            0.041333
dutch              0.014646
french             0.014509
                     ...
franco-flemish     0.000004
iran               0.000004
mexican            0.000004
bavaria,           0.000004
sri                0.000004
Name: proportion, Length: 80, dtype: float64
```

Figure 12: Schools prevalence

Nevertheless, color is described in detail in this dataset. For each painting, its 10 most common colors are registered in *color_palette*, with their corresponding occurrence registered in *color_count*. Yet, some precautions need to be taken when evaluating these color patterns. Even using the most advanced color recording techniques, a painting's color palette can vary. Additionally, colors registered as different in a fine-grained system such as RGB or Hex are essentially the same for the human eye, which leads to possible semantic false negatives, hampering similarity recognition and, therefore, classification based on color. That can be exemplified by Figure 13. In that subset of 400 paintings, 356 different colors were identified as *dominant_color*, when many of them are humanly-indistinguishable from each other and should be interpreted with similarity tending to 1, accounting as essentially the same color.

Figure 13: 400 paintings dominant color sample

8 Pre-processing

8.1 Dataset cleaning

8.1.1 *school* attribute cleaning

As presented described in Section 7, school names are not intuitively correlated to actual artistic schools or styles. Nevertheless, most of them hint towards the creation location of the art pieces. Therefore, cleaning the attribute *school* enables the possibility of data mining related to **geographic information**, which is one of this project's goals. To clean that column, a dictionary of nationality translations [2] that listed all 85 *schools* described in the dataset and associated them to a country or region name was used.

8.1.2 *decade* attribute creation

century is a long period of time to analyze modern artworks, which are marked by a faster-paced change of movements, and concurrent styles. Furthermore, specific *creation_year* are often imprecise. Therefore, to obtain granularity without restraint to specific years, the *decade* column was created.

8.1.3 *color_pallete* and *palette_count* attributes dimension definition: 10 fields

For some rows in this dataset, the number of colors cited in *color_pallete* and *palette_count* is irregular. In some rows these columns had more than 10 items, in others less; occasionally, the number of items in *color_pallete* and *palette_count* mismatched for the same row. For that reason, rows with dimensionality higher than 10 were reduced to 10, while rows with dimensionality lower than 10 were dropped. After data cleaning, the final dataset encompasses 264218 artworks, 29 less than the original 264247, and looks like Figure 14.

	artwork_name	artist_full_name	creation_year	century	school	color_pallete	dominant_color	palette_count	decade
264213	truth	circlekeeper	2008.0	21.0	modern	[#a86f61, #a09d79, #372b2e, #252125, #242d42, ...	#252125	[1312, 5526, 9605, 9971, 4030, 5524, 5562, 286...	2000
264214	yavanna queen of the earth giver of fruits-tif...	tiffany-illustration	2017.0	21.0	modern	[#ffffff, #feffff, #eef1df, #c0af95, #896962, ...	#ffffff	[15056, 255, 5087, 6846, 9838, 1577, 4027, 476...	2010
264215	unknown	kano furunobu	1700.0	18.0	Japan	[#b6ad97, #af9974, #a78d66, #a19371, #a08d67, ...	#7e4e48	[3431, 5740, 4247, 4971, 6362, 6885, 5328, 445...	1700
264216	unknown	wang yuanqi	1715.0	18.0	China	[#f8f3ed, #f4ede6, #f1e3d0, #edd9bb, #fbf8f2, ...	#e7d2b5	[6512, 6102, 860, 6846, 4826, 5036, 5117, 3297...	1710
264217	portrait of a gentleman	louis-gabriel blanchet	1725.0	18.0	France	[#a9977f, #39211a, #2d2b25, #251f1a, #1c1d1d, ...	#393227	[3673, 4253, 5392, 6641, 3191, 7046, 4680, 706...	1720

Figure 14: Clean dataset

8.2 Color representation model transfer: HEX to HSV

As explained in section 3, the color model chosen to represent colors in this project is **HSV**. To acquire that, the HEX values provided by OmniArt was translated into HSV using the function*rgb_to_hsv* from the *colorsys* library. The final dataframe adds the columns H, S and B (equivalent to V), that follow the same list formatting of previous *color_pallete*.

[2] nationality-country-dict.csv

8.3 Color clustering: dimensionality reduction

Humans can only reliably distinguish around one thousand distinct colors (Wang and Li, 2020). This perceptual threshold implies that finer subdivisions beyond this number do not contribute meaningful additional information for color analysis, as they exceed the resolution of human vision. Therefore, consolidating the 1197223 colors listed in OmniArt into approximately 1000 representative categories not only aligns with human visual capabilities but also simplifies computational processing, reducing complexity while preserving perceptually relevant distinctions. This approach ensures that the analysis remains both efficient and meaningful in terms of how colors are experienced and differentiated by observers.

8.3.1 Methodology

The features for the clustering algorithm used in this work are defined by a vector of tuples (H_sin, H_cos, S, B), where H_sin and H_cos are subdivisions of H, as hue is a circle. The algorithm also takes *palette_count* values as weights for each color's total occurrence, which helps the clustering algorithm focus on the most visually dominant or representative colors in the dataset. As 1000 colors was described in the literature as a reasonable universe for human-compatible color analysis, using **kmeans** with $\mathbf{K} = \mathbf{1000}$ was a straightforward choice.

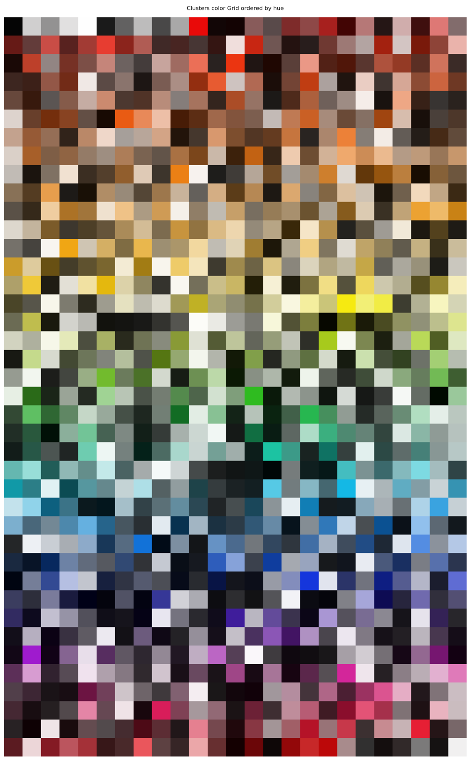

Figure 15: All 1000 color clusters ordered by hue

8.3.2 Results

The resulting clusters can be seen in Figure 15. The overall **hue**, **saturation** and **brightness** values were skewed in the original dataframe, with a prevalence for low hue, low saturation, and high brightness, as seen in Figures 16, 18 and 20, respectively. Nevertheless, after clustering, values were balanced, with more specimens of less dominant profiles, as seen in Figures 17, 19 and 21 respectively. Clustering that redistributes color samples more evenly helps ensure that underrepresented but perceptually important color ranges are captured, allowing for a more faithful representation of the full color space.

Moreover, Figures 17, 19 and 21 demonstrate that the color space was not sampled completely uniformly, the general shape of the hue and saturation functions resemble wider spectrum of their original forms. The high peak in light brightness characteristic of the original dataset was substituted by a softer one, with the creation of a more pronounced left tail, which denotes a less straightforward, but still perceivable resemblance.

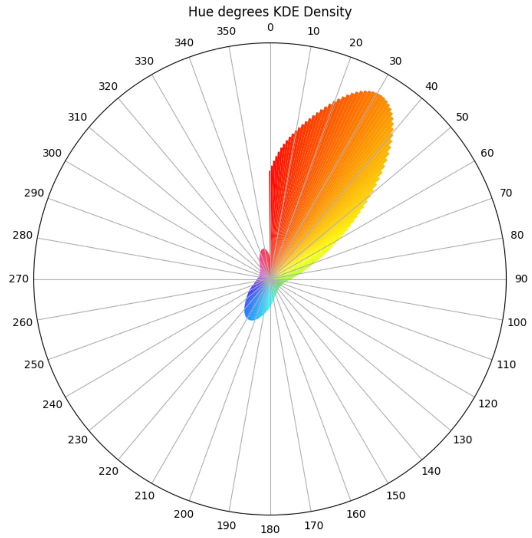

Figure 16: Original dataset's hue distribution

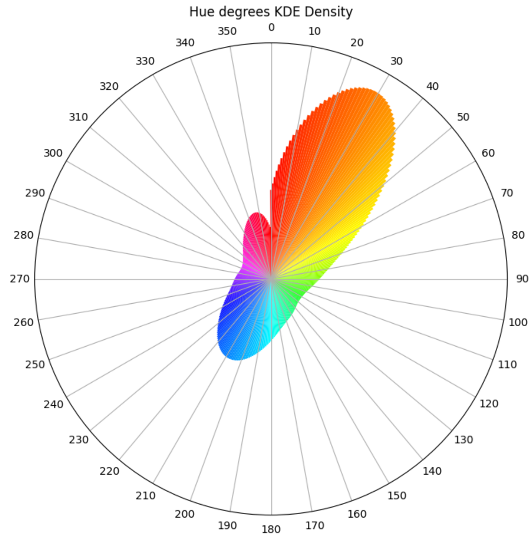

Figure 17: Clustered colors hue distribution

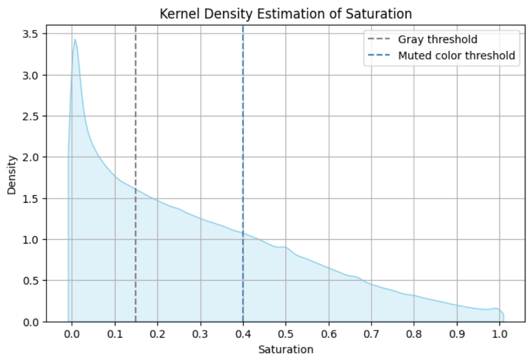

Figure 18: Original dataset's saturation distribution

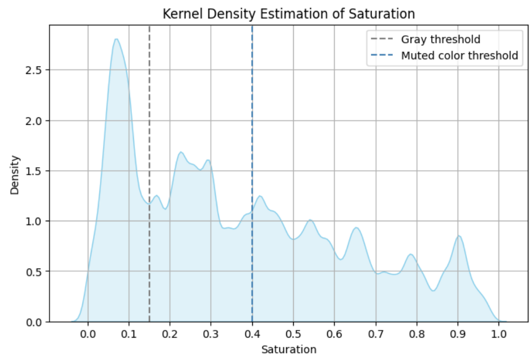

Figure 19: Clustered colors saturation distribution

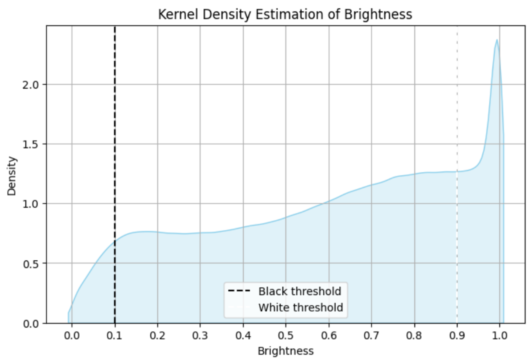

Figure 20: Original dataset's brightness (value) distribution

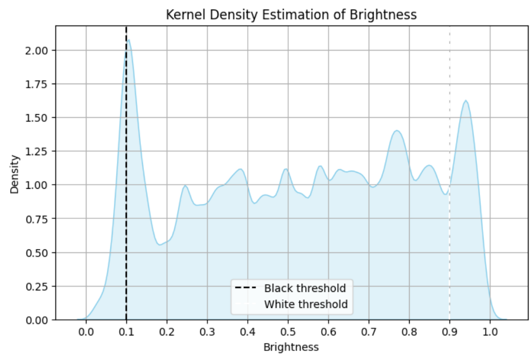

Figure 21: Clustered colors brightness (value) distribution

8.3.3 Quantified cluster evaluation

- **Silhouette Score:** The Silhouette Score is a clustering evaluation metric that quantifies how well each object lies within its cluster compared to other clusters. It ranges from -1 to 1: values near 1 indicate well-separated, dense clusters; values around 0 suggest overlapping or ambiguous clustering; and values below 0 imply misclassified samples. Typically, a higher score indicates better-defined clusters. However, in contexts like color clustering, unlike discrete, independent categories, a gradual transitions and inherent similarities are expected. For example, a light orange is naturally close to both a deep orange and a pale yellow. Therefore, some degree of cluster overlap is not only expected but necessary for an accurate representation of the color spectrum. That occurs especially when the subspace of colors is those prevalent in paintings, which are limited by material and stylistic constraints, as stated in section 4. The Silhouette Score of 0.212, while modest, is a positive value that reflects this reality. It suggests that while data points are, on average, closer to their own cluster than to neighboring ones, the boundaries are not sharply defined. This aligns perfectly with the continuous nature of color, where distinct color groups fade into one another.

- **Calinski-Harabasz Score:** The Calinski-Harabasz Score evaluates cluster validity based on the ratio of between-cluster dispersion to within-cluster compactness, with higher values suggesting clearer distinctions. In this context, the very high Calinski-Harabasz Score of 189,733.403 is particularly insightful. This score demonstrates that despite the expected overlap at the peripheries, the clusters are internally compact and their centroids are well-separated. It indicates that the clustering successfully identified distinct and dense color archetypes (e.g., a "core red," a "core blue," a "core brown").

- **Davies-Bouldin Index:** The Davies-Bouldin Index quantifies the average similarity between each cluster and its most similar one, where lower values indicate better-defined clusters. The Davies-Bouldin Index of 0.977 further supports the sillouette score interpretation. This value indicates a degree of similarity between each cluster and its nearest neighbor, reinforcing the idea of soft boundaries and a connected color space.

8.3.4 Conclusion

The color clustering process presented here achieved an effective dimensionality reduction of the original color space, condensing over one million unique color instances into a curated set of 1000 perceptually meaningful categories. This decision was guided by both human perceptual limits and computational practicality, ensuring that the resulting clusters remain interpretable and representative of actual visual experience. By leveraging a feature space tailored to the circular nature of hue and weighting colors by their frequency of occurrence, the clustering emphasized dominant chromatic elements while also preserving less frequent, yet perceptually significant, tones.

Post-clustering distributions showed a more balanced coverage of hue, saturation, and brightness ranges, correcting the biases found in the raw dataset and enabling fairer and richer analyses. The evaluation metrics further confirmed the effectiveness of this method: although color transitions are naturally continuous — leading to moderate Silhouette and Davies-Bouldin scores — the high Calinski-Harabasz score demonstrated strong internal cohesion and inter-cluster separation.

9 Artworks clustering

To analyze the color-based patterns in a large collection of 264218 artworks, clustering is a practical method for grouping artworks with similar color palettes. Given that each artwork is represented by 10 colors selected from a fixed vocabulary of 1000 possible colors— formed in Section 8.3 — the total combinatorial space of possible palettes is extremely high, far beyond what can be meaningfully interpreted or directly compared. Reducing this complexity through clustering makes the data manageable and reveals higher-level structure.

More importantly, clustering artworks by their color composition is valuable in itself, as it allows us to uncover patterns that align with artistic styles, schools, or historical periods. Color is a fundamental element of visual expression and often reflects the aesthetic conventions, cultural influences, and technological constraints of a particular time or movement. For instance, certain clusters may correspond to the muted tones of Baroque painting, the bright contrasts of Pop Art, or the earthy palette of early modernist works.

By identifying groups of artworks that share similar chromatic signatures, we can infer stylistic affinities without requiring explicit labels or metadata. This unsupervised approach enables the discovery of latent artistic patterns that may not be obvious through manual inspection or categorical analysis. Moreover, it supports comparative studies across regions, schools, and decades, offering a scalable way to explore the evolution of visual language across a massive corpus. In this way, color-based clustering not only simplifies the analytical task but also opens a window into the aesthetic structure of art history, providing insights that are both computationally tractable and art historically meaningful.

9.1 Methodology

For this clustering task, **k-means** was once again used. Each artwork in the dataset is represented by a 10-dimensional color vector, where each dimension corresponds to a color-cluster that approximates one of the original colors used in the artwork. These vectors, constructed from a fixed vocabulary of 1000 possible colors (as described in Section 8.3), serve as the features for clustering. While the theoretical space of possible color combinations is extremely large, in practice, artworks tend to share many common colors — such as natural skin tones, sky blues, and earth pigments — driven by stylistic traditions, representational goals, or material limitations. This overlap implies that the effective diversity of color palettes is much lower than the full combinatorial space, making clustering both tractable and meaningful.

Assuming no overlap between clusters, around 100 clusters would be required to span the range of distinct colors found in the dataset. Additionally, choosing $\mathbf{K} = \mathbf{100}$ provides a midpoint between the 134 possible decades and the 49 possible schools attributed to the artworks, offering a balanced level of granularity for subsequent analysis.

9.2 Results

The resulting clusters can be seen in Figure 22 and will be used in further sections to exemplify the results related to the concepts presented in section 4.1, that states that the relationships between colors in artworks help comprehend their characteristics.

The distribution of artworks per cluster can be seen in Figure 23. Given the total of 264218 artworks, the median of 2457 artworks per cluster, very close to the mean 2642, demonstrates a relatively **balanced distribution of artworks** across clusters.

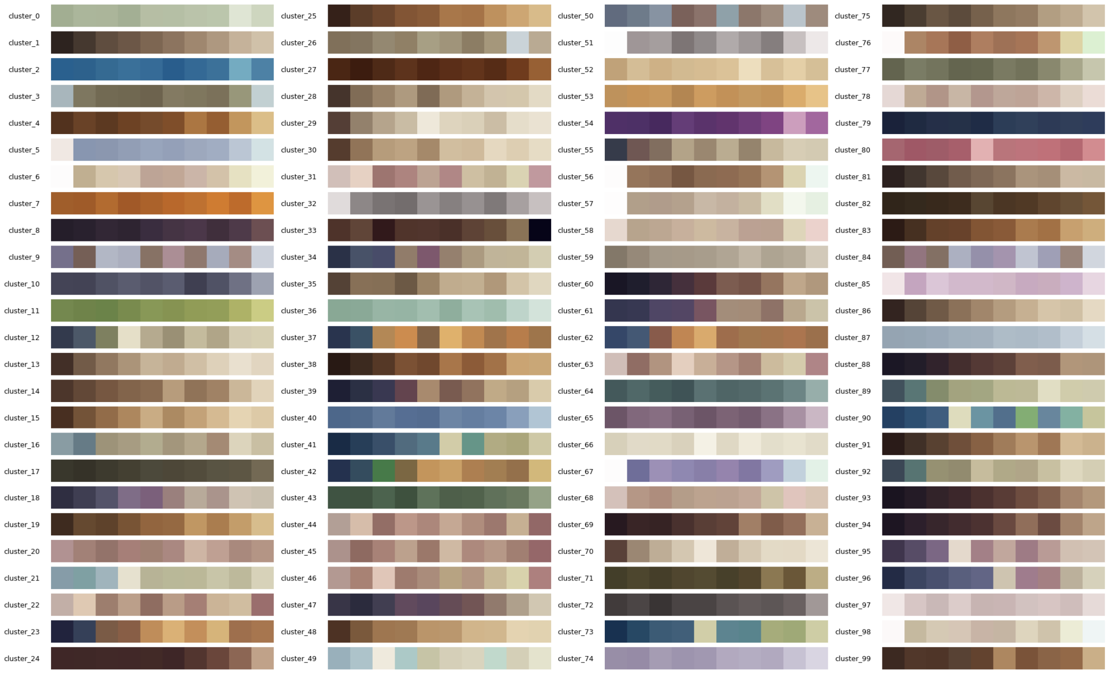

Figure 22: All 100 artwork clusters

This suggests that the clustering process did not produce highly imbalanced groupings, and that most clusters contain a comparable number of artworks, reinforcing the suitability of using K = 100 for this dataset. Likewise, this balance is beneficial for downstream pattern recognition tasks, as it prevents the dominance of a few large clusters and ensures that each cluster contributes comparably to the analysis, and ensures a larger minimum **support** for each pattern mined. When clusters are evenly populated, statistical comparisons between them — such as identifying characteristic features, measuring diversity, or detecting trends across schools or time periods — are more robust and less biased. It also improves the effectiveness of machine learning models trained on cluster-labeled data, as balanced class distributions reduce overfitting to majority classes and improve generalization. Thus, this balanced clustering enhances the interpretability and reliability of patterns extracted from the artwork dataset.

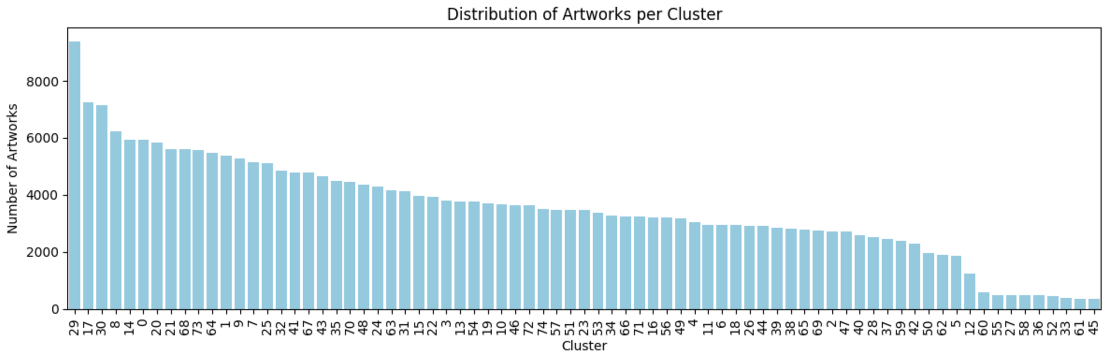

Figure 23: Distribution of artworks per cluster

9.2.1 Clusters color profile: Hue

Do the concepts of monochromatic, analogous, triadic and complementary hue profiles from 4.1 hold when clustering artworks by color?

Looking at **hue**: **Monochromatic** pieces cover less than 15 degrees of the color wheel; **Analogous** less than 60 degrees; **Triadic** cover intervals multiples of 120 degrees; and **Complimentary** cover at least 180 degrees, as seen in Figure 24. Analyzing Figures 22 and 25 (which demonstrates the hue coverage of each artwork cluster), guided by these four categories, one can count 21 monochromatic; 33 analogous; and 46 triadic/complimentary clusters (it is important to note that triadic and complimentary are difficult to differentiate, as when three triadic colors are present, the hue coverage resembles a complimentary hue coverage). Therefore, **this clustering exemplifies and demonstrates the hue-related artistic intents formalized in section 4.1**, as the paintings can be grouped by their restricted hue coverage.

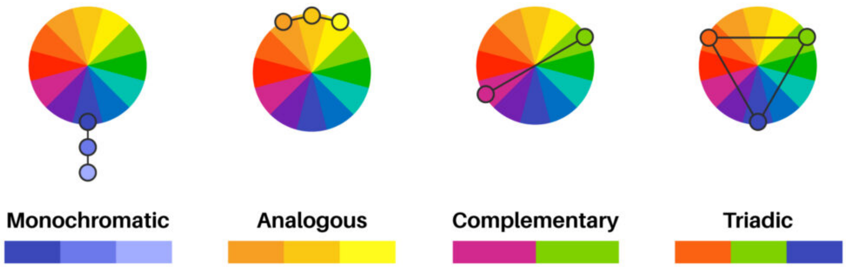

Figure 24: Hue coverage color profiles

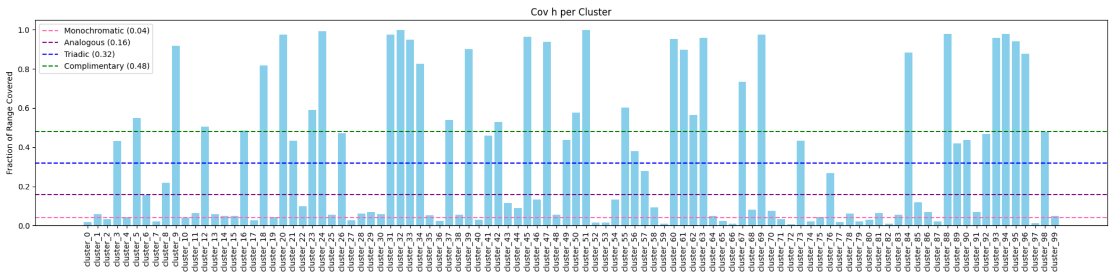

Figure 25: Clusters hue coverages

9.2.2 Clusters color profile: Saturation

Does the concept of harmonic saturation profile from 4.1 hold when clustering artworks by color?

Looking at **saturation**: a cluster is homogeneously saturated when its saturation coverage is below 30% of the saturation spectrum. As seen in Figure 26, 88 clusters are **homogeneously saturated**, while 12 are not, which denotes saturation coherence intra-cluster. This denotes that **this clustering exemplifies and demonstrates the saturation-related artistic intent formalized in section 4.1**, where the hypothesis is that saturation harmony is preferred.

Furthermore, a color is interpreted as **gray** when its saturation value is below 15%, and **muted** when its saturation value is below 40%.Figure 27 demonstrates that 76 groups have average saturation below muted, where 62 of them also have maximum saturation below that threshold, which denotes an overall prevalence of muted tones in the clusters' color palettes -

which is perceivable in the overall saturation homogeneity if 22 -. 9 of the clusters have minimum saturation above the muted threshold, denoting highly contrasted palettes. Furthermore, 17 clusters have average saturation below gray, where 5 of them also have maximum saturation below that threshold, denoting nearly exclusively gray-resemblant palettes.

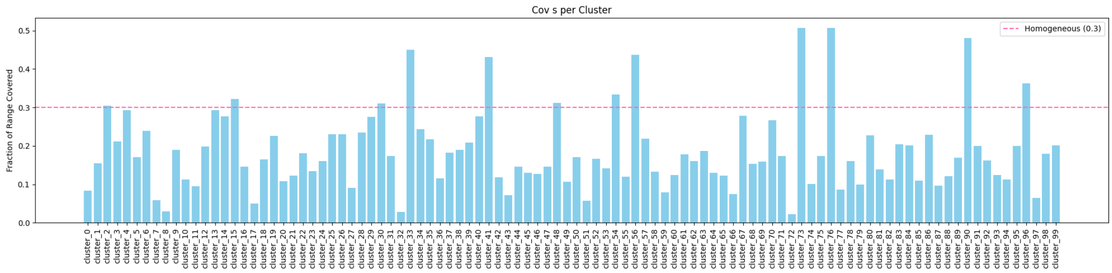

Figure 26: Saturation coverage color profiles

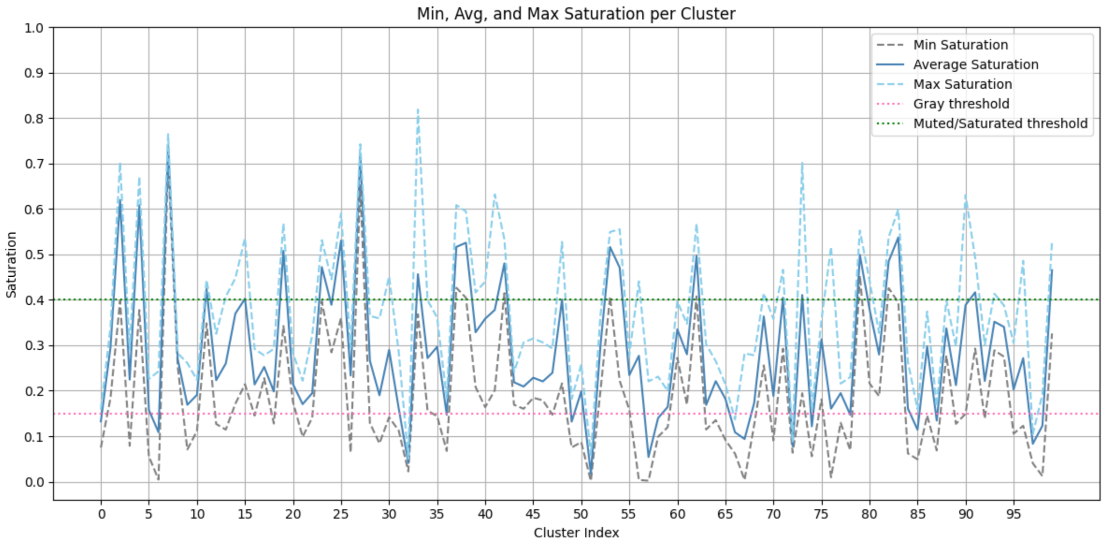

Figure 27: Clusters saturation values

9.2.3 Clusters color profile: Brightness

Does the concept of contrasting brightness profile from 4.1 hold when clustering artworks by color?

Looking at **brightness**: a cluster is homogeneously bright when its brightness coverage is below 30% of the possible saturation spectrum. As seen in Figure 28, 31 clusters are homogeneously bright, while 69 are not, which denotes a prevalence of **brightness contrast** intra-cluster. That can be perceived in Figure 22, where most palettes contain a combination of light and dark colors. This denotes that **this clustering tends to the brightness-related artistic intent formalized in section 4.1**, where the hypothesis is that artists use highly contrasting brightness, possibly to denote emotion.

Furthermore, a color is interpreted as a shade of **white** when its brightness value is over 90%, and **light** when its brightness value is over 80%. Similarly, a color is interpreted as a shade of **black** when its brightness value is below 10%, and **dark** when its brightness value is below 20%. Figure 27 demonstrates that no groups have average or maximum brightness under black or dark, only one of them have minimum brightness under black - which can be seen in Figure 22 as *cluster_ 33* - and 14 of them under dark). This denotes a small, but significant group of clusters that contain dark colors, one of them containing black itself, but no cluster that is exclusively dark. Similarly, regarding white and light, no clusters have minimum brightness above white, and only one has it above light, denoting only one exclusively light color palette.

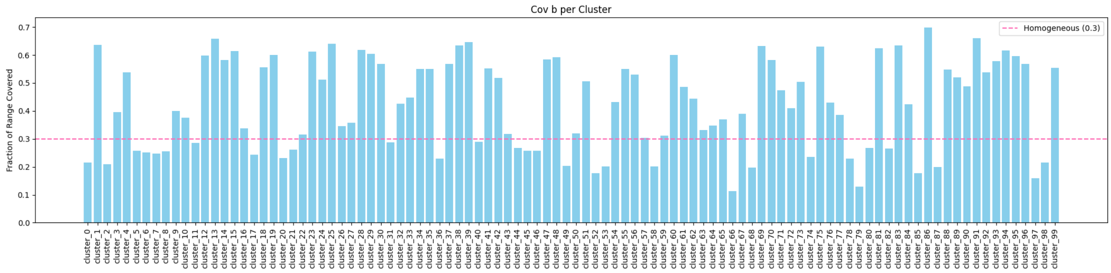

Figure 28: Brightness coverage color profiles

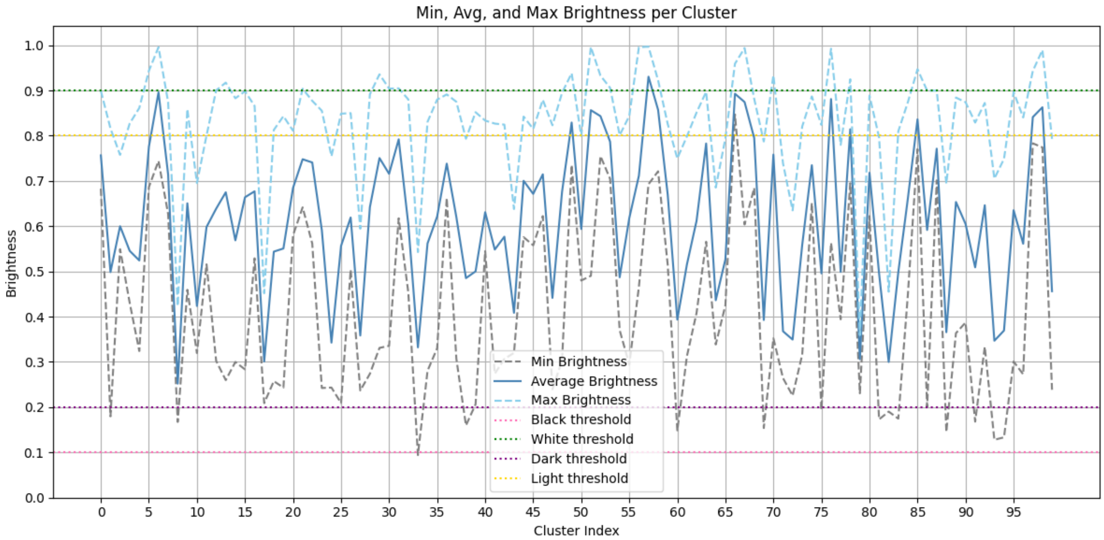

Figure 29: Clusters brightness values

9.3 Clusters relationships to attributes: insights for pattern recognition and classification

Similar to the dataset's skewness towards **21th century** *decades* and **modern** or **unknown** *schools*, artworks clusters are skewed toward those groups, as seen in Figures 30 and 31. This can limit the interpretability and representativeness of the results when performing itemset mining, reinforcing the necessity of using metrics that balance results by item popularity, and performing mining with small minimum support levels; and when performing classification, also reinforcing the necessity of balancing class prevalence, as when clusters are dominated by overrepresented categories they may fail to capture meaningful variation across underrepresented ones.

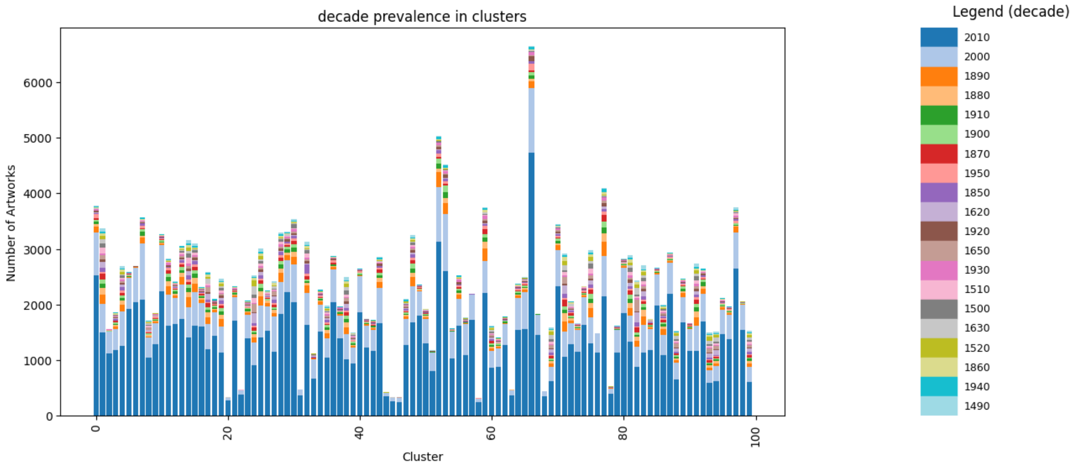

Figure 30: Decades prevalence in artworks clusters

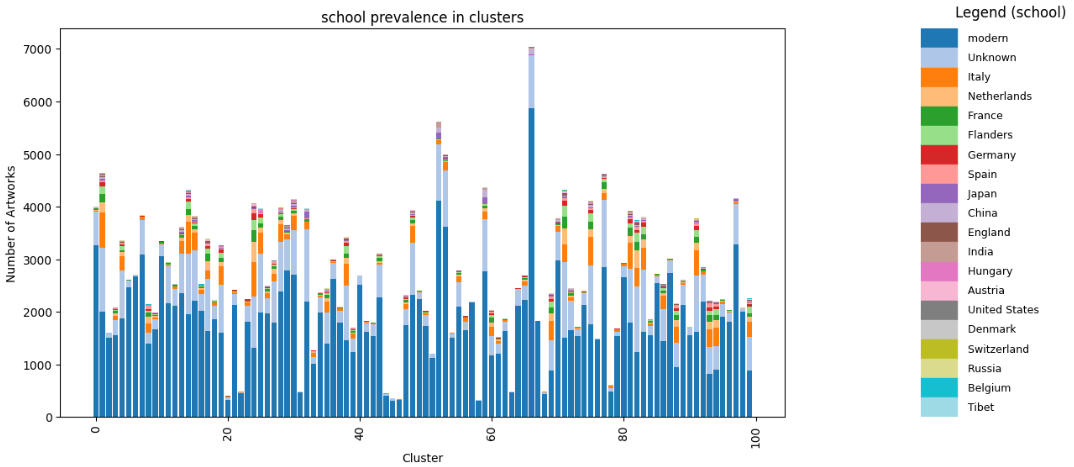

Figure 31: Schools prevalence in artworks clusters

The **Shannon Entropy**,

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i$$

, defines that the maximum entropy for a 134 *decades* universe is 7.07, while it is 5.61 for a 49 *schools*. According to Figures 32 and 33, the maximum values achieved are 3.5 and 1.75, respectively, which demonstrate a good level of homogeneity in the clusters, especially regarding the *school* attribute. Therefore, despite the necessity of working with low support when performing pattern recognition, that can be counteracted by usage of, for example, higher lift values when filtering patterns, as clusters with lower entropy indicate stronger association with specific *decades* or *Schools*.

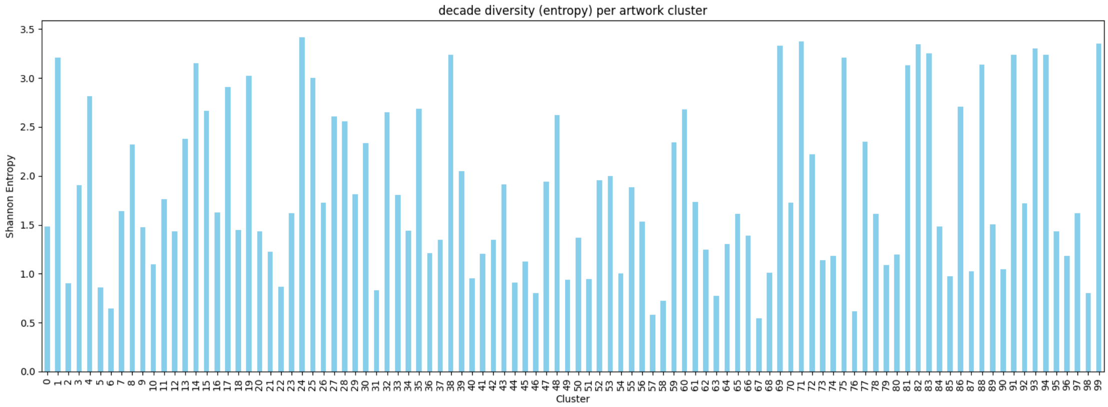

Figure 32: Decades diversity in artworks clusters

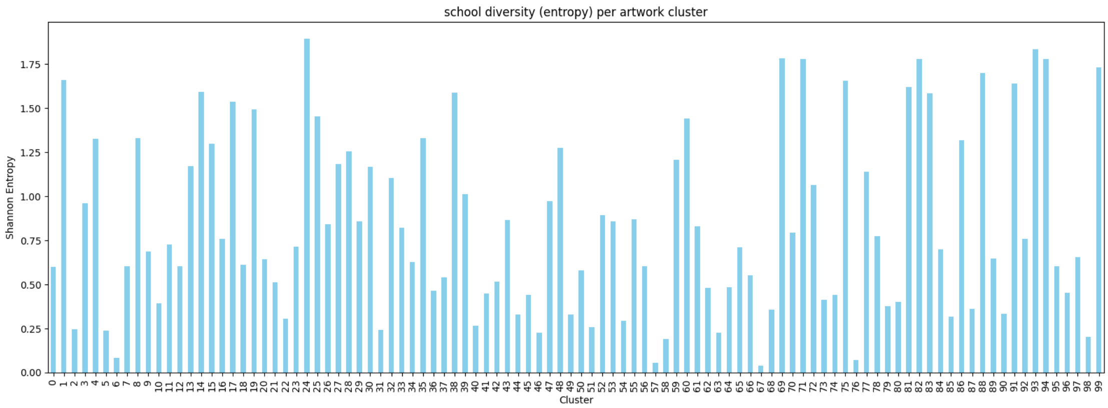

Figure 33: Schools diversity in artworks clusters

9.4 Quantified cluster evaluation

9.4.1 Sillouette Score

The **Silhouette Score** is a clustering evaluation metric that quantifies how well each object lies within its cluster compared to other clusters. It ranges from -1 to 1: values near 1 indicate well-separated, dense clusters; values around 0 suggest overlapping or ambiguous clustering; and values below 0 imply misclassified samples. Typically, a higher score indicates better-defined clusters. However, in contexts like color usage in artworks, where some hues are naturally more prevalent and clusters are expected to overlap, a low score like 0.018 is not necessarily a negative outcome. It reflects that many colors are shared across clusters, which aligns with artistic practices - certain palettes recur across styles, periods, and artists. In this case, the low score confirms the fluid and continuous nature of color use, rather than indicating poor clustering.

9.4.2 Calinski-Harabasz Score

he Calinski-Harabasz Score evaluates cluster validity based on the ratio of between-cluster dispersion to within-cluster compactness, with higher values suggesting clearer distinctions. A score of 2377.760 is relatively high, meaning that, despite the expected overlap in color distribution, the clusters still maintain internally consistent structure and meaningful separation in terms of dominant hues or usage patterns. In a dataset where color transitions are often gradual and shared, this suggests the clustering still captures salient groupings or thematic tendencies in color usage across artworks.

9.4.3 Davies-Bouldin Index

The Davies-Bouldin Index quantifies the average similarity between each cluster and its most similar one, where lower values indicate better-defined clusters. A value of 2.766 is relatively high in strict clustering contexts, but in this artistic setting, it reflects the inherent continuity and shared nature of color across artworks. Clusters that are somewhat close and similar make sense here, as certain color tones (e.g., earth tones) are commonly reused across many artworks and periods. Thus, rather than suggesting poor clustering, this DBI value aligns with the real-world expectation of soft cluster boundaries in visual art data.

9.5 Conclusion

The analysis demonstrated that the computationally derived clusters empirically validate established principles of color theory and artistic intent. The clusters naturally segregated artworks based on their hue profiles, successfully identifying groups corresponding to monochromatic, analogous, and complementary/triadic color schemes. This alignment confirms that the unsupervised clustering captured meaningful artistic choices regarding hue harmony. Similarly, the investigation of saturation revealed a strong tendency towards intra-cluster homogeneity, with a majority of clusters (88%) exhibiting harmonic saturation and a prevalence of muted tones. In contrast, the analysis of brightness pointed towards a preference for high contrast within clusters, supporting the hypothesis that artists often leverage variations in lightness and darkness for expressive impact.

Furthermore, the relationship between the derived clusters and the artworks' metadata provided valuable insights. Despite a dataset skew towards modern eras and specific schools, the low Shannon Entropy values for both decade (H max = 3.5) and school (H max = 1.75) distributions within clusters indicate a significant degree of stylistic and temporal homogeneity. **This suggests that the color-based clusters are not arbitrary but correspond to genuine art-historical groupings.**

The quantitative evaluation metrics offered a nuanced validation of the clustering structure. While a low Silhouette Score (0.018) and a moderate Davies-Bouldin Index (2.766) reflect the inherent fluidity and overlapping nature of color palettes in the history of art, the high Calinski-Harabasz Score (2377.760) confirmed that the clusters are nevertheless dense and well-separated.

In summary, this work successfully transformed a massive, unstructured dataset of artworks into an interpretable framework of color-based archetypes. The resulting clusters not only align with formal art theory but also exhibit strong correlations with historical and stylistic metadata, providing a powerful lens for exploring the evolution of visual language. This research validates color-based clustering as a potent methodology for quantitative art history and establishes a solid groundwork for future investigations into the intricate relationships between color, style, and meaning.

10 Itemset mining

Having successfully grouped the vast collection of artworks into 100 distinct, color-based clusters, the next logical step is to leverage these groupings to uncover more granular and explicit relationships within the data. The clusters, as established in Section 9, serve as powerful, data-driven proxies for archetypal palettes that exhibit homogeneity in terms of hue, saturation, and brightness, as well as significant associations with specific art-historical periods and schools. We now move from describing these clusters to actively using them as foundational elements in a pattern discovery process.

10.1 Methodology

To systematically uncover the associative rules between artwork clusters and their categorical attributes, we employed frequent itemset mining. The entire process was implemented in *Python*, utilizing the **fpgrowth** algorithm available in the *mlxtend* library.

The FP-Growth algorithm is a highly efficient method for discovering frequent itemsets without candidate generation, which is a bottleneck in alternative approaches like Apriori. It achieves this by first compressing the database into a compact, tree-based data structure known as an FP-tree, which stores the essential itemset co-occurrence information. It then mines this structure directly to extract frequent itemsets, making it particularly well-suited for large or sparse datasets like the one under analysis.

For the purpose of this analysis, each of the 264218 artworks was conceptualized as a single transaction. The items within each transaction were derived from the artwork's attributes. To investigate relationships separately, two distinct transactional datasets were prepared:

- **Cluster-Decade Analysis:** Each transaction contained two items: the artwork's assigned cluster and its decade.
- **Cluster-School Analysis:** Each transaction contained two items: the artwork's assigned cluster and its school.

In both cases, **Support** was set to a minimum threshold of **0.0005**. With 264218 artworks, this threshold corresponds to an absolute minimum count of approximately 132 artworks. This deliberately low support value was chosen for two critical reasons: (1) the dataset is inherently sparse, with thousands of potential item combinations; (2) the data is severely skewed towards 21st-century artworks (in the decade analysis) and the "Modern" or "Unknown" categories (in the school analysis). A higher support threshold would exclusively capture patterns related to these dominant groups, rendering invisible the potentially significant associations within under-represented, yet art-historically important, periods and schools. By setting a low support, we ensure the algorithm has the sensitivity to detect these rarer signals.

Lift was used as the primary metric for filtering the final association rules, with a minimum threshold of **5**. By setting a high threshold of 5, we impose a strict criterion that filters for only very strong and non-spurious correlations. This high lift value serves as a crucial counterbalance to the low support threshold; it ensures that while we may be looking at infrequent patterns, the associations we retain are highly predictive and represent a genuine relationship between a color palette and an attribute, rather than a coincidental co-occurrence.

10.2 Cluster-Decade patterns

Are artworks color patterns associated to their creation period?

Itemset mining with such low support resulted in **1086 itemset patterns**, which were then filtered into **66 association rules**. These associations can be seen in Figure **??**, where, in the grid, an intersection between a **decade** column and a **cluster** row denotes an association between the two. 16 decades are related to only one cluster, while 7 are related to 2 clusters, and only 1 is related to 3 clusters. On the other hand, 4 clusters are associated with only 1 decade, while 2 are associated to 2 decades, 1 is associated to 3 decades, 1 associated to 9 decades and 1 to 11 decades. Some relationships are 1 for 1: 1420 and cluster 25; 1540 and cluster 88. Furthermore, one interesting pattern in these associations is that, when a cluster relates to more than one decade, these decades are generally consecutive, which denote the intuitive persistence of some artistic trends through time, and more long-lasting art movements, especially as in both cases this continuity happened before the Industrial Revolution, which is characterized by a more fast-paced environment, even in the arts. Therefore, the associations found **show that time is a defining axis for artistic trends.**

When analyzing patterns between decades and clusters, the most interesting associations are cluster -> decade, as the hyphothesis is that color-based artwork clusters can help filter out the decade period when paintings were made. Furthermore, it is alluring to look at the color palettes that define a cluster, and try to compare it to real-life examples. Figure 35 demonstrates those color palettes. Two things are to be highlighted here: (1) a prevalence of earth tones; (2) the use of a palette of shades of purple representing 1600, both of which **hint towards the historical use of earth materials to fabricate paint, and the pigments trade during the Age of Discovery, proposed in Section 4.2**. The appearance of purples in this period is particularly noteworthy. Historically, purple was among the most expensive and symbolically loaded colors, due to the laborious extraction processes from sources like Murex shells or lichen. Its rarity made it a color of royalty and exclusivity throughout antiquity and the Middle Ages. However, as trade routes expanded during and after the 15th century, particularly following Vasco da Gama's opening of the sea route to India, a wider range of dye materials—purples included—became available in Europe (Dean, 2010). The emergence of purple hues in artwork palettes after 1600 thus reflects not only evolving aesthetic preferences, but also the broader impact of global trade networks on material culture, including the availability of exotic pigments once restricted to the elite.

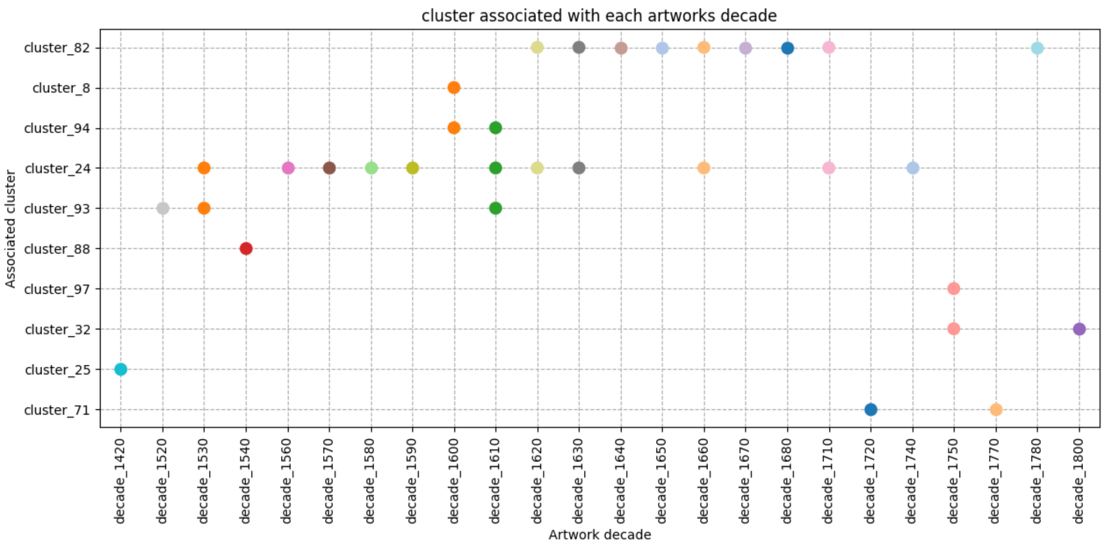

Figure 34: Decades-clusters association strip-plot

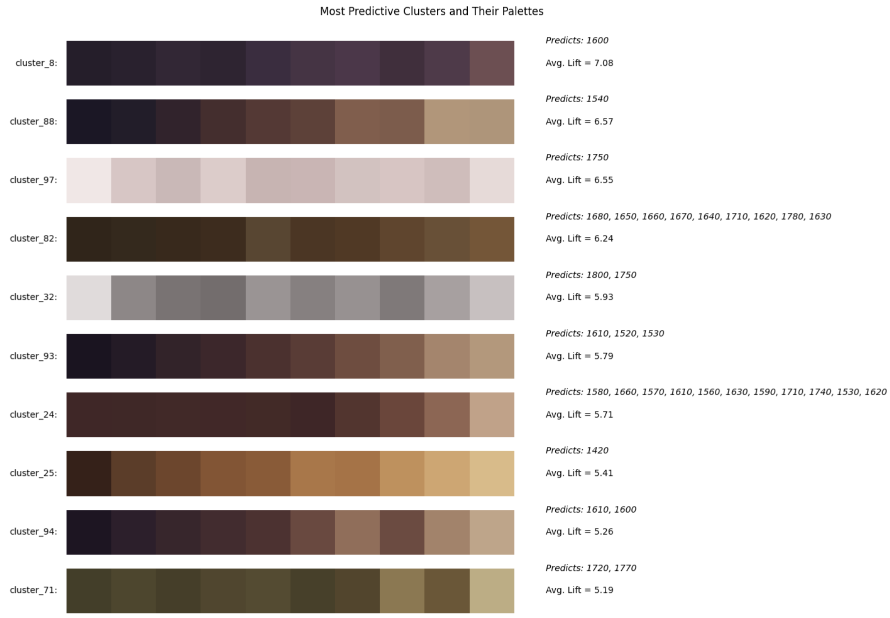

Figure 35: Cluster -> decade associations with color palette and lift value

10.2.1 Quantified association rules evaluation

1. **Support:** indicates the frequency or prevalence of an itemset in the dataset. The support values for the generated rules are low, with a **mean of 0.000307**, a **minimum of 0.000102**, and a **maximum of 0.000799**. This low support is a direct and intended consequence of the mining methodology, which set a low support threshold (0.0005) to uncover patterns in a sparse and skewed dataset.

2. **Confidence:** measures the reliability or predictive power of a rule. It is the conditional probability of finding the consequent (C) given that the transaction already contains the antecedent (A). Confidence scores range from a **minimum of 0.0066 (0.66%)** to a **maximum of 0.117 (11.7%)**, with a **mean of 0.051 (5.1%)**. On their own, these confidence values may appear modest. For instance, a mean confidence of 5.1% implies that for a given palette, there is a 5.1% chance it belongs to the associated decade. However, confidence can be misleading when the consequent is rare. The true strength of the rule is only revealed when comparing this confidence to the baseline probability of the consequent, which is the function of the Lift metric.

3. **Lift:** measures how many times more likely the antecedent and consequent are to co-occur than if they were statistically independent. A lift value greater than 1 indicates a positive correlation. The lift values are consistently and impressively high. The **minimum lift is 5.035**, the **mean is 5.899**, and the **maximum reaches 8.228**. **This is the most compelling evidence for the quality of the discovered rules.** A mean lift of approximately 5.9 signifies that the co-occurrence of the antecedent palette and consequent decade is, on average, nearly **six times more likely** than would be expected by random chance. The fact that the minimum lift is above 5 (the threshold set in the methodology) confirms that every single rule is highly significant. These are not coincidental findings; they represent strong, genuine associations.

4. **Leverage:** measures the difference between the observed frequency of an itemset and the frequency that would be expected if the items were independent. Positive leverage indicates a positive correlation. All leverage values are positive, ranging from **0.000083 to 0.000690**, with a **mean of 0.000254**. While the absolute values are small (a direct consequence of low support), their consistently positive nature confirms the positive association indicated by Lift. Leverage quantifies in absolute terms the number of transactions this rule adds above what would be expected from random co-occurrence.

5. **Conviction:** measures the degree of implication of a rule. It can be interpreted as the frequency with which the rule makes an incorrect prediction. A high conviction value means the consequent is highly dependent on the antecedent. All conviction values are greater than 1, with a **mean of 1.046** and a **maximum of 1.117**. A conviction value greater than 1 indicates a positive correlation. The results consistently show that the antecedent and consequent are dependent, further strengthening the validity of the discovered rules.

6. **Zhang's Metric:** is a robust, symmetric measure of association that ranges from -1 to +1. A value of +1 indicates a perfect positive association, -1 indicates a perfect negative association, and 0 indicates independence. **This metric provides another strong validation of rule quality.** The values are all high and positive, ranging from a **minimum of 0.804** to a **maximum of 0.891**, with a **mean of 0.835**. These scores are exceptionally high, indicating a very strong positive association for all discovered rules. A mean score of 0.835, close to the maximum of +1, suggests that the relationship between the specific palettes and decades is not just statistically significant but also very strong and reliable.

10.3 Cluster-School patterns

Are artworks color patterns associated to their creation's geographical location? Itemset mining with such low support resulted in **570 itemset patterns**, which were then filtered into **25 association rules**. These associations can be seen in Figure 36, where, in the grid, an intersection between a **school** column and a **cluster** row denotes an association between the two. 2 schools are related to only one cluster, and 2 related to 2 clusters, while only 1 is related to 3 clusters, and 1 is related to 4. On the other hand, 9 clusters are associated to only 1 school, while 2 are associated to 2 schools. Two relationships are 1 for 1: Austria and cluster 24; China and cluster 59. Furthermore, one interesting pattern in these associations is that, when a cluster is related to more than one school, these locations are geographically close and culturally intertwined: China and Japan share cluster 59; while Austria and the Netherlands share cluster 24. Therefore, the associations found **show that geographic location is a defining axis for artistic trends.**

When analyzing patterns between schools and clusters, the most interesting associations are cluster -> school, as the hyphothesis is that color-based artwork clusters can help filter out the geographic location where paintings were made. Furthermore, it is alluring to look at the color palettes that define a cluster, and try to compare it to real-life examples. Figure 37 demonstrates those color palettes. Two things are to be highlighted here: (1) a prevalence of earth tones; (2) the use of palettes containing shades of purple representing Spain, both of which **hint towards the geographically universal experience of using earth materials to fabricate paint, and enhanced paint access by globally trading nations as proposed in Section 4.2**. The appearance of purples in Spain of all places is particularly noteworthy. As explained in the previous section, purple is among the most expensive and symbolically loaded colors, therefore, if a country would have wider access to it, it should be one of the great naval traders.

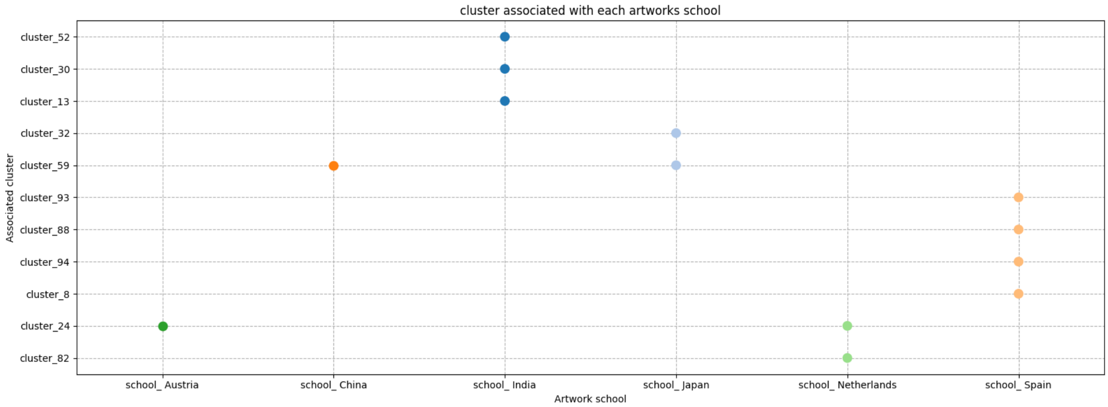

Figure 36: School-clusters association strip-plot

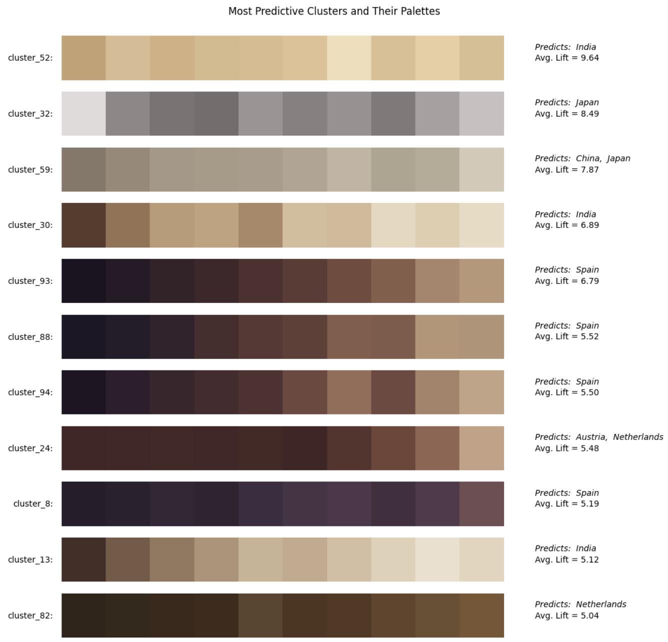

Figure 37: Cluster -> school associations with color palette and lift value

10.3.1 Quantified association rules evaluation

1. **Support:** indicates the frequency or prevalence of an itemset in the dataset. The support values for the generated rules are low, with a **mean of 0.000463**, a **minimum of 0.000102**, and a **maximum of 0.001450**. This low support is a direct and intended consequence of the mining methodology, which set a low support threshold to uncover patterns in a sparse and skewed dataset.

2. **Confidence:** measures the reliability or predictive power of a rule. It is the conditional probability of finding the consequent (C) given that the transaction already contains the antecedent (A). Confidence scores range from a **minimum of 0.0066 (0.66%)** to a **maximum of 0.2053 (20.53%)**, with a **mean of 0.0632 (6.32%)**. On their own, these

confidence values may appear modest. For instance, a mean confidence of 6.32% implies that for a given antecedent, there is a 6.32% chance of finding the consequent. However, confidence can be misleading when the consequent is rare. The true strength of the rule is only revealed when comparing this confidence to the baseline probability of the consequent, which is the function of the Lift metric.

3. **Lift:** measures how many times more likely the antecedent and consequent are to co-occur than if they were statistically independent. A lift value greater than 1 indicates a positive correlation. The lift values are consistently and impressively high. The **minimum lift is 5.042**, the **mean is 6.528**, and the **maximum reaches 9.639**. **This is the most compelling evidence for the quality of the discovered rules.** A mean lift of approximately 6.5 signifies that the co-occurrence of the antecedent and consequent is, on average, nearly **six and a half times more likely** than would be expected by random chance. The fact that the minimum lift is above 5 (the threshold set in the methodology) confirms that every single rule is highly significant. These are not coincidental findings; they represent strong, genuine associations.

4. **Leverage:** measures the difference between the observed frequency of an itemset and the frequency that would be expected if the items were independent. Positive leverage indicates a positive correlation. All leverage values are positive, ranging from **0.000085 to 0.001168**, with a **mean of 0.000386**. While the absolute values are small (a direct consequence of low support), their consistently positive nature confirms the positive association indicated by Lift. Leverage quantifies in absolute terms the number of transactions this rule adds above what would be expected from random co-occurrence.

5. **Conviction:** measures the degree of implication of a rule. It can be interpreted as the frequency with which the rule makes an incorrect prediction. A high conviction value means the consequent is highly dependent on the antecedent. All conviction values are greater than 1, with a **mean of 1.060** and a **maximum of 1.232**. A conviction value greater than 1 indicates a positive correlation. The results consistently show that the antecedent and consequent are dependent, further strengthening the validity of the discovered rules.

6. **Zhang's Metric:** is a robust, symmetric measure of association that ranges from -1 to +1. A value of +1 indicates a perfect positive association, -1 indicates a perfect negative association, and 0 indicates independence. **This metric provides another strong validation of rule quality.** The values are all high and positive, ranging from a **minimum of 0.806** to a **maximum of 0.916**, with a **mean of 0.848**. These scores are exceptionally high, indicating a very strong positive association for all discovered rules. A mean score of 0.848, close to the maximum of +1, suggests that the relationship between the antecedents and consequents is not just statistically significant but also very strong and reliable.

10.4 Conclusion

This section **successfully transitioned from descriptive clustering to quantitative pattern discovery**, demonstrating that strong, explicit relationships exist between color palettes and an artwork's spatio-temporal context. By employing a carefully calibrated itemset mining strategy—using a low support threshold to capture rare signals and a high lift threshold to ensure their significance—we successfully navigated the dataset's inherent sparsity and skew.

The analysis yielded dozens of statistically robust association rules for both creation decade and artistic school. These data-driven discoveries are not mere statistical artifacts; they resonate powerfully with established art-historical narratives. The identified palettes align with known material constraints and historical developments, such as the universal prevalence of earth tones and the emergence of exotic purples in European palettes corresponding with the expansion of global trade routes. Furthermore, the association of similar palettes with geographically and culturally linked schools provides empirical validation for theories of regional influence.

The validity of these findings is unequivocally supported by the quantitative metrics. The impressively high mean lift values—indicating that these palette-attribute pairings occur up to 6.5 times more frequently than by random chance—and the consistently high Zhang's scores confirm the strength and reliability of every discovered rule. Ultimately, this work validates frequent itemset mining as a potent methodology for computational art history, providing firm empirical evidence that specific color palettes can serve as powerful and reliable markers of an artwork's origin in time and space.

11 Classification

The classification of artworks has long been a cornerstone of their scientific and scholarly analysis, a tradition that in the contemporary era is overwhelmingly dominated by computational methods. Among these, multi-layered Convolutional Neural Networks (CNNs) have become the preeminent tool, as stated in Section 5. Shifting focus from these intricate deep learning architectures, this section details a more explainable classification methodology that instead leverages the foundational element of color.

11.1 Methodology

This study employed a **supervised machine learning** approach to classify artworks based on two distinct criteria: the decade of creation and the artistic school of origin. To accomplish this, two independent **Random Forest classifiers** were developed using the *RandomForestClassifier* from the **scikit-learn** (sklearn) *Python* library. Both models were trained on a **hybrid feature set** that combined quantitative visual information, derived from the artwork's **cluster color palette**, with its associated **contextual metadata**. The Random Forest algorithm was selected for its robustness, its strong performance on tabular data, and its ability to handle a mix of feature types while being relatively resistant to overfitting. Furthermore, in contrast to the more "black box" nature of deep learning models like CNNs, the Random Forest algorithm offers greater interpretability, allowing for an analysis of which features — be they color or metadata — are most influential in the classification decisions.

11.2 Decade Classifier

Can an explainable, color focused, classifier be successfully able to classify artworks creation decade?

The Decade Classifier finished training with an overall accuracy of 52.45%. However, this result was entirely misleading, as the model had collapsed into only predicting the majority class *('2010')* due to the dataset being extremely skewed towards 21st-century artworks. To properly evaluate the classifier's potential, a second model was trained using the same methodology but on a filtered dataset that excluded all data from the 2000s and 2010s. This approach serves as an alternative and benchmark for evaluating how the classifier performs on a more challenging and balanced, albeit smaller, dataset. This second model achieved a much lower accuracy of 9.10%, but its results provide a far more realistic assessment of the task's difficulty. Unlike the first model, this classifier did not collapse into a single prediction. Instead, the results show a more diverse, though often incorrect, range of predictions across multiple decades, with classes like '1750', '1870', and '1890' showing non-zero precision and recall.

This stark drop in accuracy, coupled with the increase in predictive diversity, demonstrates that while the features are not strong enough for high-accuracy classification of all decade classes, the model is at least attempting to find real patterns in the absence of an overwhelming majority class. The new misclassifications (e.g., mistaking 1880s art for 1890s art) are more logical than the previous model's blanket predictions of '2010'. This experiment strongly suggests that the extreme skewness towards contemporary digital art in many large-scale art datasets could be a significant confounding factor that may have influenced other works in the field. Studies that fail to account for such imbalance might report inflated accuracy metrics that do not reflect a model's true ability to classify across historical periods, but rather its ability to identify the dataset's majority class.

An analysis of the feature importances for the decade classifiers presents a seeming paradox with results from itemset mining. In both models, **the visual data derived from the color clusters was consistently ranked as more relevant than contextual metadata**. This appears to conflict with the models' poor predictive performance, especially given that a separate itemset mining analysis identified high-confidence correlations between specific features and classes. The resolution lies in the crucial distinction between feature correlation and predictive power. While certain colors are statistically associated with certain decades, these patterns are not discriminative enough for accurate classification. This is deeply rooted in the nature of art history itself: some colors, like earth tones and celestial blues, are permanently common across centuries, while other chromatic trends may be popular for a generation and then fade. This complex interplay of enduring palettes and fleeting trends creates an incredible amount of noise for a classifier. The model may correctly associate a specific ochre with the 17th century but finds that same ochre is also present in countless other periods, making it a poor feature for **discrimination**.

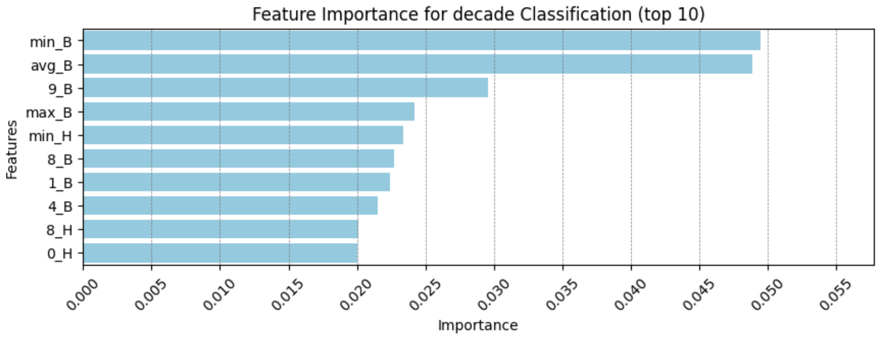

Figure 38: Decades feature importance

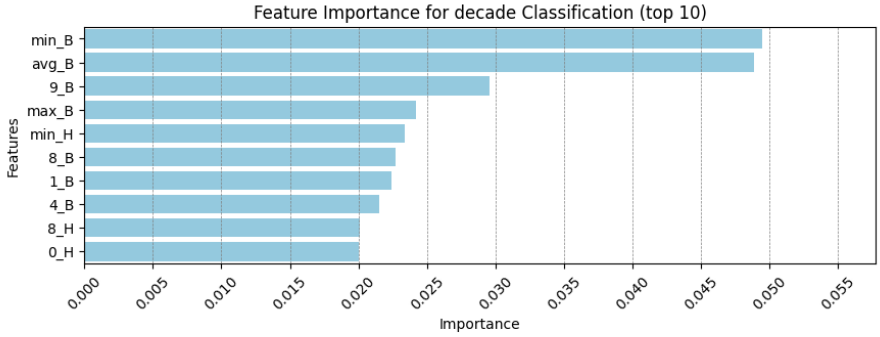

Figure 39: Decades (excluding 21th century) feature importance

11.2.1 Quantifiable classifier evaluation

A quantitative analysis of the performance metrics reinforces the qualitative conclusions. For the initial Decade Classifier, the 52.45% accuracy is deceptive, as the weighted F1-score of 0.36 and macro-average **F1-score** of 0.01 expose the model's failure. An F1-score of zero for every decade except '2010' confirms a complete collapse into predicting the majority class. The inability to calculate an **AUC score** for the '2010' class is further proof, as this occurs when a model produces no false negatives for a class, which is a direct result of predicting it for every sample. The 37688 incorrect predictions represent nearly half the dataset, quantitatively showing the model's uselessness.

In contrast, the filtered Decade Classifier, with its 9.10% accuracy, provides more meaningful metrics despite its poor performance. Crucially, multiple classes now have non-zero F1-scores and computable AUC scores (e.g., 0.66 for the 1830s), demonstrating that the model is no longer collapsed and is attempting to discriminate between classes. An AUC score greater than 0.5 indicates a performance better than random chance for that specific class, a feat the first model could not achieve for any minority class.

11.3 School classifier

Can an explainable, color focused, classifier be successfully able to classify artworks creation geographic location?

The School Classifier, when trained on the complete dataset, achieved a high accuracy of 84.72%. However, much like the initial decade classifier, this figure is profoundly misleading and masks a total model failure. The detailed classification report shows that the precision, recall, and F1-score for every single artistic school are zero, with the sole exception of a class labeled "modern". The model's weighted F1-score of 0.78 is skewed entirely by this single class, while the macro-average F1-score of 0.02 reveals the truth: the model has no ability to classify across different schools. It has learned only to predict "modern" for every input, a behavior confirmed by its classification of all "unknown" data as "modern". The list of misclassifications provides damning evidence, showing that thousands of artworks from historically significant schools like Italian, Dutch, French, and Flemish were all incorrectly swept into the "modern" category. This is a classic case of a model collapsing on a massively imbalanced dataset, where the "modern" class represents the overwhelming majority of samples.

To obtain a more realistic assessment, a second classifier was trained on a dataset with the "modern" class removed. This model's accuracy dropped significantly to 39.68%, but this lower figure represents a more genuine, albeit still flawed, attempt at classification. Unlike the first model, this classifier made predictions across a variety of schools, with classes like China, France, India, and most notably Italy now having non-zero F1-scores. However, the model has simply found the new majority class. The misclassification report shows that the classifier now defaults to predicting "Italy" for most inputs, incorrectly labeling hundreds of artworks from Dutch, French, and Flemish schools. While it no longer suffers from the absolute collapse of the first model, it has merely shifted its bias to the next largest class in the filtered dataset, demonstrating that the underlying features are still not strong enough to overcome the statistical weight of the most represented classes. When tested on "unknown" data, its predictions were limited to a more diverse, yet still biased, set of possibilities.

As with the decade classifiers, **the feature importance analysis for both school classifiers consistently ranked color-related data as the most influential.** The initial model's reliance on color led it to a simplistic solution of predicting "modern". The second model, forced to look deeper, still found color to be the most important signal but was confounded by the immense overlap in palettes between different schools. For example, the vast and varied history of the "Italian" school means its collective color profile is not a unique fingerprint but a broad amalgamation that shares characteristics with many other European schools.

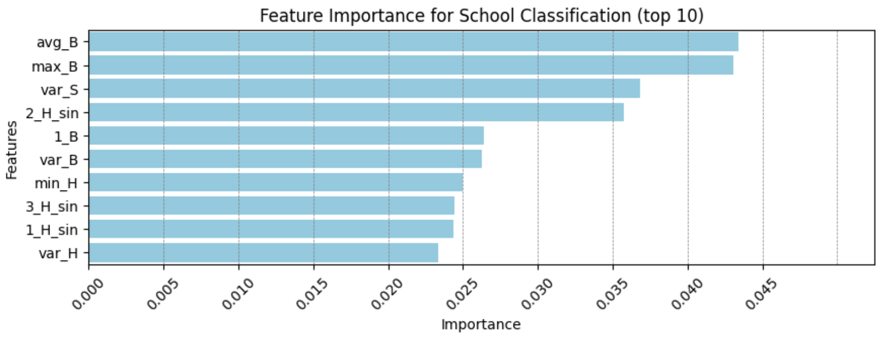

Figure 40: Decades feature importance

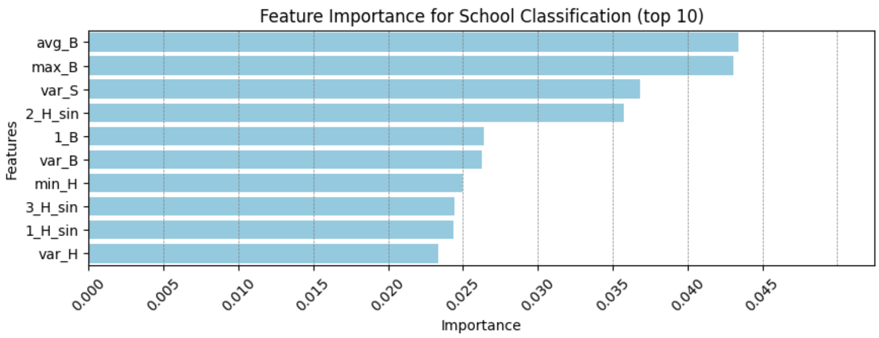

Figure 41: Decades (excluding "modern") feature importance

11.3.1 Quantifiable classifier evaluation

The quantitative metrics for the school classifiers tell a story of bias dictated by data distribution. For the full model, the 84.72% accuracy is a statistical artifact of the "modern" class's dominance. The macro-average F1-score of 0.02 is a more honest metric, quantifying the model's complete inability to identify any of the dozens of minority classes. The **AUC score** of 0.82 for the "modern" class confirms that the model is effective at identifying this one class, but this comes at the cost of ignoring all others, making it practically useless.

The filtered model's 39.68% accuracy, while poor, is a more credible baseline. The weighted **F1-score** of 0.27 and the emergence of non-zero F1-scores for multiple classes (e.g., 0.57 for Italy) show a model that is attempting to discriminate, not just default. The AUC scores further support this: scores well above 0.5 for India (0.86), China (0.81), and Japan (0.78) prove that for these specific classes, the model has learned features that allow it to perform significantly better than random chance. However, the consistent misclassification of other major European schools as "Italian" shows that even with this partial learning, the model's final output is still overwhelmingly governed by the new majority class, "Italy", highlighting the persistent challenge posed by data skew.

11.4 Conclusion

The attempt to classify artworks by decade and school using a color-centric Random Forest model served as a powerful illustration of the challenges inherent in applying machine learning to complex, imbalanced historical data. The experiments revealed that seemingly high accuracy scores (52.45% for decade, 84.72% for school) were statistical artifacts, masking a total model collapse where the classifiers only predicted the majority classes. Even after filtering the dataset, the models' performance remained poor, highlighting that their behavior was dictated by data distribution rather than a meaningful understanding of artistic styles.

Crucially, while color features were consistently ranked as most important, they lacked the discriminative power required for accurate classification. This underscores a key distinction: while a color palette may be correlated with a period or school, it is rarely a unique fingerprint due to the overlapping and fluid nature of art history. Ultimately, this section demonstrates that without addressing profound data imbalances and feature ambiguity, even explainable models may fail, evidencing that black and white classification is an inadequate tool for capturing the nuanced and interconnected evolution of art.

12 Conclusion

In conclusion, this project's journey through the OmniArt dataset highlights a fundamental dichotomy in the computational analysis of art: the distinction between finding patterns and enforcing classifications. **The first phases of this work demonstrated considerable success.** By clustering colors into a perceptually meaningful vocabulary and subsequently clustering artworks based on their palettes,**we were able to identify archetypes that empirically validated established principles of art theory.** These clusters revealed coherent groupings based on hue relationships, saturation harmony, and brightness contrast. Furthermore, itemset mining uncovered strong, statistically significant associations — evidenced by high lift values—between these color-based clusters and specific decades and schools, proving that discernible spatio-temporal patterns are indeed embedded within the chromatic data.

However, the attempt to leverage these patterns for classification starkly illustrated the critical importance of dataset structure. Both the decade and school classifiers failed, not due to a lack of sophisticated methodology, but because their performance was entirely dictated by the dataset's profound imbalances. The models collapsed, learning only to predict the overwhelmingly dominant classes ("2010" and "modern"), rendering them useless for nuanced historical analysis despite deceptively high accuracy scores. This outcome underscores a crucial point: **while clustering and itemset mining can successfully navigate and reveal patterns within a messy dataset, supervised classification is far less forgiving.** For classification, the shape of the dataset—its balance, scope, and the clarity of its labels is paramount, as it dictates the very possibility of a meaningful result.

This challenge is magnified by the intrinsic nature of art historical data. Art is not a collection of discrete, neatly-labeled specimens. It is an interconnected web of influence where styles, locations, and decades blur into one another. Artists traveled, materials were traded globally, and aesthetic ideas were revived and reinterpreted across centuries. This inherent "messiness" means that the quest to find unique, strict fingerprints for a given period or school may be fundamentally misguided, as such unambiguous signatures often do not exist. In such a fluid domain, it is arguably more insightful to group artworks and find emergent patterns than to force them into rigid classificatory boxes. The success of clustering and association rule mining, contrasted with the failure of classification, suggests that for a field as complex as art history, the most fruitful application of data mining lies in uncovering the overlapping, probabilistic relationships that define its evolution, rather than in a deterministic labeling that the history itself so often resists.

References

Albers, J. (1971) Interaction of Color. New Haven: Yale University Press.

Alimpieva, R.V. (1982) ‘Semantic significance of the color designation goluboj’, in Semantic research of lexical and grammatical systems. Sverdlovsk: Sverdlovsk State Pedagogical Institute, pp. 18–25.

Arts & Culture (n.d.) A Colorful History of Paints and Pigments. Available at: https: //artsandculture.google.com/story/ a-colorful-history-of-paints-and-pigments/_wVRps9LN6ctLQ?hl=en

Artsy (2017) A Brief History of Color in Art. Available at: https://www.artsy.net/article/the-art-genome-project-a-brief-history-of-color-in-art

Bengal1 (2021) The-Effect-of-Dataset-Type-on-Artwork-Classification. GitHub repository. Available at: https://github.com/Bengal1/The-Effect-of-Dataset-Type-on-Artwork-Classification

Berlin, B. and Kay, P. (1969) Basic Color Terms: Their Universality and Evolution. Berkeley: University of California Press.

Call, A.V. and Walter Foster Creative Team (2016) Artist’s Toolbox: Color: A practical guide to color and its uses in art. Lake Forest, CA: Walter Foster Publishing.

Computer History Museum (n.d.) Alvy Ray Smith. Available at: https://computerhistory.org/profile/alvy-ray-smith/

Crowe, D.W. (2001) ’Symmetries of Culture’, in BRIDGES: Mathematical Connections in Art, Music, and Science, pp. 1-7.

cs-chan (2018) ArtGAN/WikiArt Dataset/README.md. GitHub. Available at: https://github.com/cschan/ArtGAN/blob/master/WikiArt%20Dataset/README.md

Dean, J. (2010) Wild Color: The Complete Guide to Making and Using Natural Dyes. Revised and Updated edn. New York: Watson-Guptill.

Delamare, F. and Guineau, B. (2000) Color: Making and Using Dyes and Pigments. London: Thames & Hudson.

Deng, Y., Tang, F., Pan, X., Dong, W., Ma, C. and Xu, C. (2022) 'StyTr2: Image Style Transfer With Transformers', in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Desikan, B.S., Shimao, H. and Miton, H. (2022) 'WikiArtVectors: Style and color representations of artworks for cultural analysis via information theoretic measures', Entropy, 24(9), p. 1175.

Desikan, P. (2021) 'Transforming Information Into Knowledge: How Computational Methods Reshape Art History', DHQ: Digital Humanities Quarterly, 15(3). Available at: http://www.digitalhumanities.org/dhq/vol/15/3/000560/000560.html

Deutscher, G. (2010) Through the Language Glass: Why the World Looks Different in Other Languages. London: Arrow.

Du, J. (2025) 'The Palette of History: A historical analysis of color evolution and symbolism exists between Chinese and Western fine art painting traditions', CINEFORUM, 65(2), pp. 497–531.

Edwards, B. (2004) Color: A course in mastering the art of mixing colors. New York: Jeremy P. Tarcher/Penguin.

Gage, J. (2009) Color in Art. London: Thames & Hudson.

Getty (n.d.) Color in the Renaissance. Available at: https://www.getty.edu/education/teachers/building_lessons/formal_analysis.html

Gilbert, A.L., Regier, T., Kay, P. and Ivry, R.B. (2006) 'Whorf hypothesis is supported in the right visual field but not the left', Proceedings of the National Academy of Sciences, 103(2), pp. 489-494.

Goethe, J.W. von (1810) Zur Farbenlehre (Theory of Colours). Tübingen: J.G. Cotta'schen Buchhandlung.

He, L., Qi, H. and Zaretzki, R. (2014) 'Image color transfer to evoke different emotions based on color combinations', Signal, Image and Video Processing, 9(8), pp. 1-4.

huggan (2022) huggan/wikiart. Hugging Face. Available at: https://huggingface.co/datasets/huggan/wikiart

Khadangi, A., Sartipi, A., Tchappi, I. and Fridgen, G. (2025) CognArtive: Large Language Models for Automating Art Analysis and Decoding Aesthetic Elements. arXiv preprint arXiv:2502.04353.

Koenderink, J., van Doorn, A. and Gegenfurtner, K. (2020) 'Colors and Things', i-Perception, 11(6), pp. 1-43.

Lee, S. and Lee, H. (2021) 'The effect of the dataset on the classification of paintings', PLOS ONE, 16(3), p. e0248414.

Liao, P., Zhang, X., Li, H., et al. (2022) ArtBench: A Comprehensive Benchmark for Artwork Recognition. Available at: https://www.andrew.cmu.edu/user/peiyuanl/ArtBench.pdf

Liu, Y., Zhang, Y., Wang, C. and Li, J. (2022) 'Big transfer (BiT) learning for fine art classification', PLoS ONE, 17(6), p. e0268962.

Mohamed, Y., Abdelfattah, M., Alhuwaider, S., Li, F., Zhang, X., Church, K.W. and Elhoseiny, M. (2022) 'ArtELingo: A Million Emotion Annotations of WikiArt with Emphasis on Diversity over Language and Culture', in Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Mohamed, Y., Li, R., Ahmad, I.S., et al. (2024) No Culture Left Behind: ArtELingo-28, a Benchmark of WikiArt with Captions in 28 Languages. arXiv preprint arXiv:2411.03769.

Nassau, K. (1987) 'The fifteen causes of color: The physics and chemistry of color', Color Research & Application, 12(1), pp. 4-23.

Newton, I. (1704) Opticks: or, A treatise of the reflexions, refractions, inflexions and colours of light. London: Sam. Smith and Benj. Walford.

NGV (n.d.) Spectrum: An Exploration of Colour. Available at: https://www.ngv.vic.gov.au/essay/spectrum-an-exploration-of-colour/

OpenDataLab (n.d.) OmniArt. Available at: https://opendatalab.com/OpenDataLab/OmniArt

Papers with Code (n.d.) OmniArt Dataset. Available at: https://paperswithcode.com/dataset/omniart

Papers with Code (n.d.) WikiArt Dataset. Available at: https://paperswithcode.com/dataset/wikiart

Pastoureau, M. (2009) Blue: The History of a Color. Princeton: Princeton University Press.

Regier, T. and Kay, P. (2009) 'Language, thought, and color: Whorf was half right', Trends in Cognitive Sciences, 13(10), pp. 439-446.

Roberson, D., Davidoff, J., Davies, I.R.L. and Shapiro, L.R. (2005) 'Color categories: Evidence for the cultural relativity hypothesis', Cognitive Psychology, 50(4), pp. 378-411.

Saleh, B. and Elgammal, A. (2015) 'Large-scale Classification of Fine-Art Paintings: Learning The Right Metric on The Right Feature', in Proceedings of the International Conference on Computer Vision Workshops.

Shi, Y.Q. (n.d.) Identifying Computer Graphics using HSV Color Model and Statistical Moments of Characteristic Functions.

Smith, A.R. (1978) 'Color Gamut Transform Pairs', Computer Graphics, 12(3), pp. 12-19.

Smith, A.R. (2001) 'Digital Paint Systems: An Anecdotal and Historical Overview', IEEE Annals of the History of Computing, 23(2), pp. 4-30.

St. Clair, K. (2017) The Secret Lives of Color. New York: Penguin Books.

steubk (2021) WikiArt. Kaggle. Available at: https://www.kaggle.com/datasets/steubk/wikiart

Strezoski, G. (n.d.) OmniArt. Available at: https://isis-data.science.uva.nl/strezoski/

Strezoski, G. (2018) OmniArt in ACM TOMM. Available at: https://strezoski.com/posts/my-first-post/

Strezoski, G. and Worring, M. (2017) OmniArt: Multi-task Deep Learning for Artistic Data Analysis. arXiv preprint arXiv:1708.00684.

Strezoski, G. and Worring, M. (2018) 'OmniArt: A Large-scale Artistic Benchmark', ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 14(4), pp. 1-22.

The Metropolitan Museum of Art (2011) Impressionism: Art and Modernity. Available at: https://www.metmuseum.org/essays/impressionism-art-and-modernity

VISTORY (n.d.) OmniArt. Available at: http://www.vistory-omniart.com/

Wasielewski, A. (2022) Computational Formalism: Art History and Machine Learning. Cambridge, MA: MIT Press. Available at: https://direct.mit.edu/books/oa-monograph/5450/ Computational -FormalismArt-History-and-Machine

Wasielewski, L. (2021) 'Beyond the Universal Art Dataset: Issues and Mitigations of Western Bias in Computational Art Analysis', in Proceedings of the VISART Workshop on Computer Vision for Art Analysis.

Whorf, B.L. (1956) Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf. Cambridge, MA: MIT Press.

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R. and Boroditsky, L. (2007) 'Russian blues reveal effects of language on color discrimination', Proceedings of the National Academy of Sciences, 104(19), pp. 7780-7785.

Winsor & Newton (n.d.) History of pigments. Available at: https://www.winsornewton.com/blogs/articles/history-of-pigments

Worring, M., Rudinac, S., Efthymiou, A., Kackovic, M. and Wijnberg, N. (2021) Graph Neural Networks for Knowledge Enhanced Visual Representation of Paintings. arXiv preprint arXiv:2105.08092.

World History Encyclopedia (2020) Colour & Technique in Renaissance Painting. Available at: https://www.worldhistory.org/article/1628/colour–technique-in-renaissance-painting/