

UNVEILING COLOR DYNAMICS AND VALUE OF ANDY WARHOL'S "SHOT MARILYNS": A STUDY ON VISUAL VARIATIONS AND PERCEPTION

A PREPRINT

 **Hengyuan Liu** *


University of California, Los Angeles
hengyuanliu@ucla.edu

 **Li Yuan** *

Swiss Federal Institute of Technology, Zurich
liyuan1@ethz.ch

 **Kathy Mo**

University of California, Los Angeles
kathymo24@ucla.edu

 **Xinhui Luo**

Tufts University
xinhui.luo@tufts.edu

 **Weilin Cheng**

University of California, Davis
wncheng@ucdavis.edu

 **Erick S. Arenas V.**

University of California, Davis
esarenas@ucdavis.edu

November 2, 2024

ABSTRACT

This study investigates the visual dynamics and value of Andy Warhol's "Shot Marilyns" series through the innovative application of statistical techniques, including entropy, K-Means clustering, and K-Nearest Neighbors, alongside traditional analytical methods. This approach provides a comprehensive analysis of color distribution, regional variation, and the overall value of the masterpiece. The results reveal significant inter-painting variations and uncover intricate color dependencies that challenge assumptions of uniformity. Furthermore, the restoration of the damaged region in "Blue Marilyn" highlights the complexity of Warhol's color choices and contributes to the ongoing discussion about the potential value of the series. These findings offer new insights into Warhol's aesthetic decisions, deepening our understanding of the role of color perception in contemporary art.

Keywords shot marilyns · pop art · andy warhol · region of interest

1 Introduction

In May 2022, Andy Warhol's "Sage Blue Marilyn" portrait set a new auction record, selling for \$195 million, as reported by Vankin [2022] on the Los Angeles Times. This unprecedented sale has renewed both public and scholarly interest in Warhol's work, highlighting the enduring impact of his art on contemporary culture. The "Shot Marilyns" series holds immense value, not only monetarily but also in its profound impact on contemporary art and its reflection of societal themes. The record-breaking auction underscores its continued relevance and fascination [Vankin, 2022]. Furthermore, Marilyn Monroe remains an iconic figure whose image has permeated popular culture [Schmidt, 2022]. Her tragic life story, coupled with her enduring allure, makes her an intriguing subject for artistic exploration [Gallery, 2019].

Warhol's unique art style, characterized by his use of silkscreen printing and vibrant color schemes, offers a rich field for visual analysis. His method of mass-producing images and manipulating colors challenge traditional notions of art and celebrity, making the "Shot Marilyns" series a perfect case study for understanding his innovative approach [Lanchner, 2017]. Warhol's work satirized and celebrated materialism and celebrity culture. His focus on celebrity paintings could

*Both authors contributed equally to this research.

be a critique of the obsession with celebrities, but he also painted images of consumer products, suggesting appreciation of consumerism [BIOGRAPHY, 2022]. He embraced the idea that art was something you could "get away with," symbolizing an era when pop culture and celebrity became self-sustaining phenomena [Rozenman, 2023]. Notably, some of his most famous works are of female celebrities with tragic lives, such as Jacqueline Kennedy and Marilyn Monroe. He had a traumatic childhood and had a fear of death, so he related to the lives of the women he painted [Fallon and Warhol, 2010].

In 1964, amidst the bustling atmosphere of Andy Warhol's studio, The Factory, a significant event led to the creation of the "Shot Marilyns" series. Warhol, deeply influenced by Marilyn Monroe's tragic death from a drug overdose in 1962 [Bolton, 2002], began producing silkscreen portraits of her, capturing the iconic actress's image through repetitive, vivid depictions [Christie's, 2022]. The "Shot Marilyns" series features five portraits shown in Figure 1, each rendered in different color schemes.



Figure 1: The five portraits in Andy Warhol's "Shot Marilyns" series, each showcasing Marilyn Monroe in distinct color schemes © 2024 The Andy Warhol Foundation for the Visual Arts, Inc. / Licensed by Artists Rights Society (ARS), New York

The name "Shot Marilyns" originates from an incident involving Dorothy Podber, a performance artist and frequent visitor to The Factory. One day, Podber, accompanied by Warhol's friend and photographer Bill Name, observed the Marilyn portraits lined up against a wall. She asked Warhol for permission to "shoot" them, which Warhol, interpreting it as a request to photograph the artworks, granted. Unexpectedly, Podber pulled out a revolver and fired a shot, piercing four of the five canvases through the forehead [Ghighi, 2022]. This act of violence not only created physical damage but also added a layer of historical intrigue and controversy to Warhol's work, further embedding it into the fabric of pop culture and art history.

Examining the academic and artistic fields, there are various analyses of Andy Warhol's "Shot Marilyns" from the perspectives of history, art development, color theory, and more. However, these viewpoints can sometimes be subjective. We are pioneers in combining statistical approaches with traditional methods of analysis, and we believe this will provide a deeper and more objective understanding of Warhol's masterpiece. Specifically, through this analysis, we aim to uncover new insights into the interplay between celebrity, media, and art, enriching our understanding of both Warhol's work and Monroe's legacy.

In this paper, we aim to conduct a comprehensive analysis of Andy Warhol's "Shot Marilyns" series using several advanced techniques. First, we will analyze the relative conditional entropy of the pixel color distribution in RGB (red, green, blue) space to understand the variations in color across the different portraits. This will provide insights into the underlying patterns and complexity of Warhol's use of color. Next, we will create 3D scatter plots to visualize how each pixel color is distributed in the RGB space, enabling us to observe the distinct color palettes used in each image.

We will also apply K-means cluster analysis to identify and compare the primary color clusters within the portraits, highlighting different regions of interest such as the background, hair, eyeshadow, and face. Additionally, we will focus on digitally repairing the "Blue Marilyn" using K-Nearest Neighbors to model and analyze the RGB distribution around the gunshot-damaged area. This restoration will involve capturing the gunshot region and using K-Nearest Neighbors algorithm to reconstruct the damaged section, preserving the artwork's integrity. Through these methods, we aim to gain a deeper understanding of Warhol's artistic techniques, the visual impact of his "Shot Marilyns" series, and the series value. While our analysis strives for objectivity, we acknowledge that interpretations of art can be inherently subjective.

2 Methods

An image is composed of pixels, each containing three color components: Red (R), Green (G), and Blue (B), denoted as (R, G, B) respectively. These components determine the intensity of their respective colors, with each component represented by an integer value within the range of 0 to 255 in the RGB color space. Therefore, each color component is a discrete variable capable of assuming 256 distinct values. In the equations below, $Y = y$ or $X = x$ can be selected from any of the three color components, R, G, or B. For this study, each image in the "Shot Marilyns" series has a resolution of 960 by 960 pixels.

2.1 Entropy Calculation

As Sigaki et al. [2018] stated, "different artistic styles have a distinct average degree of entropy and complexity." Analyzing cross entropy can help in identifying which colors dominate in certain images and how these colors interact with one another. This can reveal the underlying artistic choices made by the artist.

The probability of a specific color component, $P(Y = y)$, is determined by dividing the number of pixels with color coordinates corresponding to that component by the total number of color components in the entire image. The following equations illustrate the calculation of entropy, conditional entropy, and relative conditional entropy introduced by Shannon [1948].

The entropy of a color component Y is defined as:

$$H(Y) = - \sum_{y=0}^{255} P(Y = y) \cdot \log(P(Y = y)) \quad (1)$$

The conditional entropy of Y given X is given by:

$$H(Y|X) = \sum_{x=0}^{255} P(X = x) \cdot H(Y|X = x) = - \sum_{x=0}^{255} \sum_{y=0}^{255} P(X = x, Y = y) \log_2 \left(\frac{P(X = x, Y = y)}{P(X = x)} \right) \quad (2)$$

The relative conditional entropy is calculated using the following formula:

$$HR(X|Y) = \frac{H(X|Y)}{H(X)} \quad (3)$$

Incorporating concepts from information theory, as discussed by Cover and Thomas [2012], offers a valuable framework for understanding entropy's role in visual art analysis. The principles of entropy, cross entropy, and conditional cross entropy enable us to quantify the complexity and distribution of color in the "Shot Marilyns," revealing relationships between different color palettes. Insights from Gage [1999] further emphasize how these color choices evoke emotional responses. By integrating these perspectives, we can see how entropy measures not only capture statistical properties but also illuminate the artistic intent behind these iconic works.

2.2 K-Means Clustering Analysis

In the clustering analysis, we applied K-Means clustering to examine the color dynamics in Andy Warhol's "Shot Marilyns" series. For each portrait and region of interest, we specified 15 clusters and used the "k-means++" initialization method. This initialization method, introduced by Arthur and Vassilvitskii [2007], improves the convergence speed and accuracy of the K-Means algorithm by spreading out the initial cluster centers. This method is particularly effective in avoiding poor clustering results due to the random placement of initial centroids.

Mathematically, the K-Means algorithm minimizes the following objective function:

$$J = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (4)$$

where k is the number of clusters, C_i is the set of points belonging to cluster i , x represents a data point, and μ_i is the centroid of cluster i . The "k-means++" algorithm initializes the centroids by first selecting one random data point as the first centroid. Subsequent centroids are chosen based on a probability proportional to the squared distance from the nearest existing centroid. This process can be expressed as:

$$P(x) = \frac{D(x)^2}{\sum_{x' \in X} D(x')^2} \quad (5)$$

where $D(x)$ is the distance from the point x to the nearest centroid already chosen.

This method allowed us to quantify and visualize color distribution, revealing underlying patterns and variations within the artworks. The resulting clusters were analyzed to assess the prominence of specific colors across the series, illustrated in bar charts and ribbon visualizations. These visuals highlight Warhol's distinctive color schemes, offering insights into his artistic technique and color usage.

2.3 Region of Interest Extraction

In our region of interest (ROI) analysis, we targeted specific segments of the images such as the background, hair, eyeshadow, and face. We began by converting the images to the HSV color space using OpenCV's conversion functions, which facilitate more effective identification and segmentation of specific color ranges [Flores-Vidal et al., 2022]. By manually determining the minimum and maximum HSV values within selected regions, we created color masks using OpenCV's masking functions to isolate these target areas. These masks highlighted the pixels that fell within the specified HSV range, effectively isolating the desired colors from the rest of the image. Once the masks were applied, we used image processing techniques to extract only the parts of the image that matched the mask, discarding the rest. The resultant ROIs were then processed and saved for K-Means clustering analysis. This method, enhanced by the precise capabilities of OpenCV and guided by best practices from Culjak et al. [2012], enabled us to highlight specific color features in Warhol's artwork, providing nuanced insights into his use of color and its variations, and ensuring accurate and efficient color segmentation and analysis.

2.4 K-Nearest Neighbors Restoration of "Blue Marilyn"

To repair the damaged sections of the "Blue Marilyn" image, we employed K-Nearest Neighbors (KNN) regression for image restoration, inspired by Lee et al. [2022]. This method involves identifying the coordinates of the damaged pixels, shown as a zoomed-in area in Figure 2, and using the surrounding undamaged pixels to predict their values.



Figure 2: The damaged area of "Blue Marilyn"

The KNN regression model, using 8 nearest neighbors, was trained on the RGB values of undamaged pixels. The model then predicted RGB values for the damaged pixels, effectively restoring the affected area.

Figure 3 demonstrates this process. The damaged pixel at coordinate (415, 425) originally had an RGB value of (227, 164, 183). The KNN regressor then identified the 8 nearest neighbors from the undamaged pixels, with numbers on the edges indicating their distance from the damaged pixel. Finally, the KNN algorithm averaged red, green, and blue

channel values respectively, resulting in a predicted RGB value of (223, 155, 177). This fixed the pixel at coordinate (415, 425).

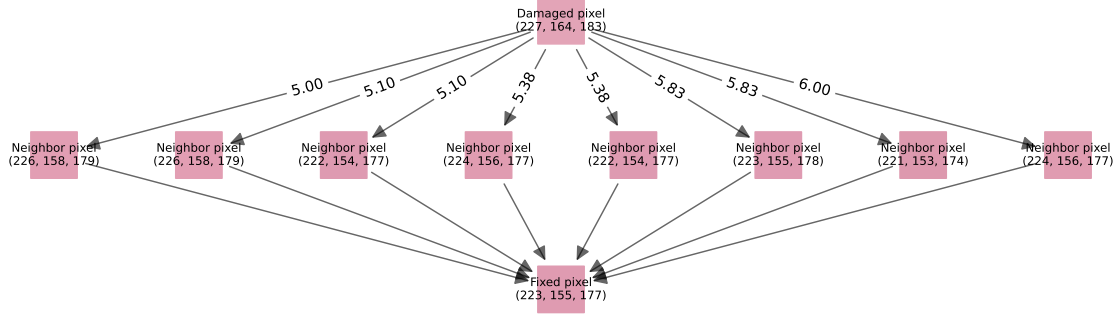


Figure 3: KNN restoration process of a damaged pixel at (415, 425) in "Blue Marilyn"

This approach preserved the visual consistency of the image by leveraging spatial color information from the undamaged regions.

3 Data Description and Visualization

Our study utilized a dataset comprising five portraits titled "Orange Marilyn," "Red Marilyn," "Turquoise Marilyn," "Blue Marilyn," and "Sage Blue Marilyn," as shown in Figure 1. Each portrait is digitally encoded in RGB (Red, Green, and Blue) color channels, which synthesize a spectrum of colors through additive mixing of red, green, and blue lights. To further analyze Warhol's color strategies in "Shot Marilyns" and enhance future data visualizations, we converted the RGB values into Hexadecimal representations. Each image measures 960 by 960 pixels, resulting in 921,600 unique data points per image, each specified by a distinct location and chromatic composition. In this additive color model, the intensity of each primary color (Red, Green, Blue) is quantized into discrete levels ranging from 0 to 255, providing a finite palette within this cubic color space. Each pixel's color is quantified based on the RGB values, making it part of a discrete color space where the combination of these three channels can reproduce a wide array of colors.

3.1 Distribution

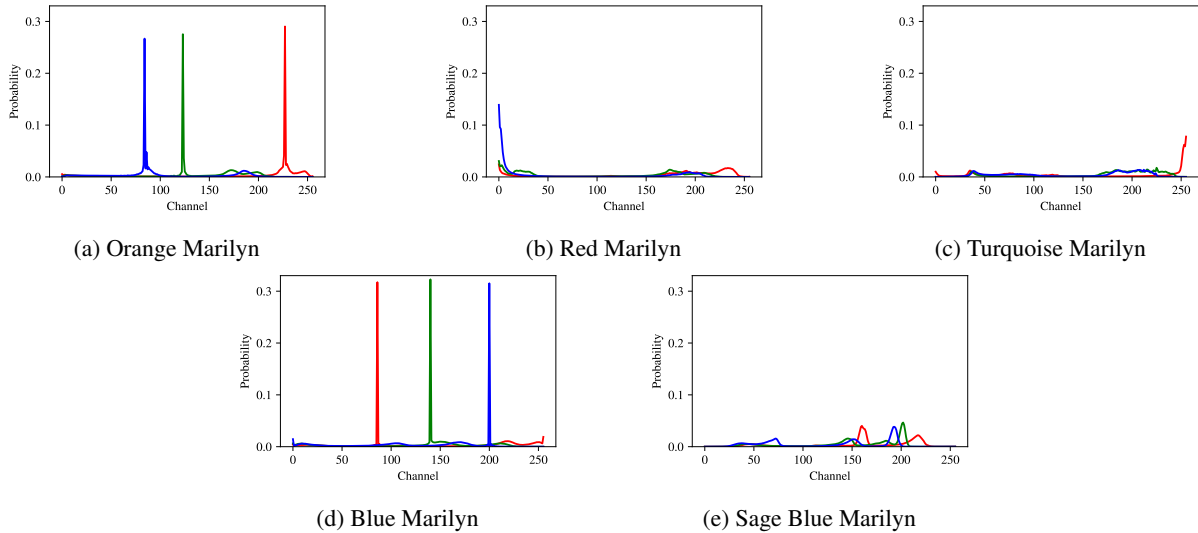


Figure 4: Distribution of RGB channel values for the five "Shot Marilyns" portraits

Our initial analysis involved examining the distribution profiles of the RGB channels in the portraits. Figure 4 illustrates the variations in the red, green, and blue distributions across the five portraits, with Figure 4a for "Orange Marilyn"