Analysis of Shot Marilyns by Andy Warhol

Hengyuan Liu, Xinhui Luo, Weilin Cheng, Erick S. Arenas V.

Abstract - **This paper explores the differences and commons between five different versions of Shot Marilyns done by Andy Warhol. We conducted a comprehensive analysis of color composition and distribution in a set of art pieces. Using various methods such as relative conditional entropy, we explored the distinct color distributions and correlations within each image. By clustering the images and examining specific regions of interest (ROIs), including backgrounds, hair, eyeshadow, and face, we gained detailed insights into the construction and differences of each image. We also observed that initial expectations of uniform colors for certain elements were not met, highlighting the complexity and deceptive nature of color perception in art.**

I. INTRODUCTION

Shot Marilyns is a renowned artwork by the iconic American artist Andy Warhol. It was produced in 1964 and consists of multiple images of Marilyn Monroe[1]. The series was created after Monroe's tragic death and reflects on her status as a pop culture icon. Warhol was deeply fascinated by the concept of celebrity culture and the mass production of images. He sought to explore the intersection of fame, consumerism, and the impact of media on society through his art. In August 1962, Andy Warhol began using silkscreening for its assembly line effect in his art. He chose photographs, enlarged them, transferred them onto silk with glue, and rolled ink across the screen to create repeated images with slight variations. Inspired by Marilyn Monroe's death that same month, Warhol used this technique to make screenprints of her face, employing photo stencils and different colored inks to reproduce the photographic images onto canvas[8]. The Shot Marilyn's artwork exemplifies Warhol's unique approach to capturing the essence of iconic figures. Warhol's idea behind Shot Marilyns was to deconstruct Monroe's image through repetition. He took a single photograph of Monroe and reproduced it multiple times. Each repetition featured variations in color and composition, allowing Warhol to experiment with different visual effects and interpretations.

The repetition of Monroe's face in Shot Marilyns can be seen as a commentary on the omnipresence and commodification of celebrities in popular culture. It reflects how the media inundated society with images of famous figures, turning them into recognizable symbols. By using bright and contrasting colors, Warhol aimed to capture the vibrancy and allure of Monroe's celebrity persona. The intense colors also convey a sense of energy and dynamism, emphasizing her impact and magnetism on society. In this project, we are interested in the pixel color distribution in RGB (red, green, blue) space among the five images of Shot Marilyns and the relationship between pairs of primary colors in every five paintings by the method of relative conditional entropy. Different distributions of pixel colors represent a different mood for the artist Andy Warhol. For example, an image with a red-dominated color can depict symbolism such as love, passion, and desire. To reach our goals, we are planning to use scatter plots to explore how each pixel color is distributed in 2D RGB space and try to observe the difference in the color for each image based on ROI (regions of interest), such as the color of the background, hair, eyeshadow, and face.

Furthermore 1964, following Andy Warhol's creation of Five Shot Marilyns, an incident occurred involving a friend and photographer named Bill Name, who accompanied Dorothy Podber to The Factory. Podber, known as a performance artist and collaborator, observed the Marilyns positioned against a wall and requested permission from Warhol to "shoot" them. Misinterpreting her request as photographing the artworks, Warhol gave his consent. To everyone's surprise, Podber removed her gloves, pulled a revolver from her bag, and fired a shot at the forehead of the Marilyn portraits. Only four out of the five canvases were pierced by the bullet, creating the series known as Warhol's "Shot Marilyns" [5]. In this project, we will repair the gunshot part for the "Blue Marilyn" image by capturing the gunshot area and sampling this area based on the RGB distribution.

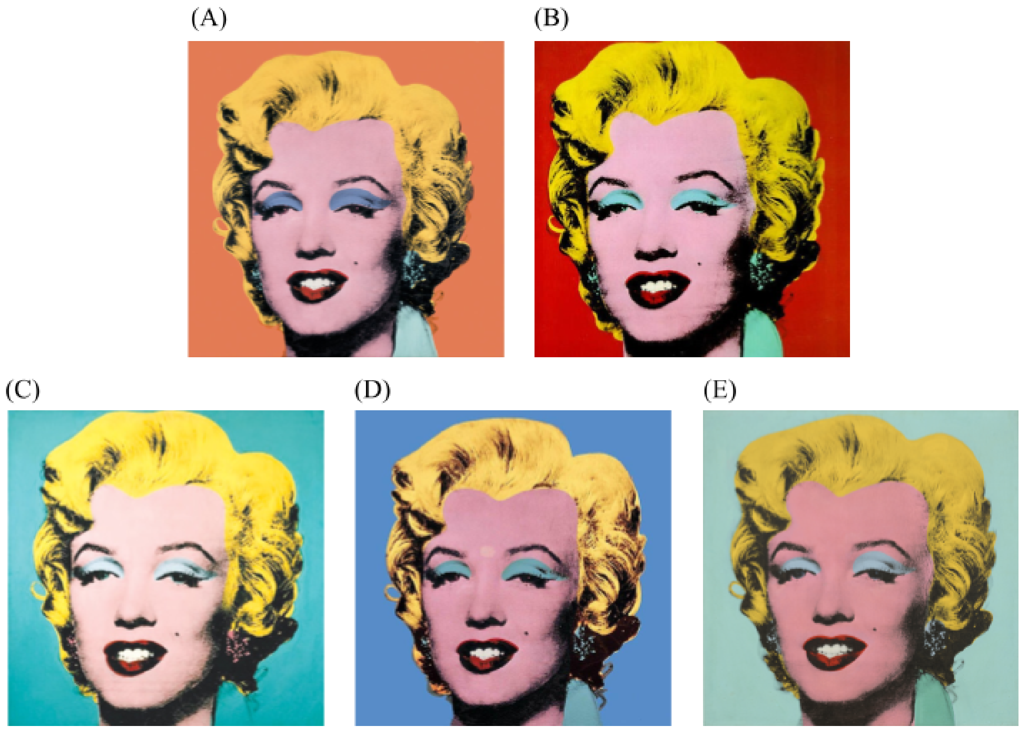

Figure 1. Picture of the five paintings of shots of Marilyn Monroe from Andy Warhol: **(A)** Orange Marilyn **(B)** Red Marilyn **(C)** Turquoise Marilyn **(D)** Blue Marilyn **(E)** Eggblue Marilyn

II. METHODS

Relative Conditional Entropy (HR)

Image is composed of pixels, and every pixel has 3 color components which, are (R, G, B), that define the amount of Red(R), Green(G), and Blue(B). In the RGB channels, each of them is an integer in the range [0, 255]. Thus, each is a discrete variable that can take 256 distinct values. Y=y or X=x can be taken from any of R, G, and B in the following equations. The probability of one color component P(Y=y) is the number of elements of the set of pixels such that its color coordinates for that color are divided by the total number of the color components in the whole image. Here are the equations calculating entropy, conditional entropy, and relative conditional entropy[4][6]:

Entropy:

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

Conditional Entropy:

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

Relative Conditional Entropy:

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

Hierarchical Clustering Analysis (HCA)

Clustering involves the task of grouping objects in a way that objects within the same cluster are more similar to each other compared to objects in different clusters. In our analyses, the objects we considered were pixels within an image. Our project employed Hierarchical Clustering Analysis as an algorithm, which considers individual pixels as the objects of interest.[7]. Each pixel, denoted as P(i), was defined by its color, c(i), represented by a combination of its components in the primary colors: red (R), green (G), and blue (B). In other words, c(i) = (r(i), g(i), b(i)), where each component belongs to the discrete 3D space $[0, 255]^3$. The similarity between two pixels, P(i) and P(j), was determined by calculating the Euclidean distance between their colors[4]:

$$d(P(i), P(j)) = \sqrt{(r(i) - r(j))^2 + (g(i) - g(j))^2 + (b(i) - b(j))^2} \quad (4)$$

In our project, our clustered color bars include two parts. The first color bar consists of all the colors that exist in the analysis and shows which of them are having a closer distance. The second color bar is based on the cluster results and refilled with an average color based on that specific cluster.

Center of RGB Cubes

To start the process, we partitioned the RGB color space from 0 to 255 in each dimension into a set of color cubes measuring 3x3x3. Subsequently, every pixel within an image was assigned to the corresponding cube based on color. For representation purposes, we chose the center of each cube to represent the color of the entire cube. We also experimented with larger cubes, such as 8x8x8, but found that these resulted in color plots with a limited number of displayed colors[10].

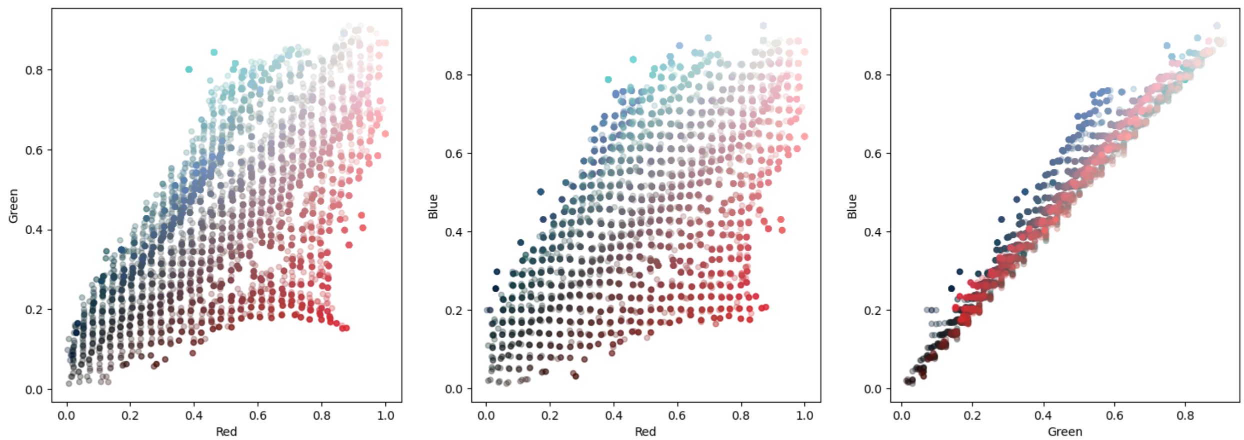

Figure 2. Scatter Plot of 8x8x8 Color Cube Distribution.

In **Figure 2**, we see the result for a cube of 8x8x8, representing each cube's centroid. However, we decided to switch to a 3x3x3 cube to capture more color variation. From this point onwards, the distribution of the artwork is based on the 3x3x3 cube.

II. DATA DESCRIPTION

We have access to the five images based on each of its URL links, which are Orange Marilyn, Red Marilyn, Turquoise Marilyn, Blue Marilyn, and Eggblue Marilyn, from "The Interior Review" website[1]. Each image is composed of pixels, and each pixel has a unique location and color. The color of a pixel is described using the RGB color model, which combines Red, Green, and Blue lights to create various colors. In digital images, the intensity of each RGB component is represented by discrete values ranging from 0 to 255. This means that the color space of a pixel is discrete and consists of a finite set of possible colors within this range. Each image is 960 in length multiplied by 960 wide, which means each image has 921,600 pixels.

III. DATA EXPLORATION/VISUALIZATION ANALYSIS

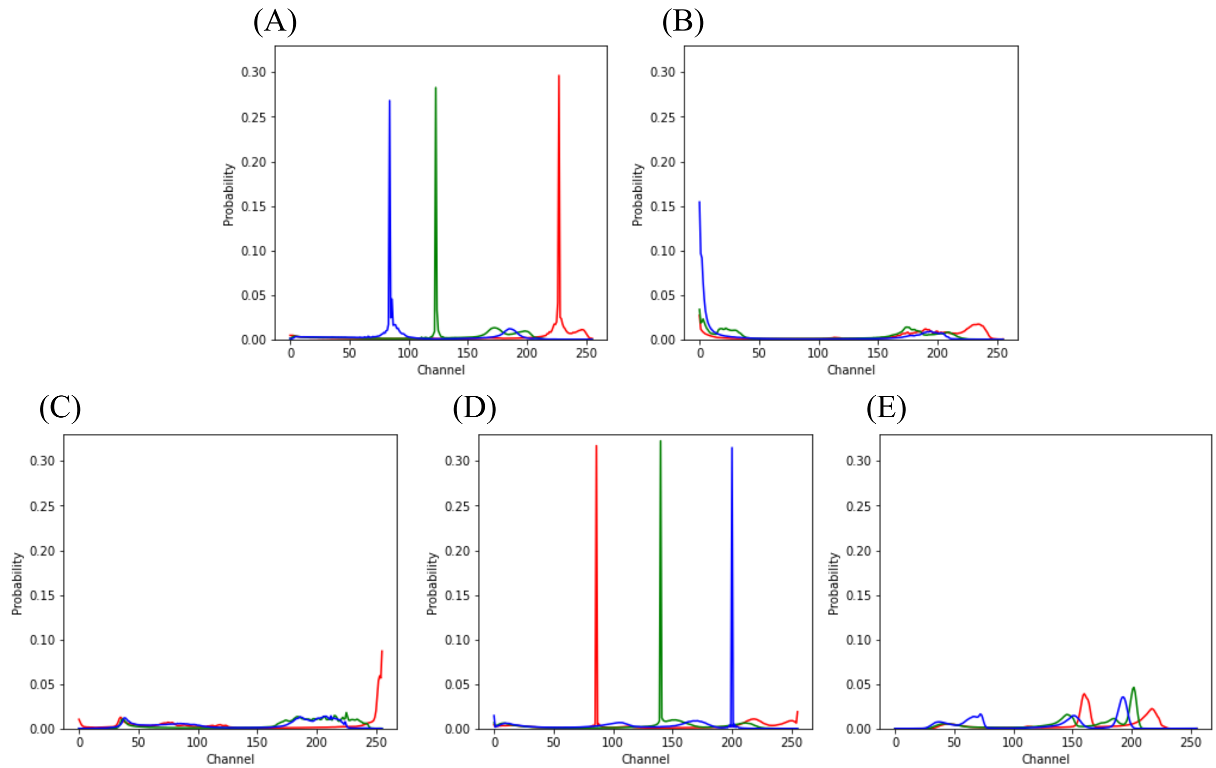

Figure 3. Distributions of values of Red, Green, and Blue channels for five images with all pixels: **(A)** Orange Marilyn RGB Distributions **(B)** Red Marilyn RGB Distributions **(C)** Turquoise Marilyn RGB Distributions **(D)** Blue Marilyn RGB Distributions **(E)** Eggblue Marilyn RGB Distributions

The first analysis we want to apply to the images is the examination of the RGB channels and the observation of the distributions. The distributions of red, green, and blue are different between the five images (see **Figure 3**). The distributions observed in images (A) and (D) exhibit noticeable distinctions compared to the remaining three images. In the (A) Orange Marilyn image, the highest probability of blue is concentrated in the range of [50, 100], with an approximate probability of 27%. The highest probability of green is focused on the range of [120, 130], with an approximate probability of 28%.

Furthermore, the highest probability of blue is concentrated in the range of [220, 240], with an approximate probability of 29%. Remarkably, in the image (D), a striking similarity can be observed with the image (A) in terms of the concentration of high probability values for the green channel, specifically within the range of [120, 130]. However, the ranges for the red and blue channels in the image (D) differ significantly. The highest probability of red in image (D) is concentrated in the range of [70, 80], accounting for approximately 32% probability. On the other hand, the highest probability of blue is concentrated in the range of [190, 210], also with an approximate probability of 32%. This discrepancy in the ranges of red and blue between image (D) and image (A) underscores their distinct characteristics within the overall distribution.

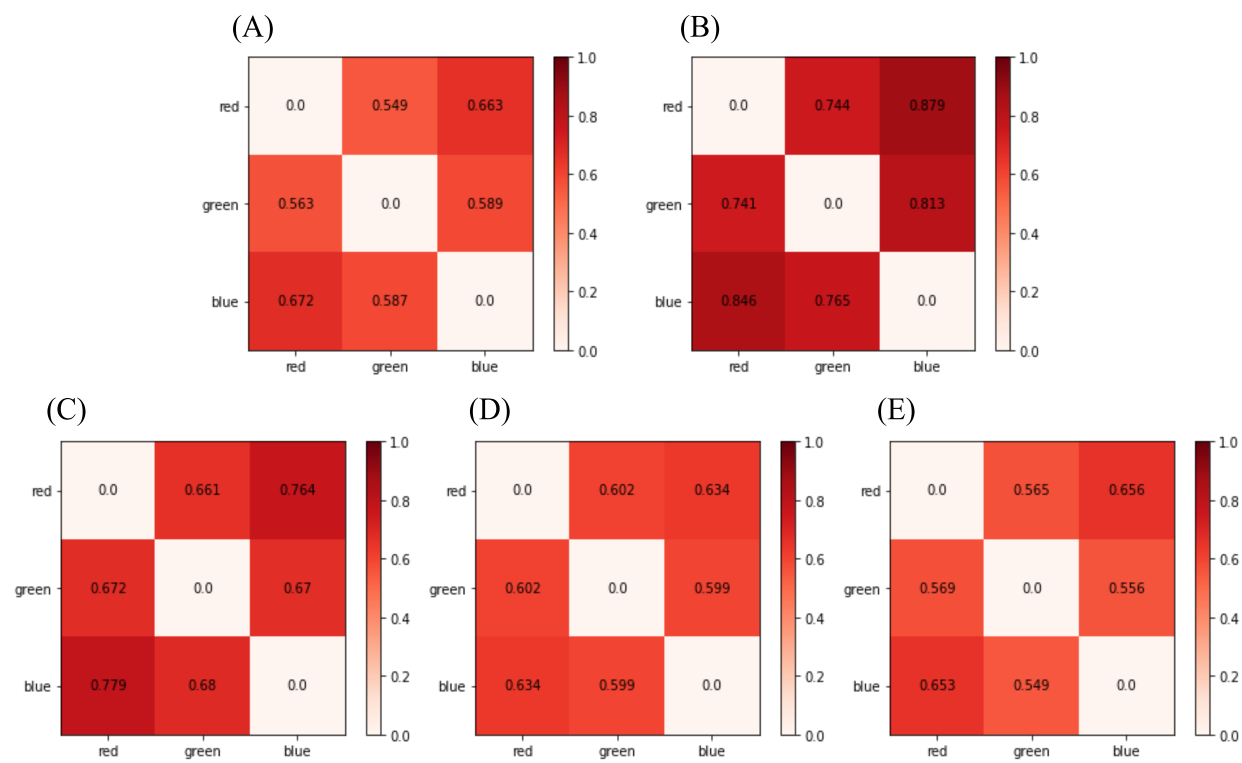

Figure 4. The relative conditional entropy values among the red, green, and blue coordinates of pixels in different images:**(A)** Orange Marilyn, **(B)** Red Marilyn, **(C)** Turquoise Marilyn, **(D)** Blue Marilyn, and **(E)** Eggblue Marilyn.

After analyzing the RGB distribution of each image, we further examined the relationship between pairs of primary colors in all five images by calculating their relative conditional entropy, denoted as HR (See METHODS). It measures how much information is shared or dependent between the two color channels. A lower relative conditional entropy indicates a stronger relationship or dependency between the color channels within images. In comparison, a higher relative conditional entropy indicates a weaker relationship or more independent color channels. The HR value falls within the range of 0 to 1, with 0 indicating complete determination between the colors and 1 representing full independence.

In **Figure 4**, we present the HR values for the nine color pairs in tabular form for each of the five images: Orange Marilyn, Red Marilyn, Turquoise Marilyn, Blue Marilyn, and Eggblue Marilyn. As anticipated, comparing the color to itself results in a conditional entropy of zero. On the other hand, the high HR values for different color pairs indicate a minimal dependence between them. For example, in the image of Red Marilyn, we found the relative conditional entropy of the Red channel given the Blue channel is 0.879, indicating strong independence between the Red and Blue channels.

IV. Clustering based on Whole Images

Next, we decided to plot our RGB colors in a 2D space for each color against each other for each whole image. In each panel below, from **Figure 5** to **Figure 9**, we present a series of dots, each representing a single pixel. These top three scatter plots illustrate how each pixel color is distributed in 2D RGB space and show each pixel on two combinations of different coordinates (Red, Green, and Blue). The bottom three graphs illustrate that the color of each dot is represented by a shade of grayscale, indicating the intensity of its red, green, or blue components. The grayscale ranges from low intensity (appearing white) to high intensity (appearing black). Specifically, the leftmost panel showcases the relationship between red and green (with blue displayed in grayscale), the middle panel demonstrates the interplay between red and blue (with green displayed in grayscale), and the rightmost panel depicts the interaction between green and blue (with red displayed in grayscale).

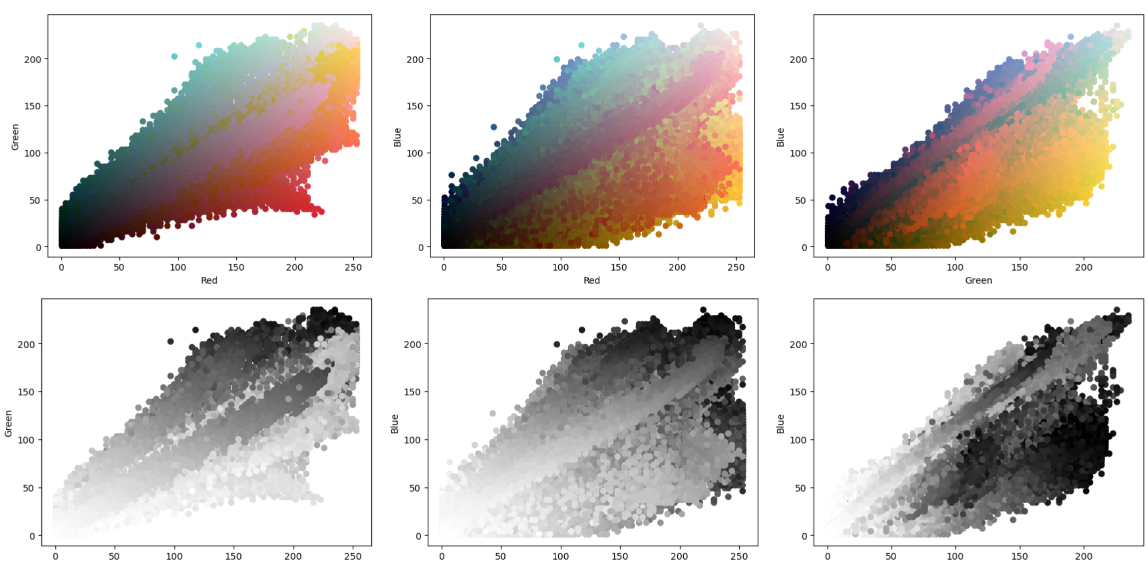

Figure 5. The RGB space occupied by the pixels for the entire image of Orange Marilyn

In **Figure 5**, we can see on the top 3 three graphs that we have the distribution of the RGB color for the Orange painting of Marilyn Monroe. On the far left plot, we have Green VS. Red; in the middle, we have Blue VS. Red; and on the far right, we have Blue VS. Green. Each point represents a centroid for each pixel in the painting. We can see that yellow and orange are prominent in all scatter plots. The bottom three graphs are the same scatter plot as the top three, but in this case, the grayscale represents the prominence of each color not present in the 2D graph. Meaning that in the far left plot, the darker the points, the more presence of blue; in the center scatter plot, the darker the color, the more prominent green it is. Finally, on the far right scatter plot, we find that the darker points represent the prominence of Red. On both the left and center scatter plots, we find that the most remarkable presence of both Blue and Green is near the right corner of the graph and is not as prominent. While we can see that in the far right scatter plot, the points are dark through most of the scatter plot. This means that red is present and is more significant than both the blue and green colors. This makes sense because this picture has an orange background and yellow hair, which both have a high content of Red to be created.

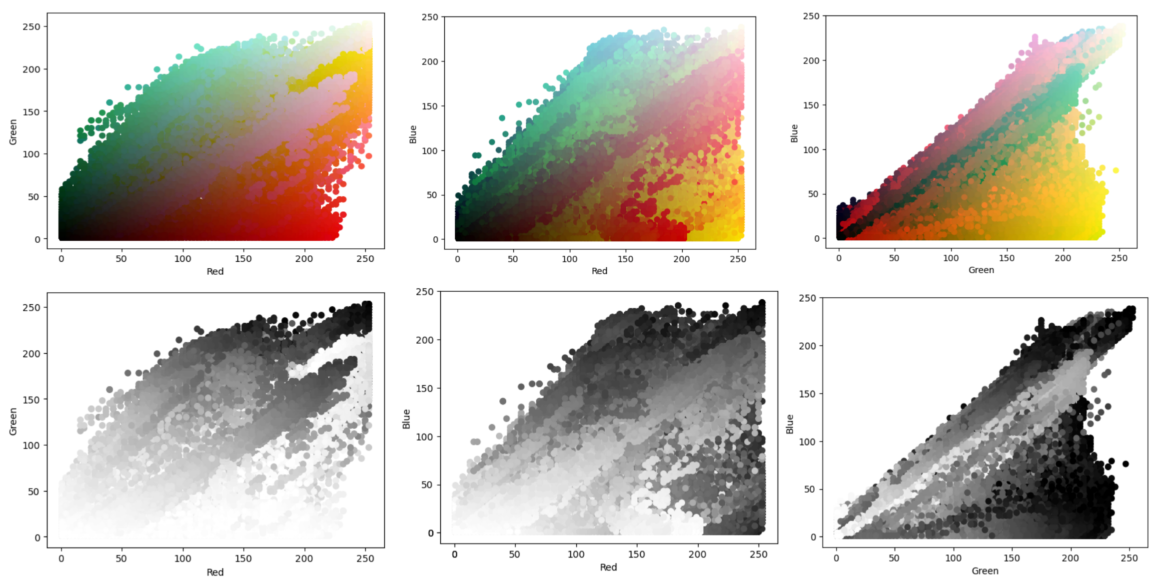

Figure 6. The RGB space occupied by the pixels for the entire image of Red Marilyn

In **Figure 6**, we can see that for the top 3 scatter plots, the distribution of color where the far left side plot is based on Green VS. Red, the center scatter plot is Blue VS. Red, and the far left field is based on Blue VS. Green for the red shot of Marilyn Monroe. We find it hard to understand the plot with all the colors present. However, the distribution of color seems to be more even in this case. In the bottom 3 scatter plots, we see the same plot in a grayscale. The darker the pixels, the more prominent the color that is not present in the axis. On the left plot, the darker the pixel, the more prominence the presence of Blue. On the center plot, we find that the darker the color, the more prominent the presence of Green. In the far right plot, we find that the darker the pixel, the more prominent the Red. We can see that on the far left side, there is no high prominence of the Blue color. While in the center plot and the right plot, we find more prominence of green and red colors. Red seems more prominent overall, which makes sense due to the red background color. However, Green still has a high prominence.

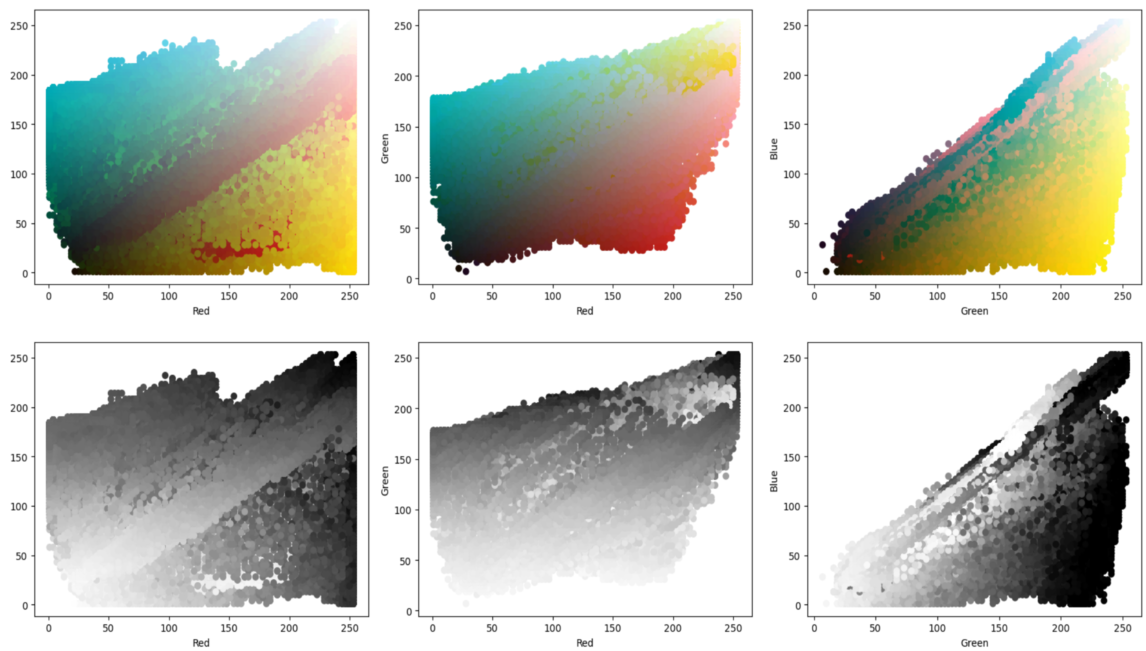

Figure 7. The RGB space occupied by the pixels for the entire image of Turquoise Marilyn

In **Figure 7**, we see the top 3 scatter plots, which are based on the color distribution of the turquoise shot of Marilyn Monroe. On the far left plot, we have Blue VS. Red. In the center plot, we have Green VS. Red; in the far right plot, we have Blue vs. Green. In this case, we see a high prominence of a blueish color closer to green for most of the plot. This likely represents the turquoise background. On the bottom 3 scatter plot, we see the same but with the grayscale based on the prominence of the color not graphed. That means that on the far left plot, the darker the pixel, the more presence of the Green color. In the center plot, the darker the pixel, the more presence of blue. And on the far right plot, we see that the darker the pixel is, the more presence of red color. In this case, the Green color seems to be high prominence on the left plot. On the far right plot, we also see a high prominence of red color. Surprisingly we do not see as much blue in the center as we would have expected to have a bigger contribution to the painting. Thus, Green and Red are the most prominent colors in the painting.

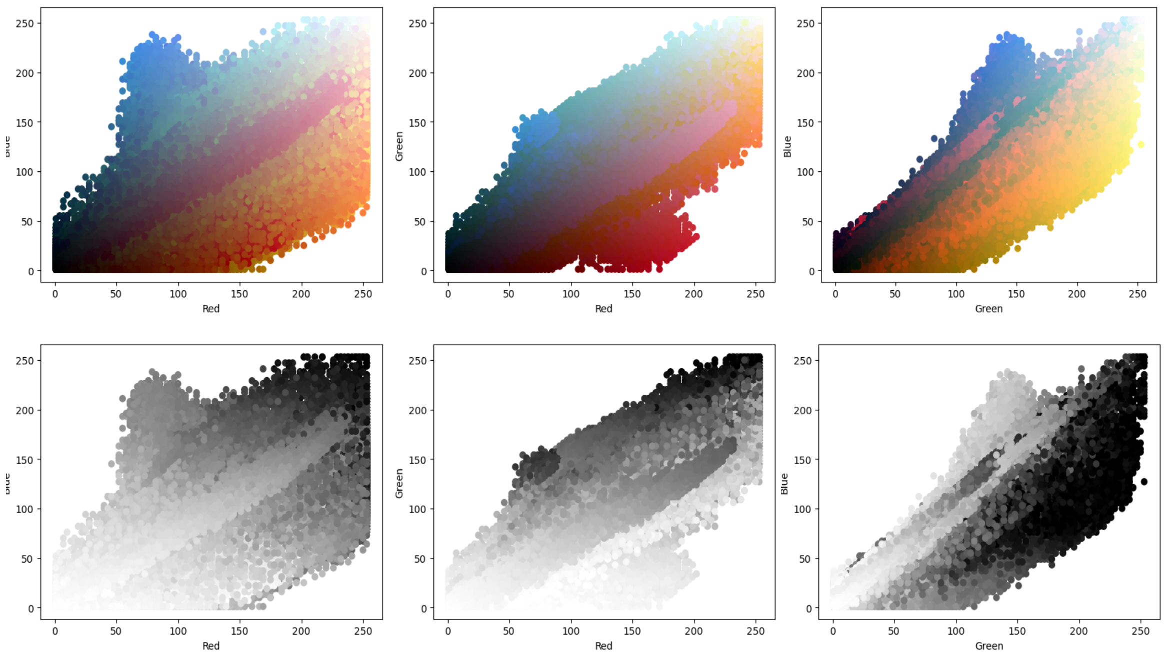

Figure 8. The RGB space occupied by the pixels for the entire image of Blue Marilyn

In **Figure 8**, we see that the top 3 plots are based on the scatter plot for the distribution of RGB color for the Blue shot of Marilyn Monroe. On the far left scatter plot, we have Blue VS. Red. In the center plot, we have Green VS. Red; in the far right plot, we have Blue vs. Green. In this case, we see a high proportion of blue in the three plots and a significant proportion of red. In the bottom 3 plots, we have the grayscale for the color that is not present. On the far left plot, the darker the pixel, the more the presence of Green color. In the center plot, the darker the pixel, the more the presence of blue, and we see that the darker the pixel, the more red color on the far right plot. In this case, we can see that in the left plot, there is a higher presence of Green color on the whole image. The far-right plot indicates an increased presence of red color in the painting. This is interesting because the background color would suggest that there should be more blue, but that is not the case. Overall, there is a contradiction between what we expect to be a higher presence of green color rather than green.

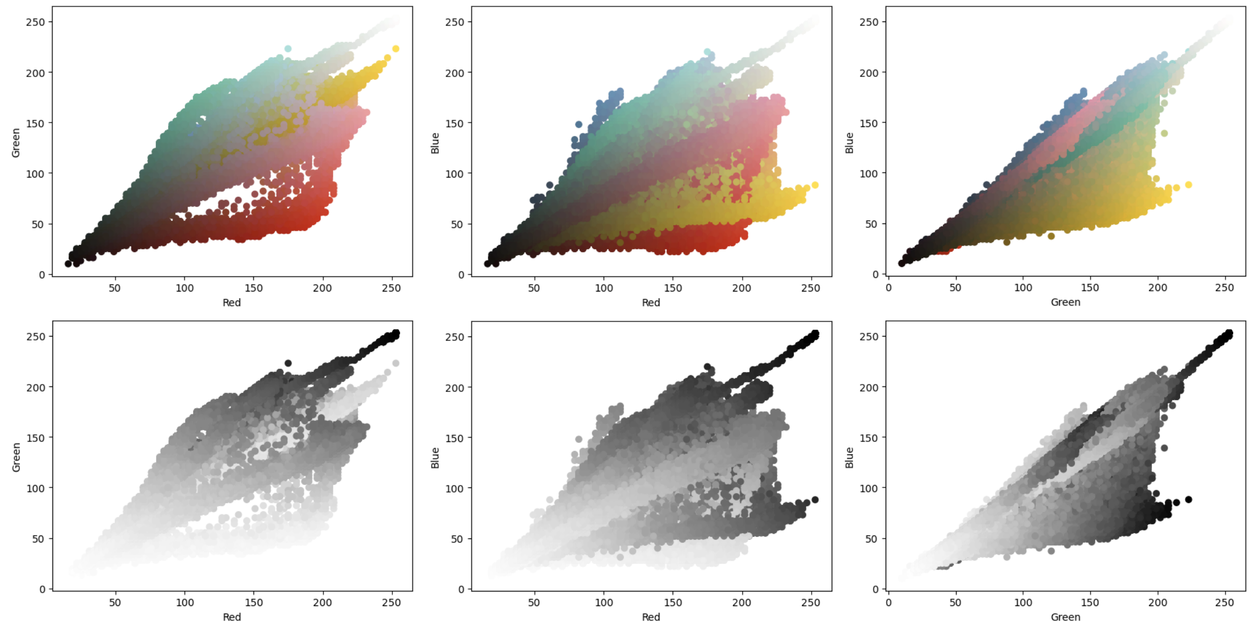

Figure 9. The RGB space occupied by the pixels for the entire image of EggBlue Marilyn

In **Figure 9**, The top three plots depict the distribution of RGB colors in the Eggblue shot of Marilyn Monroe using scatter plots. The scatter plot on the far left represents the relationship between Green and Red colors, while the center plot displays the correlation between Blue and Red. On the right side, the plot demonstrates the connection between Blue and Green. Remarkably, these plots reveal an equal distribution of all colors, with each color concentrated around the linear distribution pattern, indicating a correlation among them.

Moving to the bottom three scatter plots, they exhibit the same positioning of pixel dots in grayscale, indicating the dominance of the unrepresented color. That means that on the far left plot, the darker the pixel is, the more presence of the Green color. In the center plot, the darker the pixel is the more the presence of blue. And on the far right plot, we see that the darker the pixel is, the more the presence of red. In this instance, we can observe that as the values of the other two colors gradually increase, the corresponding pixel color also darkens gradually. No specific color dominates any particular sections of the image. This observation indicates that as the values of the other two colors rise, the overall color becomes darker, implying a simultaneous increase in the third color value. These findings suggest this image has the highest correlation among these 5 images, consistent with the correlation between colors found in the relative conditional entropy section.

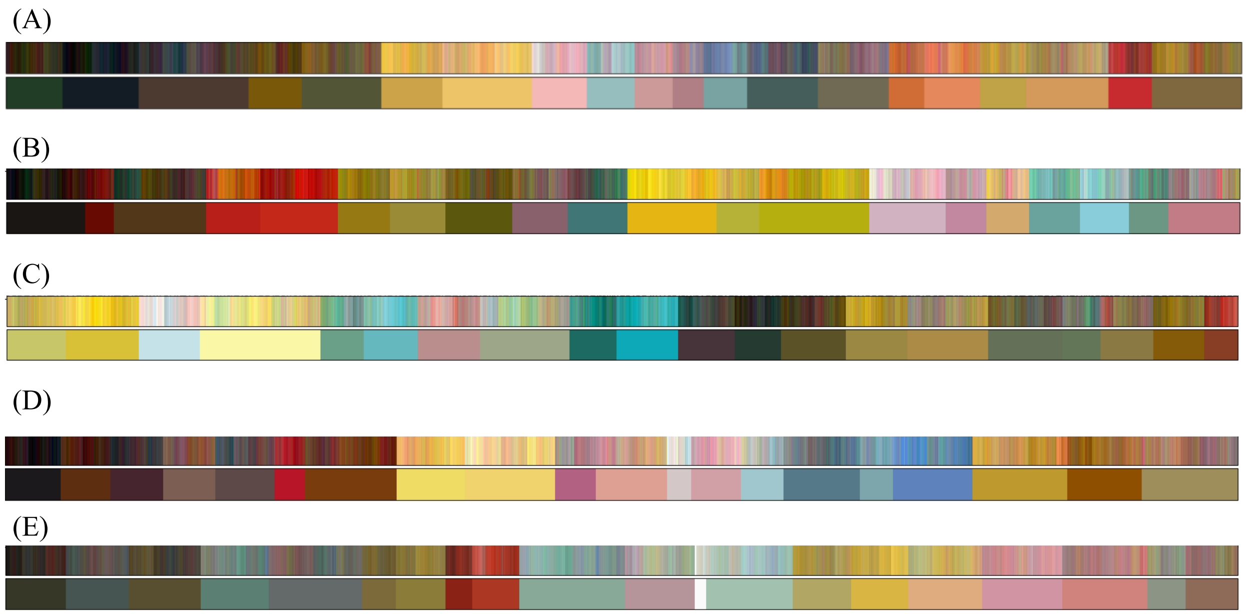

Figure 10. Distribution of Colors for 5 Images: **(A)** Orange Marilyn **(B)** Red Marilyn **(C)** Turquoise Marilyn **(D)** Blue Marilyn **(E)** Eggblue Marilyn

In plot A, we can see the hierarchical clustering of color based on 20 different bins of color. We decided to use 20 classifications to find the different colors that compose the major ones, such as the orange background, the pink face, or the yellow hair. Even though the pink color on the face or the background color would be all the same color, each of these major colors has various shades. Orange can be classified into three different clusters; the same can be applied to the hair having 3 different classifications that can be explored more.

In plot B, we find the hierarchy clustering for the Red shot of Marilyn Monroe. Again, with it being divided between 20 clusters, we find it interesting that its red background is divided into 4 different clusters indicating that the red background is not only one solid color but a range of different ones. We also see that the pink color is divided between 4 clusters, but they have a different coloration than our previous piece. This indicates that the pink palette used in the orange shot of Marilyn Monroe is not the same as the one used in the red one, even though it would appear otherwise. However, we see that the yellow of the hair color differs from the previous one, but this was not as big of a difference.

In plot C, we find the clustering for the turquoise background of Marilyn Monroe, again dividing it between 20 different bins. In this case, we see a more obvious difference between this color and the previous ones, especially since the yellow for the hair color is divided between 4 clusters, and the distinction is more recognizable. We also find that the pink color is divided into only 1 cluster, which means that even though there is a large presence of pink on the face, it seems to be the same pink color for most of the face, which is the biggest deviation from the other pieces.

In plot D, we see the hierarchy clustering for the blue shot of Marilyn Monroe. We find that the pink color for the face is divided into 4 clusters with the greatest color difference from the previous one. The background also has 4 bins of color, indicating that the blue color has a broader range. Interestingly, we found 5 yellow clusters, where we see a higher notorious distinction between the bright yellow and the darker yellow of the hair.

In plot E, we find the plot for the egg blue shot of Marilyn Monroe. It's interesting that most of the cluster colors are darker colors. We found that the only color that appears to have a bright color is pink, as well as noting the first white cluster that appeared out of all the paintings.

V. Clustering based on ROI

We decided to crop the images into distinct layers based on extensive online research. In this regard, we opted to utilize the HSV color model instead of RGB. This choice was based on the belief that HSV aligns more closely with human color perception, as it better captures how people experience color than the RGB color space[2].

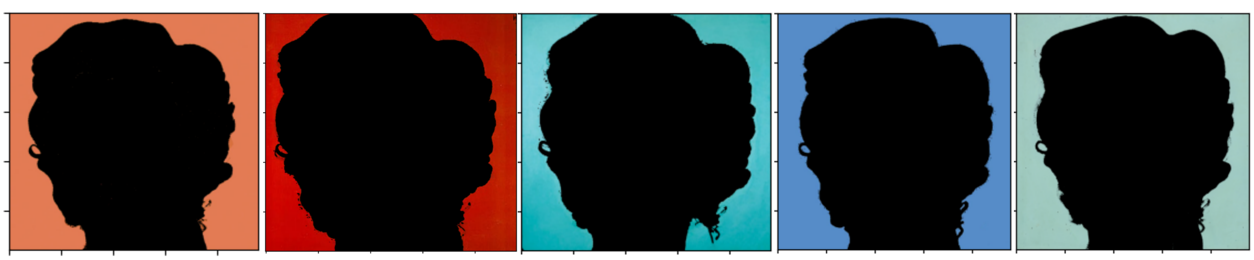

Figure 11. Region of Interest **(**ROI) **-** Backgrounds: From left to right are Orange Marilyn, Red Marilyn, Turquoise Marilyn, Blue Marilyn, and Eggblue Marilyn

We first select one or more small rectangles from the whole image to find HSV's max and min values (Hue, Saturation, and Value). By applying the max and min we saw from the rectangles, we can extract the whole background layer except for the middle image, where the color has the same color distribution as the background [9].

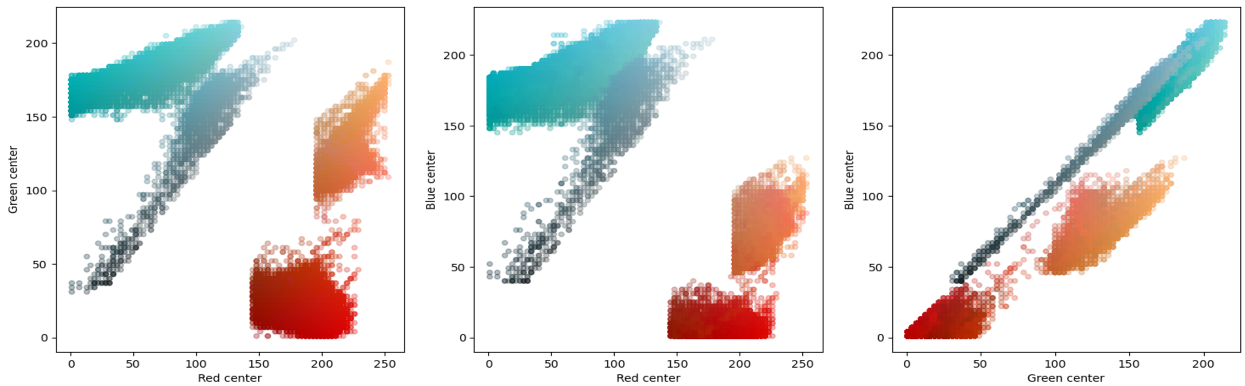

Figure 12. RGB Space for Backgrounds

In **Figure 12**, the color distribution of five backgrounds becomes visible in the 2D-RGB space. Within this space, we observe four distinguishable groups of colored dots. We hypothesize that the similarity between the background colors in images 3 and 4 is responsible for only four clusters in the scatterplot mentioned above.

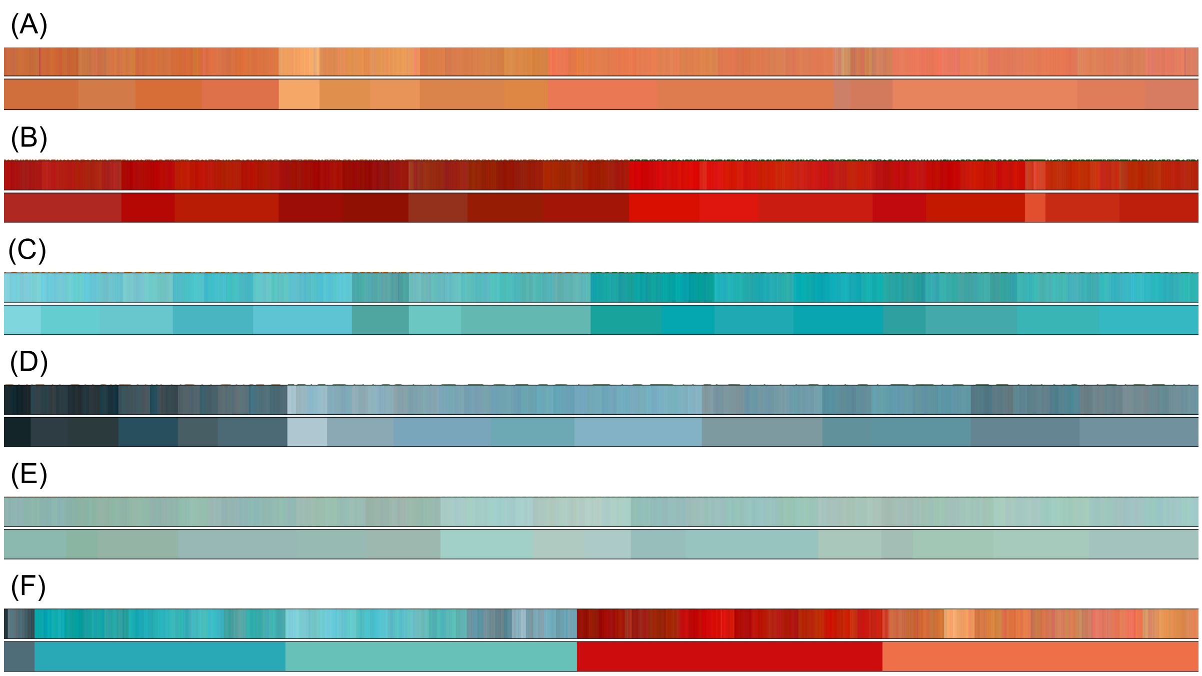

Figure 13. Distribution of Colors in the Background

Afterward, we utilized the hierarchical clustering method (See METHODS) to classify the background colors of each image to reveal more insights. Clusters (A), (B), (C), (D), and (E) represent the distinct color groups obtained from each image with 16 clustered groups, while Cluster F combines all background colors from the five images. We are trying to identify the color groups from 5 images using 5 clusters. Notably, the hierarchical clustering method successfully yielded five separate clusters corresponding to the background colors extracted from the different images. However, it becomes challenging to determine all the color groups for Turquoise Marilyn due to the significant blending of its colors with Blue Marilyn.

Figure 9. Region of Interest **(**ROI) **-** Hair: From left to right are Orange Marilyn, Red Marilyn, Turquoise Marilyn, Blue Marilyn, and Eggblue Marilyn

Using the same approach, we also tried to identify the hair as an individual part of the image. And we can observe that some hairs are brighter while others are faded into black or blurry.

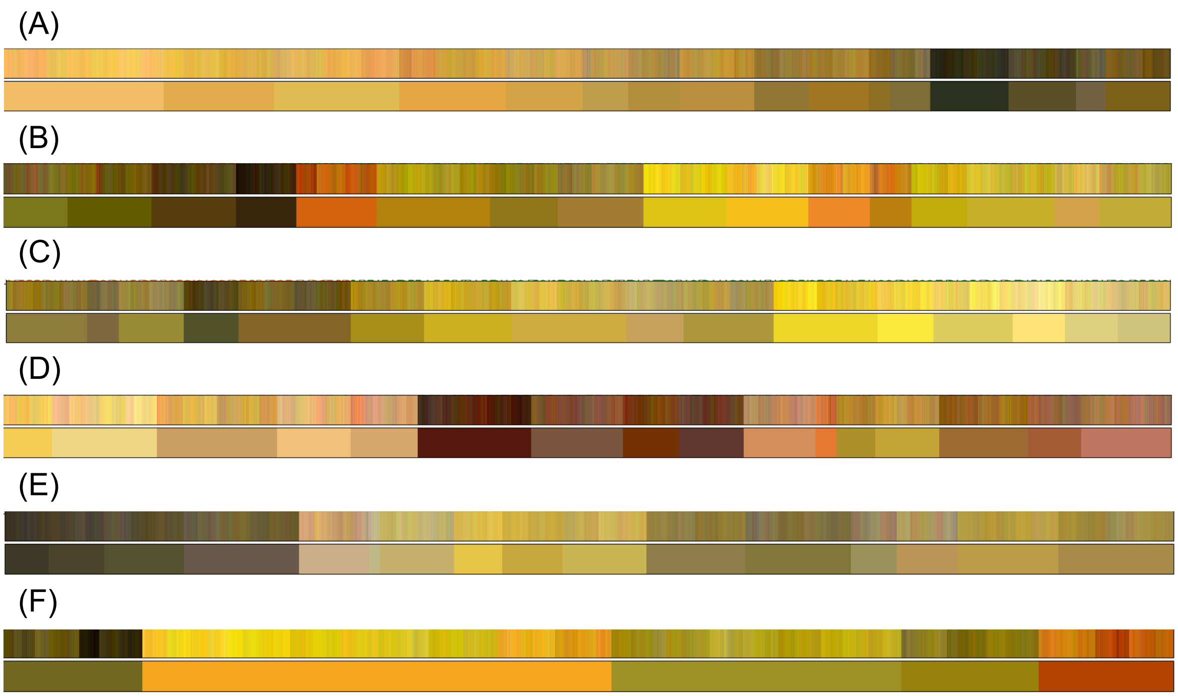

Figure 10. Distribution of Colors in the Hair

Following that, we utilized hierarchical clustering to categorize the extracted hair colors from each image to uncover additional insights. The resultant clusters, A, B, C, D, and E represent distinct color groups obtained from each image, resulting in 16 unique clustered groups. Cluster B exhibits the highest prevalence of orange tones across all images, while cluster D indicates that Blue Marilyn showcases the greatest presence of brownish hair color. Additionally, cluster C is characterized by a significant cluster of vibrant yellow color, as evidenced in the Turquoise Marilyn image.

Conversely, cluster F encompasses all hair colors observed in the five images. We are trying to identify the color groups from 5 images using 5 clusters. Our findings suggest that employing a mere five general cluster groups presents challenges in achieving effective differentiation. The clustering process primarily relied on distinguishing between dark and bright colors.

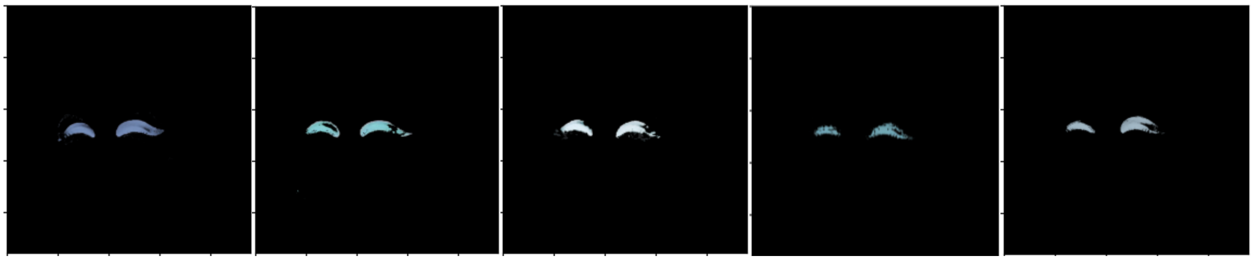

Figure 11. Region of Interest (ROI) - Eyeshadow: From left to right are Orange Marilyn, Red Marilyn, Turquoise Marilyn, Blue Marilyn, and Eggblue Marilyn

In **Figure 11**, we extracted the eye shadows from the entirety of the five images. Nevertheless, it proved to be challenging to isolate the complete eye shadow area from the entire image due to the inclusion of numerous dark areas. This made it particularly difficult to accurately crop the entire eye shadow portion, especially in the case of image 4. We can visualize so much noise in the image 4 eye shadow compared to the other images.

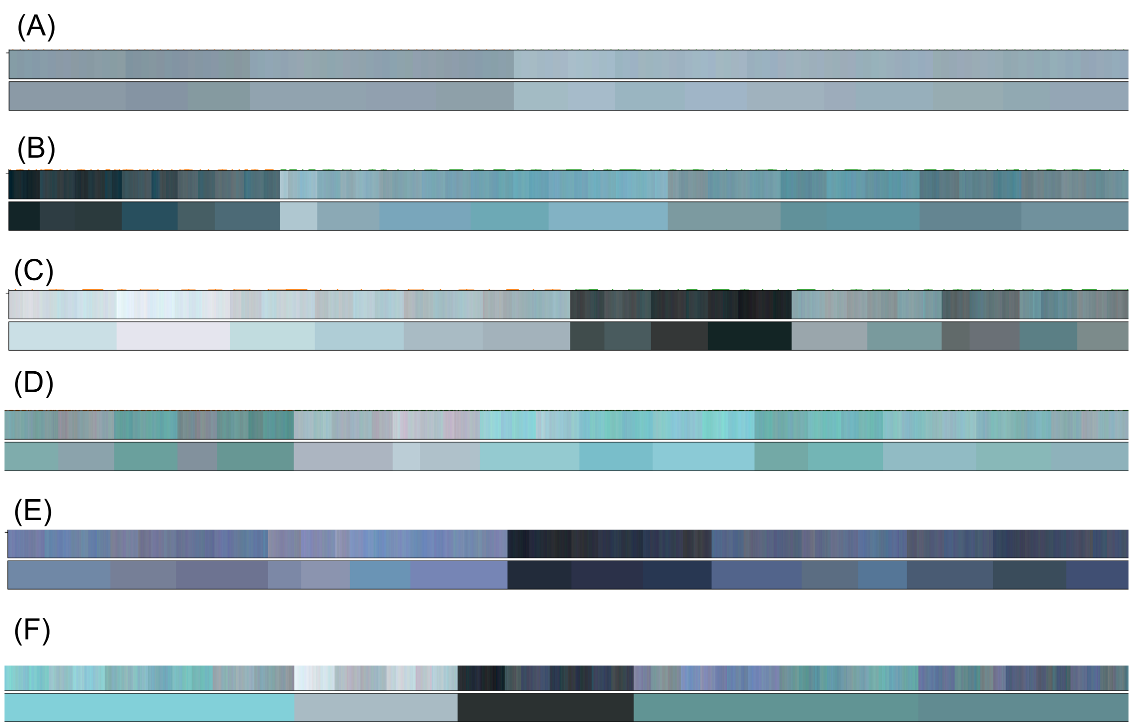

Figure 12. Distribution of Colors in the Eyeshadows

In this instance, we employed the same clustering approach to analyze eye shadow images. Clusters A, B, C, D, and E represent distinct color groups obtained from each eye shadow image, resulting in 16 clustered groups. Cluster F, on the other hand, combines all the eye shadow colors from the five images. We are trying to identify the color groups from 5 images using 5 clusters. Notably, Cluster A indicates that the Orange Marilyn eye shadow prominently features a grayish blue color, which is also present in Clusters B, C, and D. Cluster E, on the other hand, exhibits a blue shade closer to purple. Clusters B, C, and E all exhibit dark areas, indicating successful extraction of the dark regions of the eye shadows in those particular images.

However, we encountered a similar challenge with the hair analysis. The eye shadows displayed similar color tones, and Cluster F predominantly grouped the colors based on their brightness rather than distinguishing them by individual images. As a result, effectively differentiating the colors from each image using the hierarchical clustering (HC) method proved challenging.

Figure 13. Region of Interest **(**ROI) **-** Face: From left to right are Orange Marilyn, Red Marilyn, Turquoise Marilyn, Blue Marilyn, and Eggblue Marilyn.

In **Figure 13**, we extracted the face from the entirety of the five images. We notice that Turquoise Marilyn and Eggblue Marilyn have some noise in their faces for images. We think it might be because some shadow exists on the Turquoise and Eggblue Marilyn faces. The black dot between the eyebrows on Blue Marilyn results from a gunshot wound that wasn't repaired with the correct color.

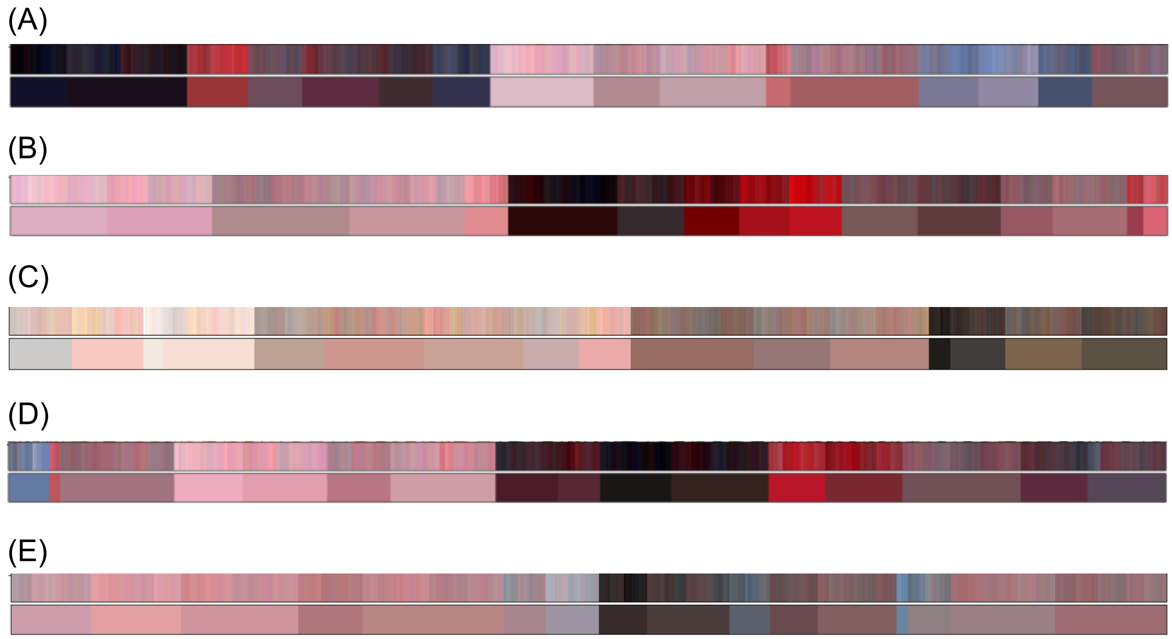

Figure 14 Distribution of Colors in the Face.

Looking at the overall result for our skin hierarchy clustering, we find that plots A, B, and D have similar clustering results, while C and E have more similar results. Plot C shows that the green color appears to be really bright. We see the same result for plot E but with a darker color. Thus, the composition of pink for C and D is made of a similar amount of color. Looking at **Figure 13**, we see that some face extraction does not completely remove the eyeshadow for the far left face. We also see that we lose some of the colors for the center face and the far right face. This impacts image clustering, such as with Plot A, where we could not remove some of the eyeshadow on the face, which is indicated by the blue cluster. There is a blue cluster in plot D, indicating that the face near the far right also has the presence of eyeshadow. Finally, the red cluster in plots A, B, and D are based on the lipstick, which was also not removed. Based on this knowledge, we would like to infer that if we could remove those elements from the image, we would have a more consistent and similar color between all three.

Nonetheless, the result brings up the idea that at a normal appearance, it seems as if the faces of each art piece are different, but they contain a similar composition of pink colors. Due to needing help to extract the face correctly, we found the clustering for all combined data sets hard to understand since there was no clear reason or explanation for some of the clustering results. We decided not to proceed with those results.

VII. REPAIR GUNSHOT OF IMAGE

The story of Shot Marilyns is set in the fall of 1964. The iconic pop artist Andy Warhol had just completed a series of five captivating Monroe paintings in his studio. Little did he know that an encounter with performance artist Dorothy Podber would forever change the fate of these artworks. When Dorothy asked to "shoot" the paintings, initially believing she meant to photograph them, Warhol agreed to Dorothy's request. However, to his astonishment, she pulled out a gun and shot the stacked canvases, piercing holes in several of them. Although the damaged pieces were subsequently repaired, they became known as the "Shot Marilyns" [5].

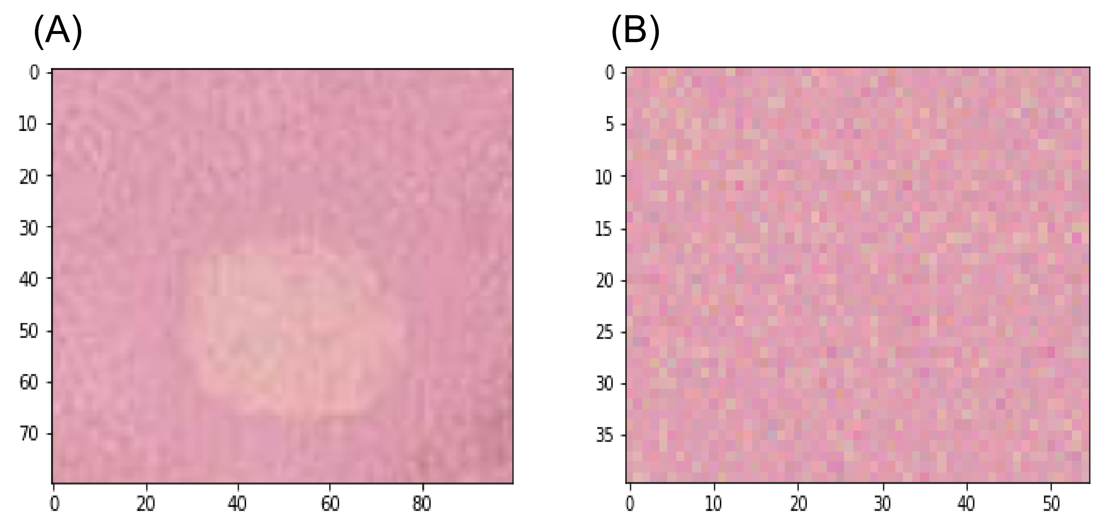

Figure 15. Region of Interest (ROI) - GUNSHOT AREA : **(A)** gunshot area before randomly sampling RGB, **(B)** gunshot area after randomly sampling RGB

While efforts were made to restore most of the damaged pieces, it was discovered that Blue Marilyn still had an evident gunshot trace between her eyebrows. Determined to achieve an authentic repair, our restoration process began by capturing the area affected, referred to as image (A). We cropped and closed the image to focus on the affected area and to ensure a natural outcome. We randomly sampled the RGB values within this region, replicating the original distribution for each color channel. The result of our work is shown on the right, labeled as the image (B).

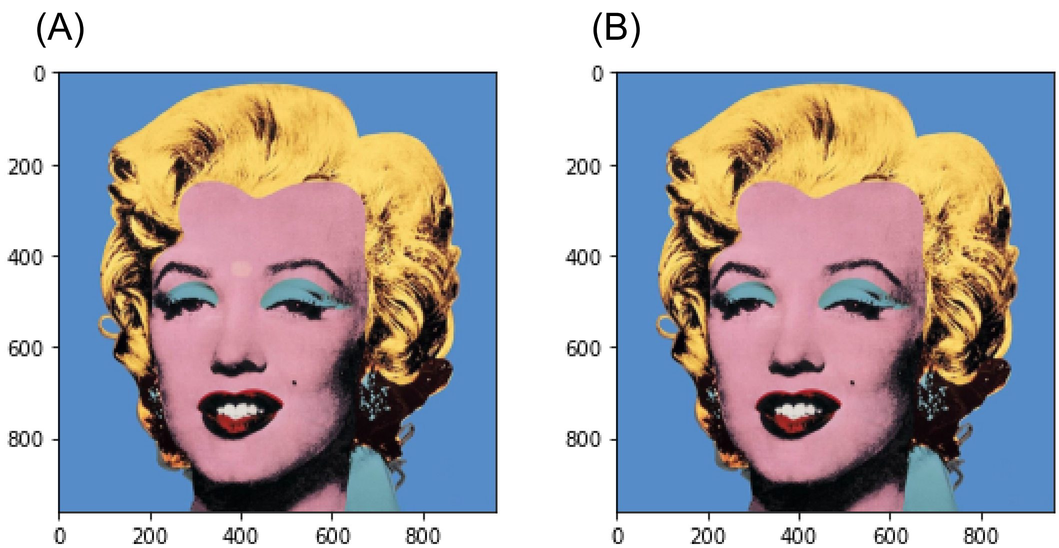

Figure 16. The comparison between the original image and repaired image: (A) Original image, (B) Repaired image

By random sampling RGB within the gunshot area, the outcome, represented by an image (B), presents a flawless repair. Upon observation, it's shown that the previously damaged area could blend in with the surrounding artwork. Through random sampling RGB within the gunshot area based on its distribution, we repaired the damaged area of the image to restore what was before the "shot" of the Marilyn images.

VII. DISCUSSION

In this project, we tried to split each part of the image, such as the background, hair, eyeshadow, and face, because we wanted to compare the different regions of images. By defining the maximum and minion range of color, we can optimize our extraction for our Region of Interest (ROI), allowing us to get the most out of the color range for each image part. However, to identify the regions of interest (ROIs), we must attempt multiple iterations of manually inputting color ranges. This process relies on the subjective selection of representative rectangles, which can result in poor outcomes if they fail to represent the ROIs accurately. It is particularly challenging to differentiate between similar color elements, such as the red background and red lips or blue eyeshadows and earrings. As a result, we often had to use an additional mask to cover some aspects with the same color. Moreover, accuracy can be affected by variations in lighting and image characteristics, and this method is only suitable for specific scenarios. While it excels at extracting the background, it may not produce satisfactory results when applied to other situations.

We used the RGB space occupied by the pixels for each entire image in the 2D scatterplot to see each image the color distributed in 3D; the advantage of using RGB scatterplots in the report is that they offer a visual representation of color distribution, allowing for a quick understanding of the overall color composition and variations within an image. Additionally, comparing scatterplots of different images enables the analysis of color patterns, trends, and specific characteristics, facilitating discussions on similarities and differences in color distributions across multiple images. Nevertheless, the limitations of using RGB scatterplots in the report include the need for capturing important color attributes like hue, saturation, and brightness, which limits the depth of color analysis. The complexity of interpreting scatterplots can arise when dealing with various color distributions or intricate patterns, requiring an additional explanation for accurate understanding. Additionally, the focus on individual pixel colors and the absence of contextual factors ignore the relationships between colors and the broader context of lighting, image content, and artistic intent, which may impact the overall interpretation of the artwork's color representation. Also, we find that clustering has some issues due to the number of pixels in the data set. Since the total painting has almost a million observations in the data set, none of our personal computers could run the code for that amount of observations. That is the reason we decided to break down the data set using 3x3x3 cubic space, which allows us to plot the RGB color for the centroid of each cube. In the plotting of the RGB space in 3D, we find it hard to understand the output; therefore, we finally break down the picture into 3 plots that can be more easily understood.

XIII. CONCLUSION

After our exhaustive examination of each piece's color composition and distribution, we find it interesting that the expectation we first had of each piece's color composition is different from the result we got. In these five images, we compared RGB relationships by distribution and relative conditional entropy, and the development we found here was also aligned with the style of pop art paintings. We found the "Eggblue Marilyn" had a higher relationship between RGB, which gives the feeling of harmony to the audience, and the "Red Marilyn" had few relationships between RGB, which gives the sense of a firm contract.

We have certain expectations for the art pieces in general, such as expecting the background for each piece to be one unique color or the scarf to have the same color for all of the pieces. The hair and face were different for each painting. This assumption can be made by just looking at the pieces and comparing them. However, the expectations still needed to be met to some extent. From our general graph distribution, we got that most pieces have a higher red composition, whereas blue seems to be the second most prominent color in most pictures.

Interestingly, we would have expected blue to have a higher prominence of color in general because the background color is mainly near the blue color distribution. However, still, red is more prominent overall. Once we start working through the specific ROI, such as the background, eyeshadow, hair, and face, we see that what would look like the same color is a spectrum of them, such as the background that can be divided into 16 different shades ors. This result is specifically shocking since, in simple view, the color seems solid for all of them. We see this result with all of our ROI, where there is not a single color that encompasses each specific area but a range of them. However, for the face, hair, and eyeshadow, the same colors are prominent throughout the five pieces, which at first deceived us into appearing different from each other. After repairing the gunshot, we find that the gunshot does not seem to affect our picture distribution much. Ultimately, the overall face, including the hair, skin, and eyeshadow color, is the same for all five pieces.

Similarly, there is a large number of different color shades that encompass each piece's background. Therefore, we conclude that mainly relying on an observational interpretation of an artwork is insufficient for fully appreciating the proper depth and intricacy that color brings to art. The human eyes are deceiving and process color in a way that can be much of it. This is the main reason we decide to break down each layer of each piece to understand better what color brings to the complexity of the art and how we can depend on our understanding of it.

Reference

[1] B. Schmidt, “A visual critique of Warhol’s shot Sage Blue Marilyn, 1964,” The Interior Review, https://www.theinteriorreview.com/story/2022/5/10/critically-assessing-warhols-shot-sage-blue-marilyn (accessed Jun. 9, 2023).

[2] Color (image processing toolbox), http://www.ece.northwestern.edu/local-apps/matlabhelp/toolbox/images/color11.html#:~:text=The%20HSV%20color%20space%20 (accessed Jun. 9, 2023).

[3] “Sold for $195 million, Andy Warhol’s ‘shot sage blue Marilyn sets new auction record,” Los Angeles Times, https://www.latimes.com/entertainment-arts/story/2022-05-09/andy-warhols-shot-sage-blue-marilyn-sets-new-auction-record (accessed Jun. 9, 2023).

[4] S. Liao, P. Koehl, J. Schultens, and F. Hsieh, “The geometry of colors iMarilynsgh’s sunflowers - heritage science,” SpringerOpen, https://heritagesciencejournal.springeropen.com/articles/10.1186/s40494-021-00608-y (accessed Jun. 10, 2023).

[5] “Andy Warhol’s ‘Shot Marilyns,’” Andy Warhol’s “Shot marilyns,” https://news.masterworksfineart.com/2019/11/26/andy-warhols-shot-marilyns (accessed Jun. 10, 2023).

[6] VishwaV, “Conditional entropy calculation in python, H(Y: X),” Data Science Stack Exchange, https://datascience.stackexchange.com/questions/58565/conditional-entropy-calculation-in-python-hyx (accessed Jun. 10, 2023).

[7] M. Foley, Unsupervised learning with HCA, https://rstudio-pubs-static.s3.amazonaws.com/489493_bf427cf192f04108bc4020f4837e8147.html (accessed Jun. 10, 2023).

[8] Andy Warhol’s Marilyn Prints, https://www.webexhibits.org/colorart/marilyns.html (accessed Jun. 10, 2023).

[9] Real Python, “Image segmentation using color spaces in opencv + py Real Python, https://realpython.com/python-opencv-color-spaces/ (accessed Jun. 10, 2023).

[10] Teacher Assistant Abdul-Hakeem Omotayo (2023, June 1). STA 160 Practice in Statistics Data Science Office Hour. The University of California, Davis.