

Supplementary Information

Large-Scale Quantitative Analysis of Painting Arts

Daniel Kim, Seung-Woo Son, and Hawoong Jeong

Correspondence to hjeong@kaist.edu and sonswoo@hanyang.ac.kr

Contents

Supplementary Information Text	2 - 3
Figures S1 to S6	4 - 9
Table S1	10

Supplementary Information Text

Binomial distribution of random sampling colors. Let us suppose that we are trying to fill N pixels using an unbiased Galton board machine with n bins, where n corresponds to the number of colors we have. We are choosing a color by dropping a ball from the top of the machine for every N steps. After picking random colors N times, the usage frequency of a color is given as follows:

$$f\left(k; N, p = \frac{1}{n}\right) = \Pr(X = k) = \binom{N}{k} \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{N-k},$$

where n is the number of colors in the palette, k is the number of chooses for a certain color ($0 \leq k \leq N$), and $p = 1/n$ because of the unbiased color-choosing assumption. That is the probability function of the binomial distribution. The cumulative distribution function of the binomial distribution can be derived in terms of the regularized incomplete beta function:

$$F\left(k; N, p = \frac{1}{n}\right) = \Pr(X \leq k) = I_{1-1/n}(N-k, k+1) = (N-k) \binom{N}{k} \int_0^{1-1/n} t^{N-k-1} (1-t)^k dt$$

The inverse function of the complementary cumulative distribution function is the rank plot.

Image size distribution. As shown in Fig. S1, over 94% of images are larger than $700 \times 700 = 490K$ pixels and the largest one is 1350×1533 . Therefore, the quality of images in the dataset is good enough to perform a statistical analysis. Furthermore, in order to discuss differences between paintings and photographs, two more datasets are gathered for hyper-realism and photographs, for each dataset, the largest one is 2974×1954 and 2070×2700 respectively.

Counting the number of pixels for each color. The usage of each color was treated as the fraction of a color in a painting $p_i = n_i / N_{pixel}$, where n_i is the number of pixels for each color i and N_{pixel} is the total number of pixels in the painting. Next, the ratio p_i of each color i for all images in a period were summed up and, the sum was normalized by the number of images in the period N_{image} . That is how we obtain the average usage ratio of a color in the painting for each period. We plotted the color usage as a function of its rank in descending order for each period.

Changing color quality and image size. When measuring box counting dimensions, we changed the number of possible colors in the RGB color space five times: $256^3 \rightarrow 128^3 \rightarrow 64^3 \rightarrow 32^3 \rightarrow 16^3 \rightarrow 8^3$, which corresponds to a reduction of a 24 bit scale to a 9 bit scale. We observed that varying the number of possible colors in the RGB color space affects a small change on the shape of the RCD for each period (see Fig. S2). If the size of an image is reduced by one quarter, four pixels shrink into one pixel and one representative color is chosen proportional to its frequency, which is similar to the spatial renormalization process. Using five interpolation kernels included in MATLAB R2011b (box-shaped, triangular, cubic, Lanczos-2, and Lanczos-3), we reduced the size of each image. Respectively, however, the result is not that different for each method.

Effects of light interference. When we convert raw paintings into a digital form, some errors in the converting process are inevitable. As an example, digitizing Georges Seurat's "*A Sunday Afternoon on the Island of La Grande Jatte*", painted by a special skill called *pointillism*, can have an error due to a unique painting technique. In pointillism, a painter uses only small distinct pure colors to draw an image, but various colors can be expressed by the *light interferences* of small dots of primary colors. The number of practically used colors is known to be few in his painting. As shown in Fig. S3, however, RCD of a Seurat's work shows that various colors can be derived from *light interference*.

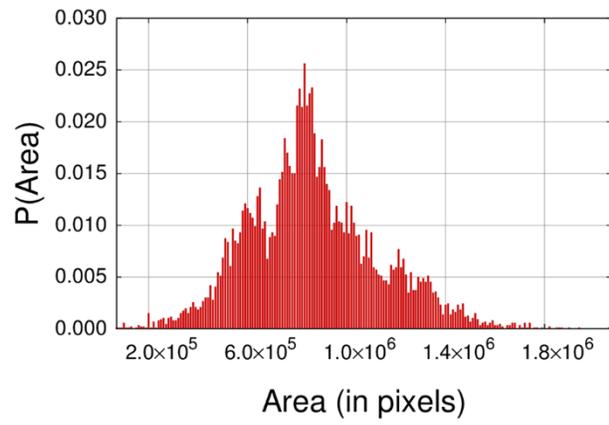


Figure S1 | The probability distribution of the image size of digitized European paintings. Area is defined by width \times height.

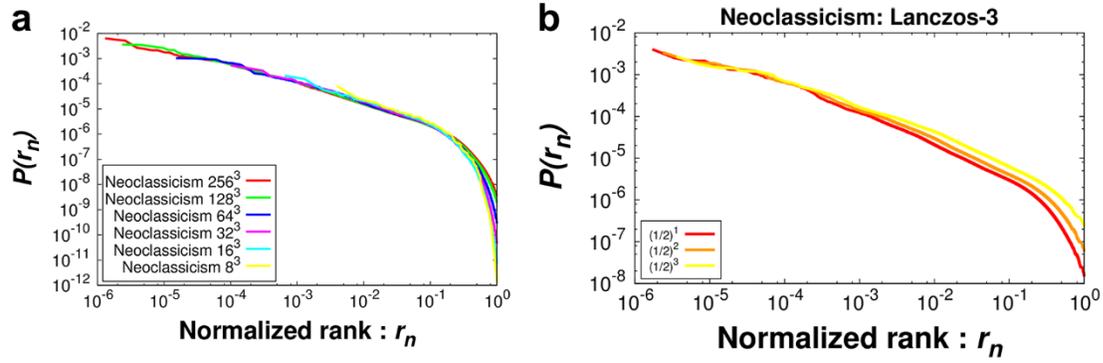


Figure S2 | Effect of changing the color quality and the size of images on the rank-ordered distribution. (a) The RCD of the Neoclassicism period with a different color quality from 24 bit (256^3 colors) to 9 bit (8^3 colors). (b) Resizing effect, $(1/2)^n$ means that the width and height of resized paintings in the Neoclassicism period are $\text{width} \times (1/2)^n$ and $\text{height} \times (1/2)^n$. For example, in the case of $(1/2)^3$, the size of image is approximately 100×100 . Resizing images and lowering images' color quality in these ranges gives insignificant changes. The horizontal axis denotes a rank of a color and the vertical axis indicates probability of using a color in the Neoclassicism period. In this case, each rank value is normalized by a maximum rank for correct comparison.

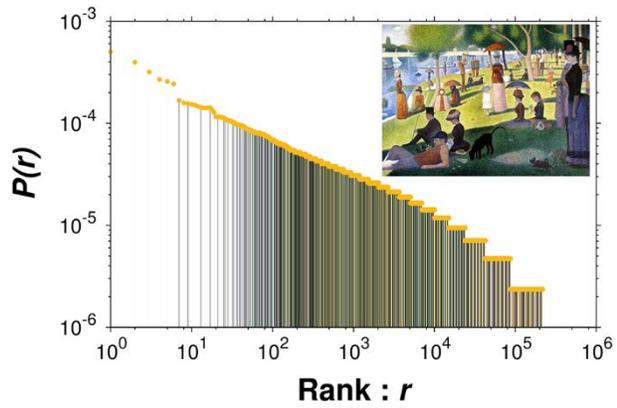


Figure S3 | The effect of the light interference. The inset image is the work of Georges Seurat (1859-1891), “*A Sunday Afternoon on the Island of La Grande Jatte*”. The RCD of the painting shows that the number of colors in digitized image is much larger than the number of colors used in the original image painted by the pointillism technique with limited colors.

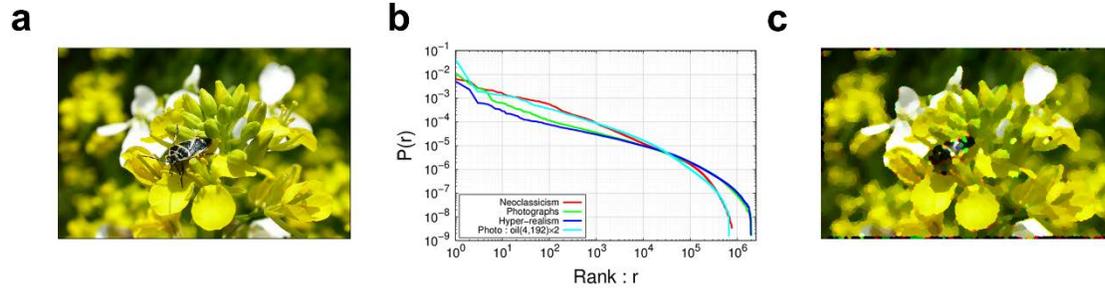


Figure S4 | Photograph vs. painting. (a) An example image is shown. This was taken by Seung-Woo Son. (b) Tail section of the RCD of neoclassicism paintings (Neoclassicism) is clearly different from that of photographs (Amateur) and hyper-realism paintings. However, tail of the RCD of the photographs with an oil painting filter applied is quite close to that of Neoclassicism. Oil (4,192) indicates two control parameters of the oil painting filter, range=4 and level=192. In order to transform a photograph into an oil painting like in this image, both the range and the level are required which correspond to the smearing radius and intensity. (c) An example of a photograph converted by the oil painting filter.

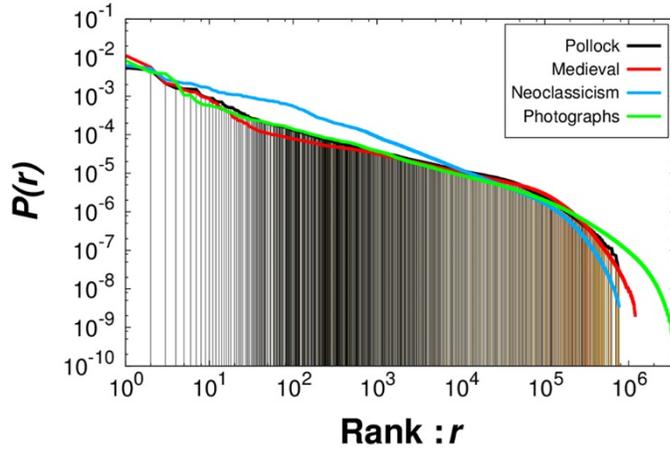


Figure S5 | Rank plot of color usage. The rank-ordered color-usage distribution (RCD) of Pollock's paintings is similar to that of medieval paintings (Medieval) and photographs except for the longer tails of the medieval and photographs, which reflects artistic details. The number of colors used in Pollock's drip paintings (Pollock) is smaller than those of Medieval and Photographs and is rather similar to that of neoclassicism paintings (Neoclassicism). However, the range of rank $[10, 10^4]$ on the RCD of Neoclassicism is different from Medieval, Photographs, and Pollock's drip paintings. Likewise, in the analysis of brightness surface, the average surface roughness of Pollock is smaller than that of medieval paintings, while in fractal analysis, the box counting dimension of Pollock is similar to Medieval.

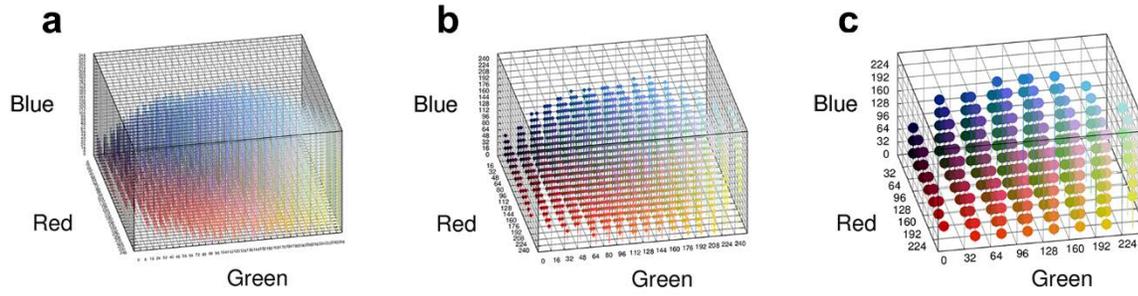


Figure S6 | Measuring fractal dimension. We partition the whole RGB space into cubic boxes of a given size ϵ . When considering box counting dimension, we need to count the number of non-empty boxes $N(\epsilon)$ out of $(256/\epsilon)^3$ boxes by varying the side length ϵ of a box. **(a)** $\epsilon=8$, **(b)** $\epsilon=16$, and **(c)** $\epsilon=32$. The slope in the $\log N(\epsilon)$ curve gives the fractal dimension. Scattered colors shown in this example are from the images of the neoclassicism period.

Table S1. The number of collected and analyzed paintings for each art historical period.

Period or movement	The number of collected paintings	The number of analyzed paintings
Medieval	351	331
Early renaissance	1,289	995
Northern renaissance	1,150	1,047
High renaissance	756	677
Mannerism	1,041	911
Baroque	3,537	3,287
Rococo	392	360
Neoclassicism	181	163
Romanticism	737	663
Realism	150	146
Hyper-realism	106	106
Photograph (National Geographic)	3,777	3,713
(Amateur)	112	112
Total	13,579	12,511