Structural stability and thermodynamics of artistic composition

San To Chan[a)] and Eliot Fried
Mechanics and Materials Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan

(Dated: 23 August 2024)

Inspired by the way that digital artists zoom out of the canvas to assess the clarity of their works, we introduce a conceptually simple yet effective metric for quantifying the visual clarity of digital images. This metric contrasts original images with progressively "melted" counterparts, produced by randomly flipping adjacent pixel pairs. It measures the presence of stable structures, assigning the value zero to completely uniform or random images and finite values for those with discernible patterns. This metric respects the color diversity of the original image and withstands image compression and color quantization. Its suitability for diverse image analysis problems is demonstrated through its effective evaluation of visual textures, the identification of structural transitions in physical systems like the Potts and XY models, and its consistency with color theory in digital arts. This allows us to demonstrate that colors in visual art function as a state variable, akin to the spin configuration in magnets, driving artistic designs to transition between states having distinct visual stability. When combined with the Shannon entropy, which quantifies color diversity, the structural stability metric can serve as a navigation tool for artists to explore pathways on the complex structural information landscape toward the completion of their artwork. As a practical demonstration, we apply our metric to refine and optimize an emote design for a video game. The structural stability metric emerges as a versatile tool for extracting nuanced structural information from digital images, enhancing decision-making and data analysis across scientific and creative domains.

Keywords: digital art, image analysis, entropy, complexity, stability

I. INTRODUCTION

Artists illustrate by gradually applying brushstrokes to plain canvases. However, placing too many structures, like lines and colors, onto a canvas would eventually result in a grayish-to-black appearance, resembling yet another plain canvas. It is the harmony and contrast between the different structures that make a painting visually informative to its viewer. Hence, the concept of structural information of an image is cyclic[1]: If the amount of structure is to be quantified using a metric, the metric must return the same value for the two extremes of complete order and disorder.

In the quest to quantify the structural information of digital images, multiple approaches have been proposed. Image compression techniques[2–4] evaluate how much an image can be compressed without structural degradation—the more structures an image contains, the less the image can be compressed. Approaches employing second-order information measures, such as delentropy[5,6] and permutation entropy[7–9], analyze the distribution of local pixel intensity arrangements, or ordinal patterns, across an image; a wider variety of these patterns leads to larger values of the entropy. Such metrics are often paired with statistical complexity, which assesses the similarity of the distribution of the observed ordinal patterns to an idealized uniform distribution, where all patterns are equally likely. Statistical complexity vanishes at the extremes of complete order and disorder. Renormalization-based methods[10,11] compare the original image with its coarse-grained versions generated through multi-step spatial averaging; images maintaining consistent structural details across scales show greater overlap between the original and simplified versions.

The methods mentioned above are supported by credible theoretical foundations such as statistical physics and information theory. Whether they are suitable for analyzing paintings is, however, another matter. For instance, metrics based on image compression largely depend on the image resolution, which is not a property of the structures contained in a painting. Delentropy and permutation entropy consider only the intensity of the pixel but not its color, even though colors are central to most artworks. Also, the number of pixels used to define an ordinal pattern is somewhat arbitrary. Metrics based on renormalization require averaging of adjacent pixel values. For styles such as pixel art, which heavily relies on patterns such as lines and dithering, this would destroy most structural information in one single step, giving a very low overlap between the original image and its coarse-grained counterpart. These observations underscore a significant gap in current methodologies when applied to the structural analysis of art. There is a compelling need for a new metric that not only encompasses the broad spectra of shape and color used in art but also respects the integrity of various artistic styles.

STRUCTURAL INFORMATION OF DIGITAL IMAGES

To ensure their depicted subject remains recognizable across different viewing scales, such as on a smartphone screen, digital artists frequently zoom out from their canvas and inspect their drawings. This practice effectively simulates a blurring of details, hinting at a way to quantify the amount of visually stable structures within digital images. As an illustrative example, Fig. 1A shows a pixel art image of width $L_x = 555$ and height $L_y = 290$ featuring Hololive Production talent Himemori Luna and her knight mascots on a

[a)]Corresponding author san.chan@oist.jp

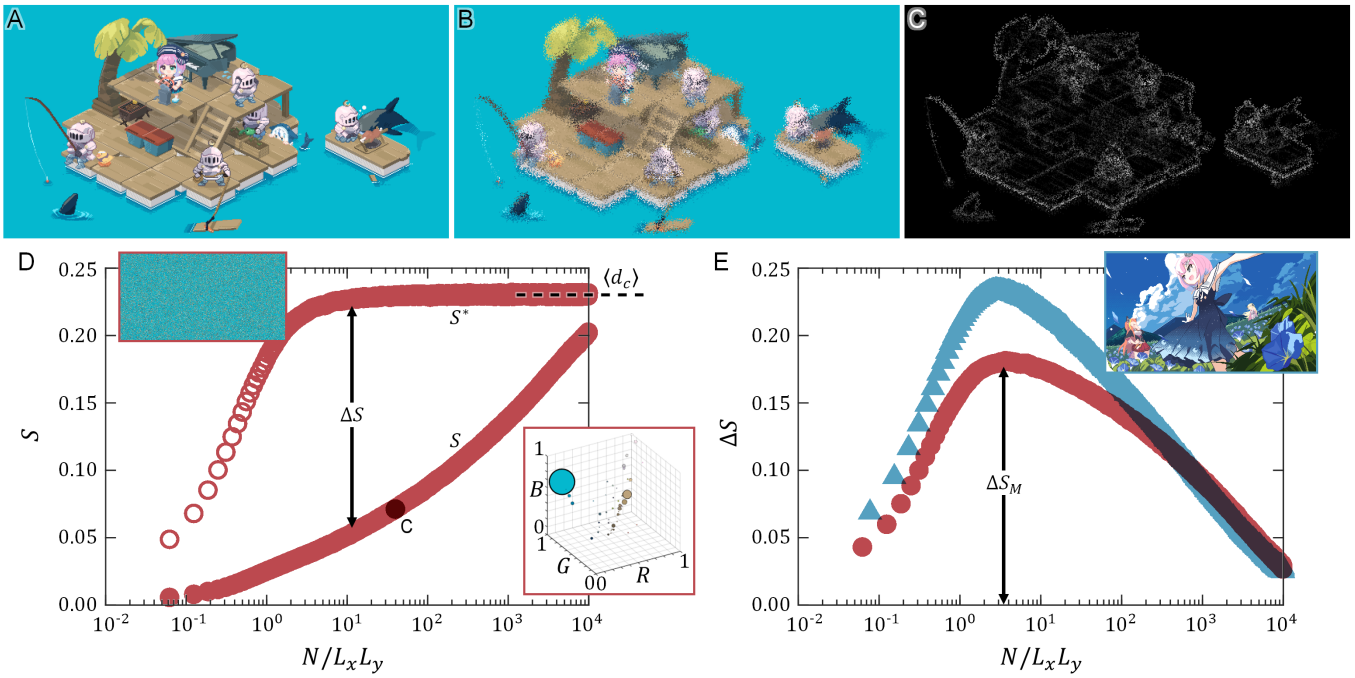

FIG. 1. Structural stability analysis of digital images. (A) The original pixel art of width $L_x = 555$ and height $L_y = 290$ depicting Hololive Production talent Himemori Luna and her knight mascots on a raft. (B) The "melted" version of the artwork after 650,000 iterations of random adjacent pixel pair flipping, demonstrating the degradation of structures with the subjects obscured, leaving the shape of the raft visually recognizable. (C) A grayscale image, referred to as the color distance map, representing structural stability. Black regions are robust against structural change due to pixel-flipping, while white regions are susceptible. (D) The calculated structural degradation score S (filled symbols) of the original artwork as a function of the number of flipped pixel pairs N, normalized by the image size L_xL_y, plotted with the score S^* (empty symbols) of a completely shuffled version of the original artwork (inset image). The filled symbol labeled C corresponds to the structural degradation score of the grayscale image in C. The difference $\Delta S = S^* - S$ quantifies the amount of stable structures in the original artwork relative to that in the shuffled version. The inset three-dimensional scatter plot shows the distribution of the fifty most frequent colors within the RGB space for the artwork. The axis labels R, G, and B represent the color intensity in the red, green, and blue channels, respectively. Each data point represents a unique color. The diameter of each data point is proportional to the probability of that particular color within the artwork. The mean distance between all pairs of data points, scaled by their sizes, is the mean color distance $\langle d_c \rangle$. (E) The differences ΔS of two visually distinct images as functions of N/L_xL_y, which are maximized consistently for $N/L_xL_y \sim 1$, leading to our definition of the structural stability metric $\Delta S_M = \max(\Delta S)$, which is independent of N. Circles represent the artwork depicted in A, and triangles represent the artwork in the inset. Artworks by Ronin (X/Twitter ID: @zeth_total). Used with permission.

raft. By flipping a random pair of adjacent pixels per iteration, we progressively degrade the structures in the image. After 650,000 iterations, the resultant "melted" version of the painting, shown in Fig. 1B, primarily obscures the subjects, leaving the overall shape of the raft still visually recognizable.

The structural degradation is quantified by calculating the normalized Euclidean color distance d_{ij} between each pixel at horizontal positions $i = 1, 2, ..., L_x$ and vertical positions $j = 1, 2, ..., L_y$ in the original image and its melted version within the RGB (red-green-blue) space:

$$d_{ij} = \frac{1}{\sqrt{3}} \sqrt{(\Delta R_{ij})^2 + (\Delta G_{ij})^2 + (\Delta B_{ij})^2}. \tag{1}$$

Here, ΔR_{ij}, ΔG_{ij}, and ΔB_{ij} are the intensity differences at position (i, j) between the two images in the red, green, and blue channels, respectively. This computation yields a grayscale image, referred to as the color distance map, for each iteration of pixel-flipping, as illustrated in Fig. 1C. In the color distance map, black areas ($d_{ij} = 0$) correspond to regions highly resistant to pixel-flipping, indicating visually stable structures such as the predominantly uniform background color. In contrast, white areas ($d_{ij} = 1$) highlight regions that are particularly susceptible to pixel-flipping, such as the subjects.

The structural degradation score S is obtained by summing the non-zero pixel values in the color distance map and normalizing by the image size L_xL_y:

$$S = \frac{1}{L_xL_y} \sum_{i,j} d_{ij}. \tag{2}$$

This score logarithmically increases with the number of flipped pixel pairs N, as shown in Fig. 1D (filled symbols). For a completely shuffled version of the original image (Fig. 1D, inset image), all large-scale structures exhibit deterioration. As each small-scale structure is equally prone to pixel-flipping, the score S^* rises more rapidly than S, reaching a plateau for $N/L_xL_y \sim 1$ (Fig. 1D, empty symbols), where on average all pixels are flipped once, essentially generating another shuffled image. Granted that shuffling disrupts

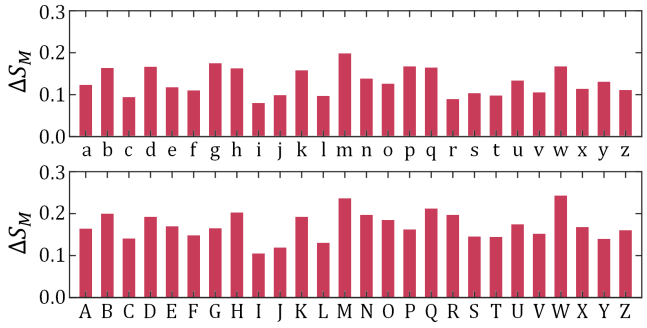

FIG. 2. The structural stability metric ΔS_M for different upper and lowercase English letters. The font used is Cambria Math.

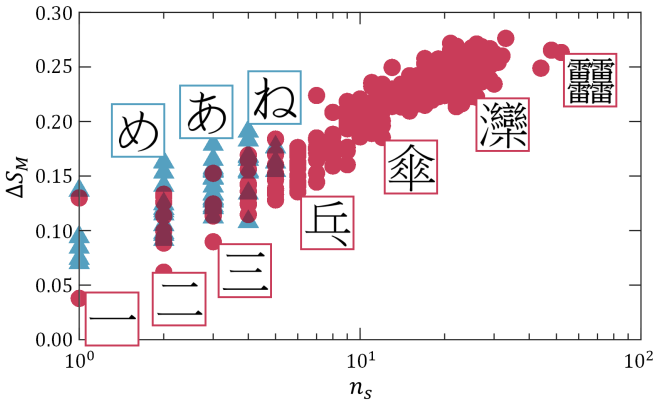

FIG. 3. The structural stability metric ΔS_M as a function of the number of strokes for different Chinese characters (circles) and Japanese hiragana (triangles). The font used is MingLiU.

all spatial color relationships, the average of d^*_{ij}, namely S^*, with sufficient pixel pairs between the two shuffled images converges statistically to the mean color distance $\langle d_c \rangle$ found when considering all possible color combinations in the color histogram of the original image (Fig. 1D, inset plot), scaled by their frequencies. This allows us to derive an expression relating the plateau value of S^* to $\langle d_c \rangle$. Given K distinct values d_k of color distance obtainable from the color histogram, with corresponding probability p_k, the plateau value of S^* is determined by

$$S^* = \frac{1}{L_x L_y} \sum_{i,j} d^*_{ij} = \sum_k \frac{n_k}{L_x L_y} d_k = \sum_k p_k d_k = \langle d_c \rangle, \qquad (3)$$

where n_k is the number of pixels with value d_k in the color distance map.

The fully melted, shuffled version of the original image represents a benchmark of complete disorder, allowing us to tally the number of stable structures, that is, the degree of structural stability in the original image, relative to this disorder state, quantified by $\Delta S = S^* - S$. For sufficiently large $N/L_x L_y > 1$, ΔS can be expressed as

$$\Delta S = \sum_{i,j} \Delta S_{ij} = \frac{1}{L_x L_y} \sum_{i,j} \langle d_c \rangle - d_{ij}, \qquad (4)$$

which represents the deviation of the local color distance d_{ij} at pixel position (i, j) from the mean color distance $\langle d_c \rangle$. Regions having similar colors in the original image have small values of d_{ij}, leading to a higher value of local structural stability ΔS_{ij} and hence contribute more to the global structural stability ΔS of the image. Therefore, large values of ΔS indicate that many regions have color distances lower than the expected average in the state of complete disorder, highlighting the presence of order, or stable structures, in the image.

As illustrated in Fig. 1E, ΔS is consistently maximized for $N/L_x L_y \sim 1$, even for visually distinct images. This leads us to define a structural stability metric $\Delta S_M = \max(\Delta S)$, which is independent of N. Notably, ΔS_M is zero for both a plain color image, where $\langle d_c \rangle = d_{ij} = 0$, and a completely shuffled image, where $S^* = S$, and thus properly incorporates the cyclic nature of structural information. Given a set of images sharing the same color histogram, the maximum possible value of ΔS_M for those images would be $\langle d_c \rangle$ according to [4]. The upper bound of ΔS_M can be obtained by considering an image filled with colors of equal proportions located at the eight vertices in the RGB cube (cf. Fig. 1D, inset plot), yielding $\langle d_c \rangle = (\sqrt{3} + \sqrt{6} + 1)/8 \approx 0.648$. For an image filled with colors of equal proportions located at only two of the non-adjacent vertices in the RGB cube, for instance, black and white, $\langle d_c \rangle = 0.5$

Our formulation of ΔS_M is directly tied to the color histogram of the original image, whose randomness is quantified by the Shannon entropy $H = -\sum_c p_c \log_2 p_c$, where p_c is the probability that a pixel exhibits a color vector $\mathbf{c}$ in the RGB space[12]. Let $\mathbf{c}_\alpha$ and $\mathbf{c}_\beta$ denote two color vectors in the color histogram, with respective color probabilities p_α and p_β. The mean distance of all colors in the color histogram can then be expressed as

$$\langle d_c \rangle = \sum_{\alpha,\beta} p_\alpha p_\beta |\mathbf{c}_\alpha - \mathbf{c}_\beta|. \qquad (5)$$

To fairly compare the structural information of two images, it is therefore necessary to consider both ΔS_M and H. In essence, this mirrors the approach of quantifying structural information via permutation entropy and statistical complexity[7–9], except that our method focuses on how colors, instead of ordinal patterns, are distributed in an image. As an important side note, to measure H, the color space must first be discretized to allow for the construction of a color histogram. Hence, all images used in the current study are pre-processed using minimum variance color quantization[13], limiting the number of colors to 2^{16} and, thus, the maximum possible value of H to 16.

VALIDATION OF THE STRUCTURAL STABILITY METRIC

To empirically test the reliability of the structural stability metric ΔS_M, we now examine a variety of test cases chosen to assess whether the metric corresponds with our innate perceptions regarding structural information.

For ΔS_M to be a valid structural measure, it must assign similar values to visually similar symbols. Fig. 2 shows the

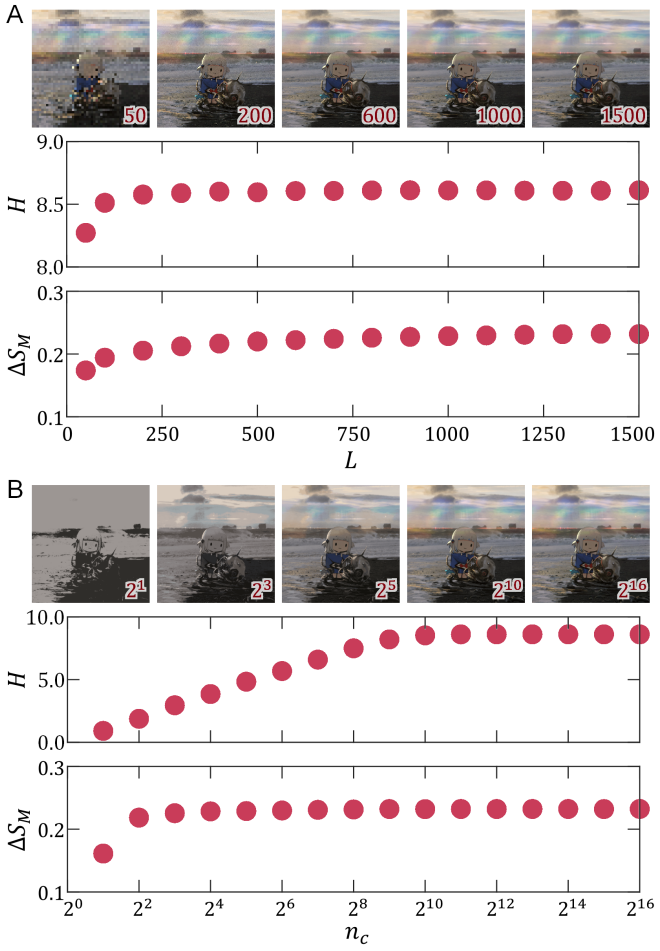

FIG. 4. Sensitivity to image compression and color quantization. (A) Square illustrations of different sizes L, depicting Hololive Production talent Gawr Gura. The image with $L = 1500$ is the original artwork; images with smaller L are obtained by compressing the original one using bicubic interpolation. The plots below show the corresponding Shannon entropy H and structural stability metric ΔS_M plotted as functions of L. (B) The same depictions of Gawr Gura with different numbers of colors n_c obtained by processing the original artwork using minimum variance color quantization. The plots below show H and ΔS_M as functions of n_c. Artwork by Advarcher (X/Twitter ID: @Anonamos_701). Used with permission.

values of ΔS_M for different upper and lowercase English letters. We observe that ΔS_M attains similar values for visually similar letters such as "b", "d", "h", "p", and "q", "l" and "I", as well as "M" and "W". Conversely, ΔS_M distinctly differentiates between letters with more pronounced structural differences, such as "l", "m", and "n". Fig. 3 shows ΔS_M for different Chinese characters (circles) and Japanese hiragana (triangles), with numbers of strokes n_s. As n_s increases, the structural features of the symbols increase, leading to larger values of ΔS_M. Japanese hiragana generally has smaller values of ΔS_M in comparison to Chinese characters with large values of n_s. This agrees with the historical understanding that the designs of hiragana were simplified from Man'yōgana, Chinese characters that were used to transcribe Japanese phonographically[14,15]. It is worth noting that ΔS_M of the characters appears to be a monotonically increasing function of n_s, indicating that the design of fonts might be optimized for their legibility. Also, the nominally logarithmic increase of ΔS_M in response to an increasing n_s resonates with Fechner's law in experimental psychology, which postulates that the relationship between stimulus and perception is logarithmic[16,17]. These observations highlight the sensitivity of the structural stability metric ΔS_M to the visual similarity within a set of symbols, effectively mapping our perception of structural information onto a quantifiable scale.

As a second requirement, ΔS_M must be insensitive to operations such as image compression and color quantization, granted that the dominant structures depicted in the image are visually preserved. Fig. 4A shows several square painterly illustrations of different sizes L. The Shannon entropy H and the structural stability metric ΔS_M as functions of L are also shown. The illustration with $L = 1500$ is the original artwork; the compressed versions with smaller values of L are obtained from the original one by bicubic interpolation[18]. The Shannon entropy and the structural stability metric are both highly robust against image compression, with H and ΔS_M showing roughly the same values even though the original artwork is reduced more than ten times in size. Fig. 4B shows the same artwork but with the number of colors n_c reduced by minimum variance color quantization[13]. The value of H decreases as n_c is reduced to 2^{10}, approaching a value smaller than 1 as n_c is further reduced to 2. On the other hand, ΔS_M retains its value down to $n_c = 2^2$, demonstrating a robust resistance against color quantization. Of course, the above observations cannot apply to images like pixel art. Nonetheless, they underscore the applicability of ΔS_M across a wide variety of digital images, particularly those where color gradations are essential to conveying the structures depicted in an image.

STRUCTURAL ANALYSIS OF DIGITAL IMAGES

The Shannon entropy H and structural stability metric ΔS_M together quantify the amount of structural information that is contained in a digital image. In what follows, we showcase how the two metrics can be applied together in a number of settings.

Kylberg–Sintorn rotation dataset

Given that H and ΔS_M can quantify the diversity of colors and their spatial distribution in a digital image, we expect that these metrics might effectively differentiate between the textures in the Kylberg–Sintorn rotation dataset, which includes one-hundred 122×122 images for each of its diverse texture classes[19]. Fig. 5A displays examples, arranged top to bottom, left to right: cane sugar, canvas, couscous, fabric1, fabric2, fabric3, flaxseed, knitwear, lentils, oatmeal, pearl sugar, rice, rug, rye flakes, seeds, sprinkles, tile, towel, wheat, and wool. Some of the images, such as fabric1 and towel, which share a fibrous texture, and lentils and oatmeal, which share a speck-

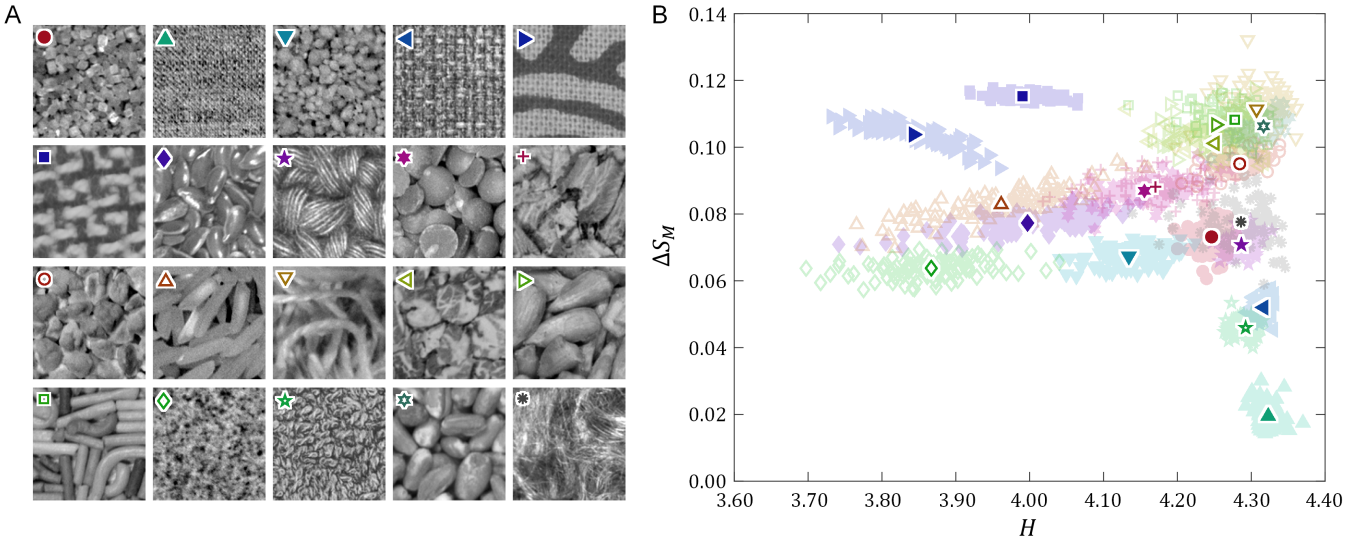

FIG. 5. Structural analysis of the Kylberg–Sintorn rotation dataset. (A) Grayscale images of dimension 122×122 depicting 20 textures from the dataset. The images are arranged in spatial order (top to bottom, left to right): Cane sugar, canvas, couscous, fabric1, fabric2, fabric3, flaxseed, knitwear, lentils, oatmeal, pearl sugar, rice, rug, rye flakes, seeds, sprinkles, tile, towel, wheat, and wool. (B) Scatter plot mapping the 100 sample images for each of the 20 textures on the entropy-stability plane defined by the Shannon entropy H and the structural stability metric ΔS_M. Half-transparent symbols represent raw data. Darker symbols represent the mean values of the raw data.

led appearance, show visual resemblance when viewed from afar. On the other hand, images such as fabric2 and fabric3, due to their unique patterns, appear visually distinct from all other textures, even when viewed from afar.

In Fig. 5B, the scatter plot of H and ΔS_M illustrates the placement of different texture classes on the entropy-stability plane. Textures like fabric3 (filled rightward triangles), with its distinctly binary and expansive patterns, are positioned towards the top-left quadrant, indicating lower color diversity and higher visual stability. Meanwhile, textures like canvas (filled circles) that consist solely of fine details are positioned towards the bottom right, suggesting higher color diversity and lower visual stability. Textures like rug (empty downward triangles), rye flakes (empty leftward triangles), seeds (empty rightward triangles), sprinkles (empty squares), and wheat (empty hexagrams), with rather shiny surfaces and pronounced gaps, are situated in the top-right quadrant, reflecting a higher degree of structural complexity. Notably, the clustering of ΔS_M values around their mean for most texture classes underscores that the metric is indeed robust in distinguishing visually similar images. Our observations regarding the Kylberg–Sintorn rotation dataset affirm the effectiveness of H and ΔS_M in capturing the nuanced differences between textures, confirming their potential for robust textural analysis of digital images.

Potts model

A natural follow-up question concerns whether H and ΔS_M can be used to identify phase transition in physical systems, in particular, solely by quantifying the structural information contained in the images depicting the system configuration without any a priori knowledge about the system. To answer that, we next consider the four-state Potts model[20], also known as the Ashkin–Teller model[21], an extension of the Ising model[22] to four spin states.

Originating in statistical physics, the Potts model allows each site on a lattice to assume one of four distinct states interacting with its nearby sites. The Hamiltonian of the Potts model reads

$$E = -J \sum_{\langle i,j \rangle} \delta(\sigma_i, \sigma_j), \tag{6}$$

which is the energy of a particular spin configuration. The sum is taken over the adjacent neighbors $\langle i, j \rangle$ of all sites. The spin state of site i, σ_i, can take on a value of either $1, 2, 3$, or 4. The symbol $\delta(\sigma_i, \sigma_j)$ is the Kronecker delta, which equals unity if $\sigma_i = \sigma_j$ and zero otherwise. The parameter J is the coupling constant, which determines the strength of the interaction between neighboring sites.

When simulating the Potts model, there are two ways to update the spin states on the lattice. In Glauber dynamics[23], a new state is assigned to a randomly chosen lattice site at each time step. In Kawasaki dynamics[24,25], the states of a randomly chosen pair of adjacent lattice sites are exchanged at each time step. Both updating schemes effectively induce changes ΔE of the Hamiltonian. However, whereas Kawasaki dynamics preserve the amounts of each spin state, Glauber dynamics allow those amounts to change. Temperature T is incorporated into the model via the Metropolis–Hastings algorithm[26,27] for accepting a new spin configuration by two criteria: Energetically favorable changes ($\Delta E \leq 0$) are always accepted. For energetically unfavorable ones ($\Delta E > 0$), a probability $p = \exp(-\Delta E/kT)$ is calculated and a random number $0 \leq n \leq 1$ is drawn; if $n \leq p$, the new configuration is ac-

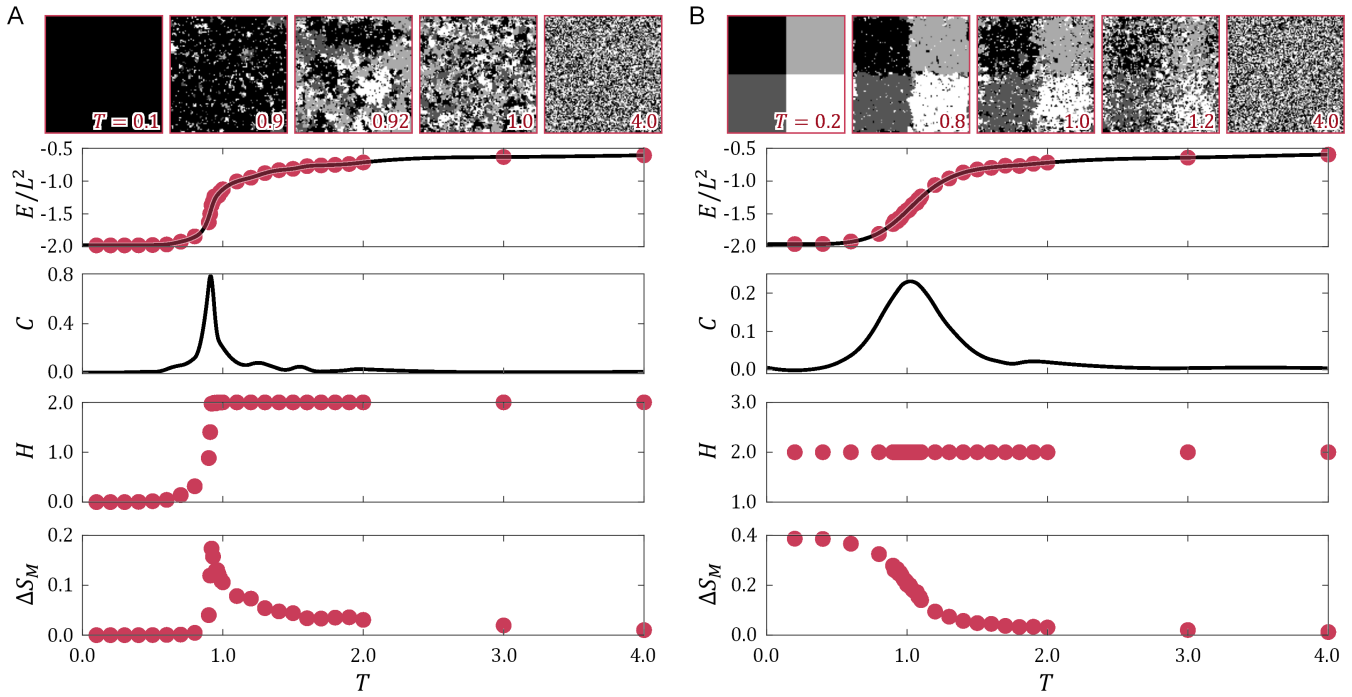

FIG. 6. Structural analysis of the four-state Potts model with $J = 1$ using the (A) Glauber and (B) Kawasaki dynamics. Top panels display spin configurations on a square lattice of size $L = 100$ at various temperatures T. Subsequent panels from top to bottom show the normalized Hamiltonian E/L^2, specific heat $C = (1/L^2)dE/dT$, Shannon entropy H, and structural stability metric ΔS_M, all as functions of T. Symbols represent simulation results. For E/L^2, the lines represent the cubic smoothing spline interpolation of the simulation results. For C, the lines represent the numerical derivative of the interpolation of E/L^2 with respect to T. To ensure equilibration, all data are obtained after 10^8 iterations.

cepted. The conversion factor k is the Boltzmann constant; to simplify the presentation in the current study, we set k to unity such that ΔE and T are both dimensionless.

Figure 6A shows the simulation results of the four-state Potts model via Glauber dynamics, a square lattice of size $L = 100$ with periodic boundary conditions and $J = 1$. To simulate the gradual degradation of structures, an initial condition of uniform spin states is assumed. At low temperatures such as $T = 0.1$, the probability of accepting an energetically unfavorable update is extremely low ($p \approx 0$), causing the spin configuration to remain in its initial uniform state. As T increases, new spin states, visualized as distinct colors in the image, gradually appear. At high temperatures such as $T = 4$, $p \approx 1$, most updates are accepted, leading to a noise-like appearance of the spin configuration. The normalized Hamiltonian E/L^2 increases as T increases. Notably, it shows a transition at $T = 0.92$, corresponding to a sharp peak of the specific heat $C = (1/L^2)dE/dT$ at the same T, in agreement with the seminal work by Binder[28]. As a side note, the slight oscillation in the specific heat curve for $1 < T < 2$ is due to our use of cubic smoothing splines[29] when fitting the data for E/L^2 and subsequent numerical differentiation needed to calculate C, prioritizing computational efficiency over precision.

The Shannon entropy H follows a trend similar to E/L^2; however, the transition appears more abrupt at $T = 0.92$. The value of H approaches 2, the maximum value for a four-state system, indicating that the four spin states become equally probable. Notably, the structural stability metric ΔS_M exhibits a sharp peak at $T = 0.92$, akin to C, indicating the emergence of stable structures in the system. This finding is significant. Traditional methods often employ C as an indicator of a phase transition. This quantity can be obtained by computing the numerical derivative of E with respect to T or by employing the fluctuation-dissipation theorem, which respectively entails curve fitting[29] or sampling large numbers of spin configurations[30] to average out thermal fluctuations. In sharp contrast, ΔS_M can identify the critical point of phase transition simply by analyzing single screenshots of the spin configurations captured at different T, complementing the traditional approaches.

Fig. 6B shows the simulation results for the same four-state Potts model but with Kawasaki dynamics. An initial condition with the four spin states occupying each quadrant of the lattice is used. As T is increased, structures on the lattice degrade; eventually, the spin configuration becomes noise-like as T becomes sufficiently high. The transition of E/L^2 and the peak of C appear less sharp compared to those depicted in Fig. 6A. This contrast is attributed to a unique property of Kawasaki dynamics. As any exchange of spins influences six adjacent lattice sites, ΔE is, on average, higher than that determined by Glauber dynamics, in which only four neighboring sites are affected. Due to the larger value of ΔE, the temperature T has a less pronounced effect on the acceptance probability. This hinders the system from exploring alternative configura-

tions, limiting energy fluctuations near the critical point and leading to a relatively blunt peak of the specific heat. Since spin states are conserved, the Shannon entropy H assumes a constant value of 2 for all T. However, the structural stability metric ΔS_M still exhibits a transition reflecting the phase transition, showcasing its capacity to identify structural changes in the system. Nonetheless, its trend now mirrors that of E/L^2 instead of C, indicating that the ability of ΔS_M to identify phase transition is context-dependent, which is common for order parameters[31].

Structural stability as energy difference

For the Potts model, when the Shannon entropy H is not conserved, the structural stability metric ΔS_M peaks at the critical temperature akin to the specific heat C (Fig. 6A). However, when H is conserved, ΔS_M mirrors the trend of the Hamiltonian E/L^2 (Fig. 6B). These observations can be explained if we consider the similarity between the structural degradation score S (cf. [2]) and the Hamiltonian, or energy, E (cf. [6]). Both expressions involve a quantitative comparison between the states of two entities and the summation of those quantities over all entities in the system. In [2], a comparison is made between the color vectors $\mathbf{c}$ of two pixels in the original image and its melted counterpart. Similarly, in [6], a comparison is made between the states σ of an adjacent pair of spins. This similarity suggests that S can be interpreted as a form of "structural energy" in the context of digital images. The normalized number of flipped pixel pairs, N/L_xL_y, can be viewed as a parameter functionally analogous to temperature T, controlling the extent of perturbation applied to the image. The score $S^* = \langle d_c \rangle$ (cf. [3]) for the completely shuffled version of the original image corresponds to the structural energy for $N/L_xL_y \gg 1$, where all structures are effectively melted. The expression [4], central to defining the structural stability metric ΔS_M, can be interpreted as the structural energy difference between the slightly perturbed ($N/L_xL_y \sim 1$) and completely randomized ($N/L_xL_y \gg 1$) states of an image. Hence, the similarity between the trends of ΔS_M (maximal structural energy difference), E (energy difference from a state of zero energy), and C (energy difference per unit temperature difference) in the Potts model is not a mere coincidence but is rooted in the similar formulations of the three quantities.

Classical XY model

Having established an energetic interpretation of ΔS_M, here, we consider the classical XY model[32–34], in which the spins can freely rotate in two dimensions, leading to continuous spin orientations. We encode the spin states using hues instead of intensities. This encoding strategy deliberately introduces a challenge that existing methods based on permutation entropy[7–9] cannot adequately address due to their reliance on ordinal relationships among pixel values. These modifications together offer a more demanding test for ΔS_M, requiring that it discerns the subtle structural changes in the system. The

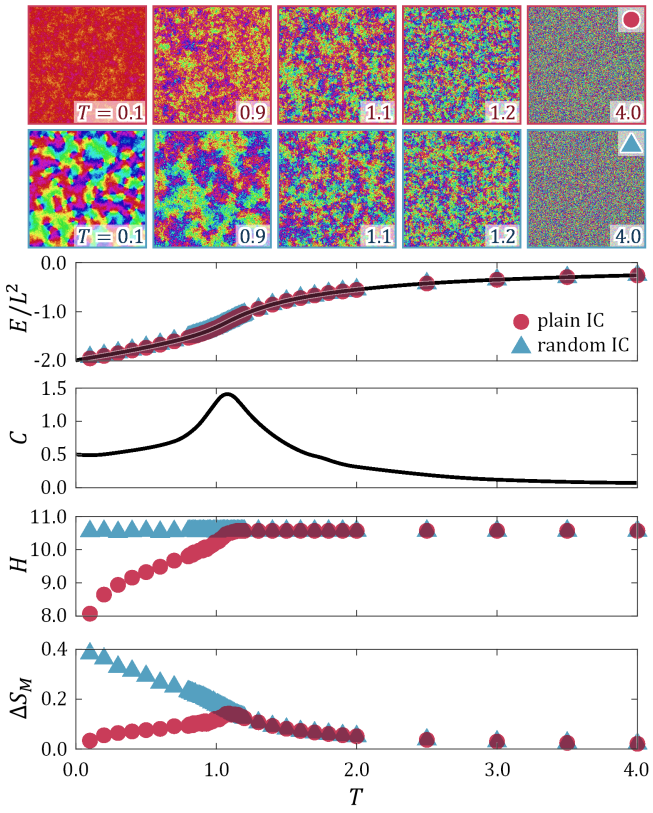

FIG. 7. Structural analysis of the classical XY model. The top two rows display spin configurations on a square lattice of size $L = 300$ at various temperatures T, assuming a plain (top) and a random (bottom) initial condition (IC). Subsequent panels from top to bottom show the normalized Hamiltonian E/L^2, specific heat C, Shannon entropy H, and structural stability metric ΔS_M as functions of T. For E/L^2, the lines represent the cubic smoothing spline interpolation of the simulation results. For C, the lines represent the numerical derivative of the interpolation of E/L^2 with respect to T. To ensure equilibration, all data are obtained after 10^8 iterations.

Hamiltonian of the XY model is

$$E = -J \sum_{\langle i,j \rangle} \cos(\theta_i - \theta_j), \tag{7}$$

where $\theta_i \in [0, 2\pi)$ is the spin orientation at site i. Notably, the XY model is known to undergo the Berezinskii–Kosterlitz–Thouless transition[32–34] marked by the unbinding of topological defects like vortex and antivortex pairs, as the temperature is raised above $T \approx 0.89$[35–41]. The interactions between vortices suppress energy fluctuations, shifting the peak of the specific heat to a higher temperature of $T \approx 1.02$–1.04[35–39].

Fig. 7 shows our simulation results of the XY model with $J = 1$ using a square lattice of size $L = 300$ with periodic boundary conditions. The spin configuration is updated by assigning a small change of angle $\Delta\theta \in [-\pi/2, \pi/2]$ to θ_i per iteration. Notably, for $T < 1.1$, the system distinctly depends on the initial spin configuration: A uniform initial state tends to retain its structural coherence; in contrast, a random initial state exhibits emergent clusters of varying hues, representing different spin orientations. For $T \geq 1.1$, these differences

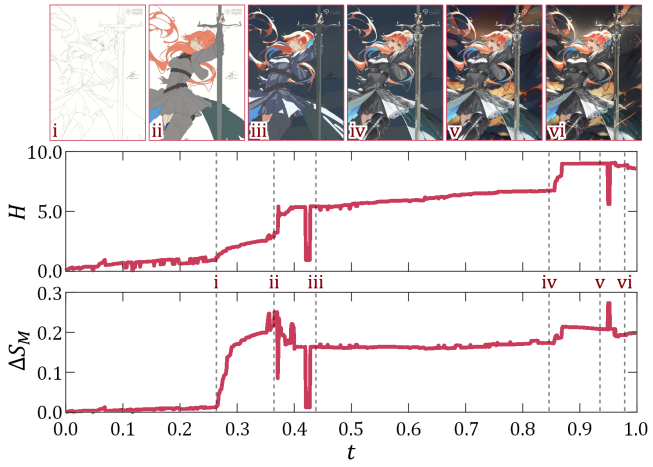

FIG. 8. Evolution of structural information during a digital painting of Hololive Production talent Takanashi Kiara. Top: Screenshots of a timelapse video illustrating the creative stages: (i) drafting, (ii) base coloring, (iii) background base coloring, (iv) detailing, (v) background enhancement, and (vi) final touches. Bottom: Plots of the Shannon entropy H and structural stability metric ΔS_M as functions of normalized number t of drawing steps. Artwork by Fadingz (X/Twitter ID: @FadingzZ). Used with permission.

diminish. This indicates that thermal agitation has become strong enough to overcome the interactions between vortices and antivortices.

Despite the structural differences at lower temperatures, the Hamiltonian E/L^2 exhibits consistent trends regardless of the initial condition used[42]. The specific heat C, showing a peak at $T \approx 1.07$, aligns closely with the critical temperatures $T \approx 1.02$–1.04 reported in previous studies[35–39].

In contrast to E/L^2 and C, the Shannon entropy H and the structural stability metric ΔS_M are sensitive to the initial conditions. For a uniform initial state (circles), H is not conserved at low temperatures $T < 1.07$; ΔS_M peaks at $T \approx 1.07$ akin to C, reflecting the emergence of stable structures, likely due to the nuanced balance between thermal agitation and vortex-antivortex interactions. Meanwhile, for a random initial state (triangles), H is conserved across all T, and ΔS_M mirrors the trend of E/L^2. These observations mirror those derived from Fig. 6 for the Potts model. The applicability of ΔS_M as a reliable order parameter for discerning transitions in a physical system is evidently contingent upon the conservation of H. Nonetheless, our study of the XY model demonstrates how H and ΔS_M can quantify structural information that traditional metrics such as C cannot reveal.

STRUCTURAL INFORMATION LANDSCAPE OF PAINTING

Our exploration of the Shannon entropy H and the structural stability metric ΔS_M across diverse domains—from textural images to physical models—underscores their capacity to quantify and trace transitions in structural information. Artistic paintings, with their diverse palettes and styles, offer a unique arena to extend our investigation further. Metaphorically speaking, the painting process can be likened to a phase transition, starting from a simple plain canvas and evolving into a complex composition filled with colors. This evolution, whether through iterative refinement or direct detailing, offers distinct trajectories on the *structural information landscape*, akin to how physical systems evolve following trajectories on their potential energy landscapes[43–45]. Such a landscape is expected to be complicated. After all, painting is a creative activity involving the artist's craftsmanship, intents, and decisions. Nonetheless, by examining how H and ΔS_M evolve from the initial blank state to the completed artwork, researchers could gain insights into and, more importantly, formulate more precise questions about the dynamic interplay of elements that constitute artistic complexity.

As an illustration of a trajectory on the structural information landscape, Fig. 8 shows a series of screenshots (i–vi), along with the evolution of H and ΔS_M throughout the drawing steps t, normalized against the total number of frames in a digital painting timelapse. Contrary to systems like the Potts and XY models (cf. Figs. 6 and 7), H and ΔS_M evolve in largely nonlinear way. Yet, several regimes can be identified. For instance, an initial transient during which the values of H and ΔS_M are relatively low can be seen for $t < 0.25$, corresponding to the drafting phase (i) of the illustration. As the base colors are added at $t \approx 0.25$, H and ΔS_M both increase abruptly, with the increase of ΔS_M being more pronounced, indicating the formation of visually stable structures on the canvas. This is also evidenced by the higher degree of the visual clarity of screenshot (ii) compared to (i). At $t \approx 0.35$, as the flat background color is added (iii), H further increases; in contrast, ΔS_M decreases due to the reduced contrast between the subject and the negative space. The value of ΔS_M is particularly sensitive to the background color: large fluctuations in the metric can be seen as the artist is choosing what color to use. Meanwhile, the entropy H, as a measure of color diversity, is rather stable against the choice of the flat background color. For $t > 0.35$, as details such as lighting, shadows, and textures are gradually added to the subject (iv), the colors become more diverse, increasing H further. On the other hand, ΔS_M is insensitive to the further addition of details, indicating equilibration in the quantity of visually stable structures, as reflected by the observation that screenshots (iii) and (iv) appear visually similar when viewed from afar. At $t \approx 0.85$, as the artist decorates the background to enhance the silhouette of the subject (v), H and ΔS_M again show a step increase, leading to the equilibration of both metrics. Minor adjustments during the finishing phase (vi) have no significant impact on H and ΔS_M.

Mapping the structural information landscape

Our timelapse video analysis has illuminated the complex landscape of structural information inherent in creative processes, with H and ΔS_M functioning as navigational tools to track an artist's specific chosen journey toward completing a painting. This exploration prompts us to consider how these

tools can be applied to estimate or partially map the structural information landscape of specific artworks, especially those in which timelapse is not available.

The paintings by Piet Mondrian, renowned for his pivotal role in the De Stijl movement and his advocacy for Neoplasticism, offer an ideal testbed. The art style of Mondrian is characterized by an emphasis on fundamental forms (horizontal and vertical lines) and a restricted palette (white, gray, black, red, blue, and yellow), striking a balance between order and disorder[46–49]. "Composition with Red, Blue and Yellow" stands out as a prime example of Mondrian's artistic vision. It features a deceptively simple arrangement of vertical and horizontal lines that partition the nominally square canvas into three color blocks and four white blocks. This simplicity provides a unique opportunity for us to generate digital variants, enabling targeted inquiries into Mondrian's choice of composition, in particular, his placement of the three colors in the seven blocks.

Fig. 9 presents a scatter plot mapping the $C_3^7 \times 3! = 210$ digital variants of "Composition with Red, Blue and Yellow", each having a distinct placement of colors, onto the entropy-stability plane. Each data point corresponds to a different variant. The cluster of points underscores the diverse range of structural information generated by these variants. Notably, the digital replica of the original artwork, denoted by a star, is positioned close to the global maximum of ΔS_M. This suggests that Mondrian's strategic color placement enhances the structural stability of the composition. Recall that a larger value of ΔS_M implies a greater degree of balance between order and disorder; such observation may reflect Mondrian's neoplasticist ideals of harmony and equilibrium. Furthermore, the scatter plot reveals that while some variants share the same value of H, indicating that for a fixed diversity of colors, their values of ΔS_M can vary significantly. This highlights the nuanced effect of color placement on the visual impact of an artwork.

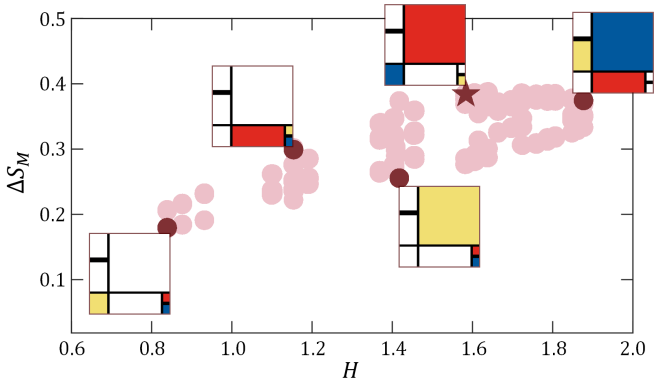

FIG. 9. Structural analysis of digital variants of Piet Mondrian's "Composition with Red, Blue and Yellow". The scatter plot maps 210 digital variants, each with unique color placements with the seven blocks, onto the entropy-stability plane defined by the Shannon entropy H and the structural stability metric ΔS_M. The star symbol (★) represents the digital replica of the original composition. Circles represent other variants. Inset images depict the color placements of the replica and several representative variants.

Navigating the structural information landscape

Our analysis in the previous section is tailored to a single artwork of Mondrian, and we acknowledge that the findings do not extend to his entire body of work. Nonetheless, the insights gained hold significant implications for the compositional analysis and design of visual communication elements like emotes, logos, comic panels, and frames in animations or movies. These media demand that their content be immediately graspable, often at reduced sizes or within split seconds[50]. In such a scenario, maximizing the visual stability of an image, quantifiable through ΔS_M, becomes crucial.

As a practical demonstration, we utilize H and ΔS_M to refine the composition of a pixel art emote for the video game Holo X Break. The goal is to ensure that the identity, posture, and expressions of the subject are discernible, even when displayed on small screens typical of video game interfaces. Fig. 10A depicts several compositions of the emote featuring Hololive Production talent Sakura Miko at different levels z of zoom and rescaled to a dimension of 512×512 by nearest neighbor interpolation[51]. In plots below the compositions, H and ΔS_M are presented as functions of z. For the original composition ($z = 1$), the portrait only occupies a small portion of the whole canvas, leading to a small value of structural information content for the emote, as revealed by the relatively low values of H and ΔS_M. Subsequent inspection of H and ΔS_M for various choices of z reveal that, while ΔS_M plateaus for $z \geq 5$, H peaks at $z = 5$, suggesting optimal color diversity at this scale. This leads us to select the composition with $z = 5$ for subsequent refinement.

In our pursuit of an optimal background color for the emote design, we select cyan as our initial guess. We set its hue, h, as our reference point at 0°. Fig. 10B presents this reference case alongside variants with different hues. The plots below show H and ΔS_M as functions of h. From a distance, the variants with $h = -160^\circ$ and 180° result in backgrounds that merge with the subject's pink hair, obscuring her silhouette. These options are readily excluded from consideration. Meanwhile, the distinction among the variants with $h = -120^\circ, -60^\circ, 0^\circ$, and 60° are subtler. Here, ΔS_M proves invaluable. While H remains constant across all hues—since altering the background color does not diversify the color palette of the emote—ΔS_M identifies three peaks at $h = 0^\circ$, -60°, and 60°, corresponding to the hues cyan, green, and blue. This reflects the complementary and triadic relationships to pink, underscoring the utility of ΔS_M in refining the choice of color to enhance visual clarity. It is worth noting that the value of ΔS_M obtained for the emote when triadic hues are used as the background is larger than that obtained if the complementary hue is used. This is because the Euclidean color distance between the subject's white outfit and green or blue is greater than the distance between white and cyan. Nonetheless, since the outfit is not our focus, we simply retain cyan ($h = 0^\circ$) as the background hue, leveraging its complementary relationship to pink. Further enhancements are explored by adjusting the saturation s

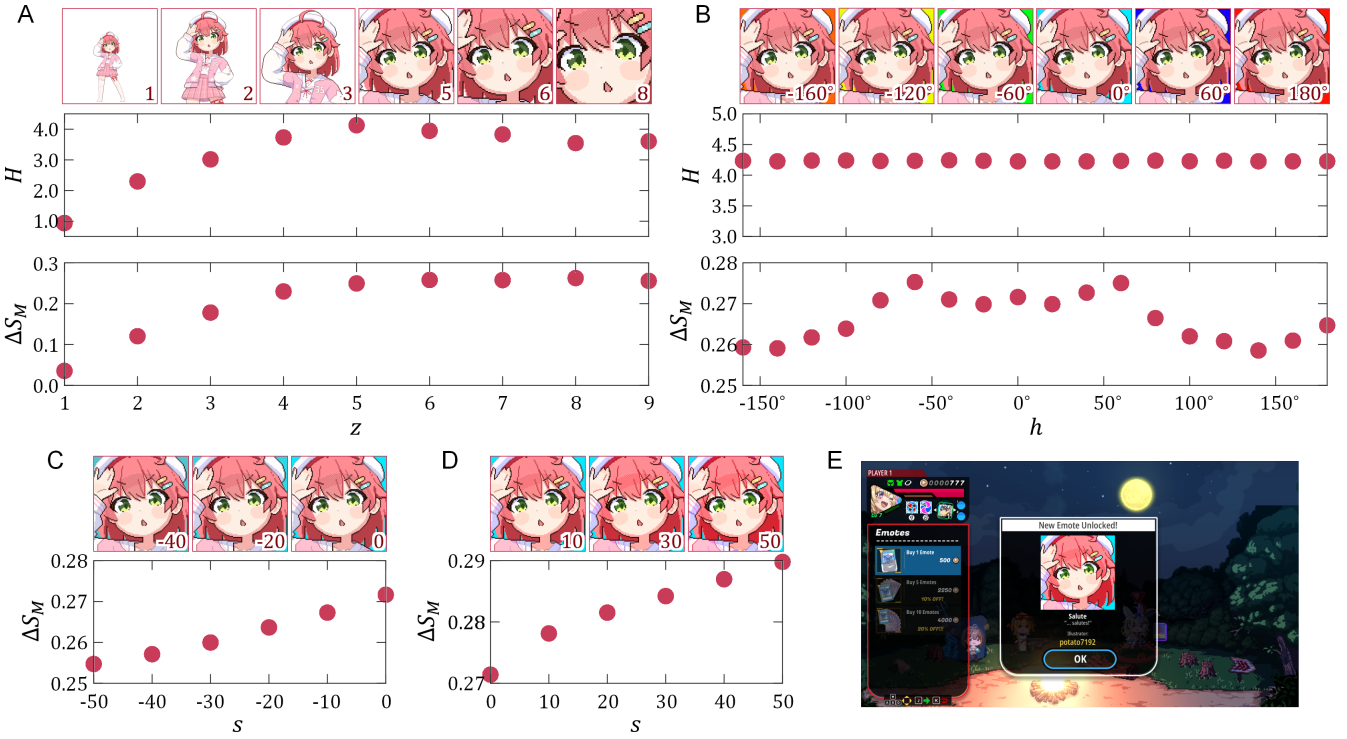

FIG. 10. Refinement of video game emote featuring Hololive Production talent Sakura Miko. (A) Emote designs with various levels of zoom z. The plots beneath illustrate the Shannon entropy H and the structural stability metric ΔS_M as functions of z. (B) Emote designs with various background hues h. The hue of cyan is marked as zero for convenience. The plots beneath show H and ΔS_M as functions of h. (C and D) Emote designs with various levels of saturation s of the cyan background and the overall composition. The plots below show H and ΔS_M as functions of s. (E) A screenshot of the finalized emote in the video game Holo X Break. Used with permission from Team HoloCure (X/Twitter ID: @HoloCureGame).

of the cyan background and the overall composition, as depicted in Figs. 10C and D, respectively. An increase in s directly correlates with a rise in ΔS_M. This leads us to select a higher saturation for the final emote design, as depicted by the in-game screenshot (Fig. 10E).

To inspect how our refined emote compares to other emotes, we construct a scatter plot (Fig. 11) mapping the 149 emotes in Holo X Break created by us (denoted as star symbols) and 23 artists (other symbols) onto the entropy-stability plane. The mean values of H and ΔS_M for the ensemble of emotes are 4.63 ± 1.15 and 0.275 ± 0.057, respectively. The star linked by a solid line represents the refined emote, with the line visualizing the refinement process (Figs. 10A and D) as a trajectory on the entropy-stability plane. The other three stars represent emotes drawn solely based on our artistic intuition, without using H and ΔS_M as guidance. The refined emote has $H = 4.18$ and $\Delta S_M = 0.287$, which are situated in proximity to the ensemble mean values, demonstrating the relevance of H and ΔS_M to the informed decision-making during the creative process. Meanwhile, for the other three emotes, $\Delta S_M = 0.189, 0.201$ and 0.278, two of which are a standard deviation below the ensemble mean, signifying that there is room for refinement in terms of their visual clarity.

Ascending the structural information landscape

Intriguingly, emotes by chibi artists Chroneco, Keenbiscuit, and Misa, who portray body proportions in a super-deformed, stylistically distorted way, tend to have larger values of ΔS_M (Fig. 11). These values can be as large as 0.4, two standard deviations in excess of the ensemble mean, suggesting that these artists are effective mountaineers in the structural information landscape. This observation prompts us to investigate the factors that contribute to the larger values of ΔS_M attained by their emotes.

Figure 12 compares eight 512×512 emotes—three created by us (A, B, and H) and six by chibi artists Chroneco (C and D), Keenbiscuit (E and F), and Misa (G)—along with their color distributions in the RGB space. The axis labels R, G, and B indicate the color intensity in the red, green, and blue channels, respectively. Each data point corresponds to a unique color $\mathbf{c}$, with its size proportional to the probability p_c of pixels of that color. The mean distance between all data point pairs, scaled by their sizes, gives the mean color distance $\langle d_c \rangle$. The values of ΔS_M for each emote are also provided. Recall, from [4], that $\langle d_c \rangle$ is the theoretical maximum ΔS_M of an image given its color histogram. The ratio $\phi = \Delta S_M / \langle d_c \rangle$ quantifies the effectiveness of the color spatial arrangement in maximizing ΔS_M. For all emotes in Figs. 12A–G, ϕ ranges from

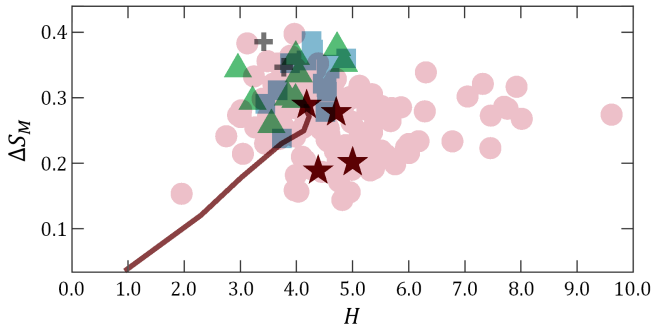

FIG. 11. Structural analysis of emotes featured in the video game Holo X Break. The scatter plot maps 149 emotes created by the authors and 23 artists onto the entropy-stability plane defined by the Shannon entropy H and the structural stability metric ΔS_M. The values of H have a mean of 4.63 ± 1.15, while ΔS_M has a mean of 0.275 ± 0.057. The star symbol (★) linked by a solid line represents the refined emote shown in Fig. 10E. The line visualizes the trajectory of the refinement process (Figs. 10A and D) of that emote on the entropy-stability plane. The other three stars represent emotes that are drawn by the authors purely based on artistic intuition. Plus signs, squares, and triangles correspond to works by chibi artists Chroneco (X/Twitter ID: @chrone_co), Keenbiscuit (X/Twitter ID: @keenbiscuit), and Misa (X/Twitter ID: @Misamisatotomi), respectively. Circles denote works by other contributing artists.

0.769 to 0.882, indicating that their ΔS_M values are rather close to the maximal values achievable with their respective color histograms.

Our refined emote (Fig. 12A), consisting mostly of reddish colors, has the most data points positioned near the $R = 1$ plane, occupying larger portions of the RGB space than brown from the line art and cyan from the background. The position $(0,1,1)$ of cyan results in larger color distances with its complementary reddish colors, which are located close to $(1,0,0)$. This contributes to a larger value of $\langle d_c \rangle$ and hence higher ΔS_M of the emote. In our other emote (Fig. 12B), the data points form a nominally diagonal curve in the RGB space between the origin and $(1,1,1)$, as we used the color-trace technique, aka line coloring, to blend the black contours with their nearby colors for a softer appearance. The data point sizes are rather uniform along the curve, except for the whitish colors. The dominance of the whitish colors, which have small color distances from each other, as well as the rarity of their complementary blackish colors, result in a lower value of $\langle d_c \rangle$. In turn, this leads to a lower ΔS_M compared to our refined emote.

On the other hand, emotes by Chroneco (Figs. 12C and D) demonstrate drastically different distributions of the data points. In particular, brown occupies a considerably larger portion of the RGB space compared to our emotes, which can be attributed to the artist's heavy use of thick brown lines to illustrate their subject. Brown, located close to the origin, has large color distances with whitish colors, contributing to a larger value of $\langle d_c \rangle$. On the other hand, the artist's use of simple shapes effectively clusters the colors into several large areas in the emote, minimizing the local color distances in those regions, leading to a lower value of S and, hence, a larger value of ΔS_M. The same observation applies to the emotes by Keenbiscuit (Figs. 12E and F) and Misa (Fig. 12G).

These findings mentioned above afford the development of a holistic optimization workflow, more effective than the greedy refinement strategy we employed in Fig. 10, to create artworks that have high visual clarity as measured by ΔS_M. Perhaps the most reasonable way is to initially select a manageable number of main colors with a large value of $\langle d_c \rangle$. Then, illustrate the artwork using simple shapes composed of those main colors. Finally, add details such as highlights, shadows, gradation, or textures and fine-tune the balance of different elements until completion. We utilized this recipe to create an optimized version of our emote, shown in Fig. 12H, which contrasts with the original depicted in Fig. 12A. The emote has a structural stability metric of $\Delta S_M = 0.426$, which is higher than the values shown in Fig. 11 for the 149 emotes featured in Holo X Break, verifying the effectiveness of our proposed workflow.

Information thermodynamics of digital art

A key finding of this study is the realization that colors in visual art serve as state variables. This realization is rooted in the similarity between ΔS, energy E, and specific heat C, which points to an energetic interpretation of ΔS_M as the maximal structural energy difference between the slightly melted and fully melted versions of the original image. This affords a thermodynamic perspective on the analysis of digital art. Our examination of emotes produced by chibi artists further bolsters this perspective, as those artists consistently achieve remarkably high values of ΔS_M. This suggests that their use of simple shapes and carefully chosen color palettes effectively minimizes the local color distances d_{ij} (and, hence, structural energy S) while maximizing the global mean color distance $\langle d_c \rangle$ (and, hence, S^*). This approach echoes certain optimization principles of thermodynamics[52], where systems tend to maximize their entropy, thus attaining thermally stable configurations.

The thermodynamic perspective on digital art opens up exciting avenues for future research and application. By leveraging the tools and concepts of thermodynamics, such as phase diagrams, free-energy landscapes, and optimization principles, researchers and artists can gain deeper insights into the dynamics of artistic composition, the factors influencing visual stability, and the strategies for creating effective digital artworks. This interdisciplinary approach promises to enrich our understanding of the creative process and to provide a framework for the systematic analysis and design of digital art.

COMPARISON WITH OTHER STRUCTURAL MEASURES

The introduction of the structural stability metric ΔS_M in this study marks a significant advancement in the quantification of the visual clarity of digital images. Distinguished by its non-parametric nature and mathematical simplicity, ΔS_M tran-

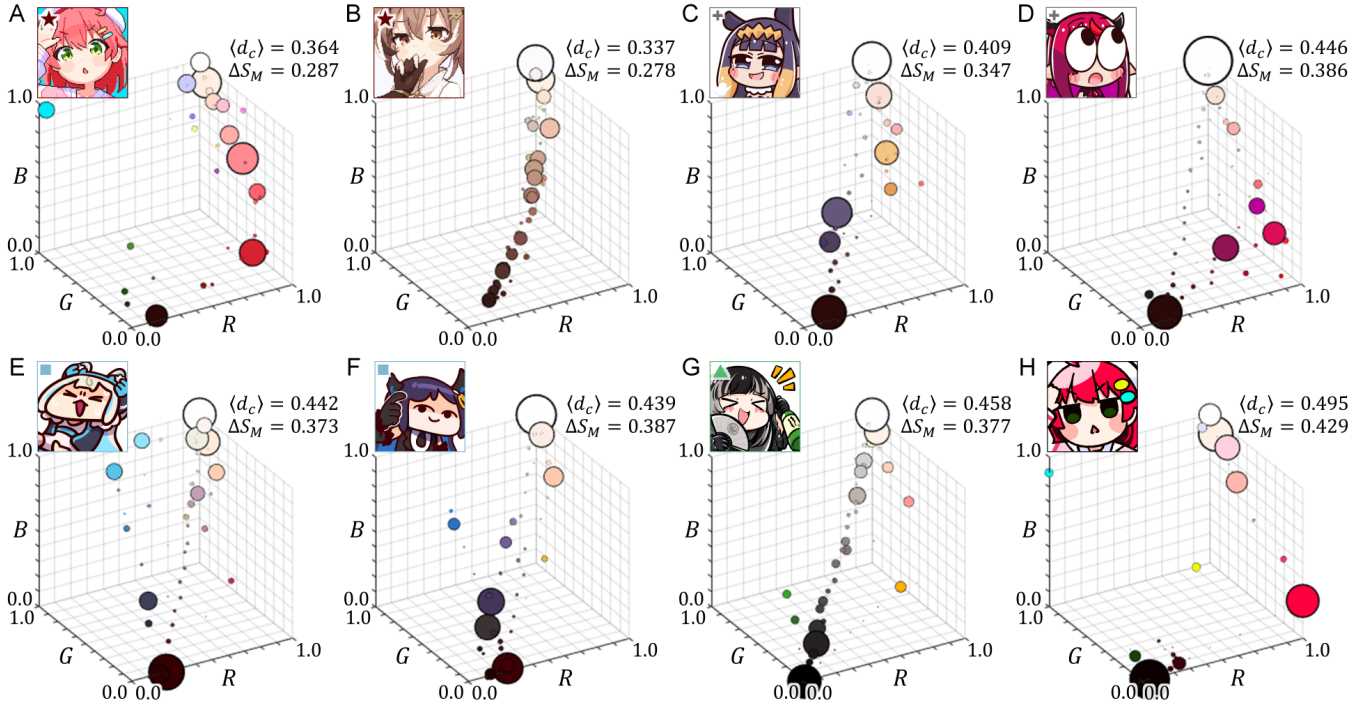

FIG. 12. Three-dimensional scatter plots visualizing the distribution of the fifty most frequent colors within the RGB space for eight different emotes of dimension 512×512. The emotes, which are featured in the video game Holo X Break, are created by (A and B) the authors, (C and D) Chroneco, (E and F) Keenbiscuit, and (G) Misa. (H) An optimized version of the emote in A. From A to G, the emote depicts Hololive Production talent Sakura Miko, Nanashi Mumei, Ninomae Ina'nis, IRyS, Fuwawa Abyssgard, Nerissa Ravencroft, and Juufuutei Raden, respectively. The mean color distance $\langle d_c \rangle$ and structural stability metric ΔS_M of the emotes are also shown for comparison. All artworks are used with permission from the artists.

scends the limitations of several existing approaches, offering a more intuitive and comprehensive measure of structural information.

One of the key advantages of ΔS_M is its independence of adjustable parameters. This is in contrast to existing methods that require the tuning of parameters, such as the quality factor of the JPEG file format in image compression techniques[2,4], the grid size used to sample ordinal patterns in delentropy[5,6] and permutation entropy[7–9], or the number of spatial averaging steps in renormalization-based methods[10,11], all of which are typically based on subjective criteria. This renders ΔS_M more objective and universally applicable across various image types and artistic styles.

Another significant advantage of ΔS_M is its mathematical simplicity, which not only facilitates straightforward computation but also makes the concept easily understandable to a wide audience, including art practitioners who may not have formal mathematical training. This simplicity has been confirmed through our communications with artists in the field, who have been able to grasp and appreciate the method without difficulty. The capacity of ΔS_M to convey complex ideas about structural stability and visual clarity in an intuitive manner will be particularly valuable in collaborative settings where scientists and artists aim to bridge the gap between quantitative analysis and creative expression.

Moreover, ΔS_M takes into account both the intensity and color of pixels, making it more suitable for analyzing the structural stability of color images, which are central to most artworks. This is in contrast to image compression-based measures and permutation entropy or delentropy-based measures, which neglect the role of color in structural information. The structural stability metric ΔS_M is also robust against changes in image resolution, making it more suitable for analyzing intrinsic structures depicted in images. This is particularly important for artworks where the content and composition, rather than the resolution, are the primary carriers of structural information. In comparison, image compression techniques assess structural information based on compressibility, which is largely dependent on image resolution. Furthermore, ΔS_M uses the flipping of pixel pairs as a means of perturbation, which preserves the integrity of visually stable structures and the color histogram of the artwork. This approach avoids the issues faced by renormalization-based methods, where spatial averaging can destroy most structural information in certain artistic styles, such as pixel art.

Despite its numerous advantages, ΔS_M has some limitations that are worth noting. One limitation is that ΔS_M relies solely on pixel pair flipping as a means of perturbation, which may not capture all types of distortions that can affect the visual clarity of an image. Additionally, while ΔS_M considers both the intensity and color of pixels, it does not explicitly account for the perceptual aspects of color, such as the varying sensitivity of the human visual system to different colors. These limitations may affect the extent to which ΔS_M can cap-

ture certain types of structural information in digital images. Nonetheless, these limitations can be easily overcome by incorporating other types of distortion as perturbation and by replacing RGB color space with perceptually uniform color spaces[53].

CONCLUSIONS AND OUTLOOKS

To summarize, we have examined the application of the structural stability metric ΔS_M across various domains, from the analysis of textures and physical models to the exploration of artistic expression and the refinement of emotes. Our findings demonstrate that, together with the Shannon entropy H, ΔS_M serves as a powerful tool for dissecting the intricate landscape of structural information. Through analyzing the Potts and XY models, the artwork of Piet Mondrian, and the design of video game emotes, the pivotal role of structural stability in the scientific and creative realms has been illuminated. These diverse applications underscore the versatility of ΔS_M as a tool for extracting nuanced structural information from digital images, enhancing decision-making and data analysis in both scientific and creative fields.

In the scientific realm, the potential applications of ΔS_M extend well beyond the cases presented in this paper. A compelling avenue for future research involves applying the metric to dissect the dynamic evolution and composition of r/place, an online collaborative art project and social experiment hosted on Reddit[54,55]. This platform transforms a blank canvas into a mosaic of collective creativity, offering a unique playground for one to examine social dynamics and community interaction[56]. Additionally, employing ΔS_M to analyze natural patterns, such as animal skins[57–60], cracks in drying colloidal films[61–63], and flow instabilities[64–68], could yield new insights, from the standpoint of structural information, into the mechanisms that govern their formation.

In the artistic realm, future research could leverage ΔS_M to scrutinize the composition of film and animation frames, offering insights into the visual aspect of emotional arc in storytelling[69,70]. It could also be interesting to prepare images of various values of ΔS_M and test those images against human observers or different saliency models in computer vision[71–73], thereby achieving a deeper understanding regarding the relation between the structural stability and perception aspects of images. Another promising direction could involve the application of the metric to study digital replicas of historically and culturally significant artworks[74–77], as we did for Mondrian's "Composition with Red, Blue and Yellow", enhancing our comprehension and appreciation of human artistry in an era that is increasingly dominated by machine-generated imagery[78,79].

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support from the Okinawa Institute of Science and Technology Graduate University with subsidy funding from the Cabinet Office, Government of Japan. C.S.T. thanks Ronin, Advarcher, Fadingz, Yasu, Chroneco, Keenbiscuit, Misa, and Team Holocure for granting permission to analyze or showcase their works in this study. All artworks depicting Hololive Production talents are used following the Derivative Works Guidelines by COVER Corporation, with confirmation from customer support.

[1] M. M. France, A. Hénaut, and J. Mandelbrojt, "Art, therefore entropy," Leonardo **27**, 219–221 (1994).
[2] P. Machado, J. Romero, M. Nadal, A. Santos, J. Correia, and A. Carballal, "Computerized measures of visual complexity," Acta Psychol. **160**, 43–57 (2015).
[3] S. Martiniani, P. M. Chaikin, and D. Levine, "Quantifying hidden order out of equilibrium," Phys. Rev. X **9**, 011031 (2019).
[4] S. Lakhal, A. Darmon, J. P. Bouchaud, and M. Benzaquen, "Beauty and structural complexity," Phys. Rev. Res. **2**, 022058 (2020).
[5] K. G. Larkin, "Reflections on shannon information: In search of a natural information-entropy for images," arXiv preprint arXiv:1609.01117 (2016).
[6] T. M. Khan, S. S. Naqvi, and E. Meijering, "Leveraging image complexity in macro-level neural network design for medical image segmentation," Sci. Rep. **12**, 22286 (2022).
[7] H. V. Ribeiro, L. Zunino, E. K. Lenzi, P. A. Santoro, and R. S. Mendes, "Complexity-entropy causality plane as a complexity measure for two-dimensional patterns," PLoS One **7**, 1–9 (2012).
[8] H. Y. D. Sigaki, M. Perc, and H. V. Ribeiro, "History of art paintings through the lens of entropy and complexity," Proc. Natl. Acad. Sci. U. S. A. **115**, E8585–E8594 (2018).
[9] C. Bandt and K. Wittfeld, "Two new parameters for the ordinal analysis of images," Chaos **33** (2023).
[10] A. A. Bagrov, I. A. Iakovlev, A. A. Iliasov, M. I. Katsnelson, and V. V. Mazurenko, "Multiscale structural complexity of natural patterns," Proc. Natl. Acad. Sci. U. S. A. **117**, 30241–30251 (2020).
[11] V. V. Mazurenko, I. A. Iakovlev, O. M. Sotnikov, and M. I. Katsnelson, "Estimating patterns of classical and quantum skyrmion states," J. Phys. Soc. Jpn. **92**, 081004 (2023).
[12] J. Rigau, M. Feixas, and M. Sbert, "Informational aesthetics measures," IEEE Comput. Graph. Appl. **28**, 24–34 (2008).
[13] S. J. Wan, P. Prusinkiewicz, and S. K. M. Wong, "Variance-based color image quantization for frame buffer display," Color Res. Appl. **15**, 52–58 (1990).
[14] J. R. Bentley, "The origin of man'yōgana," Bull. Sch. Orient. Afr. Stud. **64**, 59–73 (2001).
[15] K. Heffernan, "The role of phonemic contrast in the formation of Sino-Japanese," J. East Asian Linguist. **16**, 61–86 (2007).
[16] G. T. Fechner, *Elemente der psychophysik*, Vol. 2 (Breitkopf u. Härtel, 1860).
[17] D. Algom, "The Weber–Fechner law: A misnomer that persists but that should go away." Psychol. Rev. **128**, 757 (2021).
[18] R. G. Keys, "Cubic convolution interpolation for digital image processing," IEEE Trans. Acoust., Speech, Signal Process **29**, 1153–1160 (1981).
[19] G. Kylberg and I. M. Sintorn, "On the influence of interpolation method on rotation invariance in texture recognition," EURASIP J. Image Video Process. **2016**, 1–12 (2016).
[20] R. B. Potts, "Some generalized order-disorder transformations," Math. Proc. Camb. Philos. Soc. **48**, 106–109 (1952).
[21] J. Ashkin and E. Teller, "Statistics of two-dimensional lattices with four components," Phys. Rev. **64**, 178 (1943).
[22] E. Ising, *Beitrag zur theorie des ferro-und paramagnetismus*, Ph.D. thesis, Grefe & Tiedemann Hamburg, Germany (1924).
[23] R. J. Glauber, "Time-dependent statistics of the Ising model," J. Math. Phys. **4**, 294–307 (1963).
[24] K. Kawasaki, "Diffusion constants near the critical point for time-dependent Ising models. I," Phys. Rev. **145**, 224 (1966).
[25] K. Kawasaki, "Diffusion constants near the critical point for time-dependent Ising models. II," Phys. Rev. **148**, 375 (1966).
[26] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," J. Chem. Phys. **21**, 1087–1092 (1953).

[27] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," Biometrika **57**, 97–109 (1970).

[28] K. Binder, "Static and dynamic critical phenomena of the two-dimensional q-state Potts model," J. Stat. Phys. **24**, 69–86 (1981).

[29] G. Wahba, "Smoothing noisy data with spline functions," Numer. Math. **24**, 383–393 (1975).

[30] M. E. J. Newman and G. T. Barkema, *Monte Carlo methods in statistical physics* (Oxford University Press, USA, 1999).

[31] J. P. Sethna, *Statistical mechanics: entropy, order parameters, and complexity*, Vol. 14 (Oxford University Press, USA, 2021).

[32] V. L. Berezinskii, "Destruction of long-range order in one-dimensional and two-dimensional systems having a continuous symmetry group I. classical systems," Sov. Phys. JETP **32**, 493–500 (1971).

[33] J. M. Kosterlitz and D. J. Thouless, "Ordering, metastability and phase transitions in two-dimensional systems," J. Phys. C: Solid State Phys. **6**, 1181 (1973).

[34] J. M. Kosterlitz, "The critical properties of the two-dimensional xy model," J. Phys. C: Solid State Phys. **7**, 1046 (1974).

[35] J. Tobochnik and G. V. Chester, "Monte Carlo study of the planar spin model," Phys. Rev. B **20**, 3761 (1979).

[36] R. Gupta and C. F. Baillie, "Critical behavior of the two-dimensional XY model," Phys. Rev. B **45**, 2883 (1992).

[37] J. Xu and H. R. Ma, "Density of states of a two-dimensional XY model from the Wang-Landau algorithm," Phys. Rev. E **75**, 041115 (2007).

[38] J. F. Yu, Z. Y. Xie, Y. Meurice, Y. Liu, A. Denbleyker, H. Zou, M. P. Qin, J. Chen, and T. Xiang, "Tensor renormalization group study of classical XY model on the square lattice," Phys. Rev. E **89**, 013308 (2014).

[39] P. Jakubczyk and A. Eberlein, "Thermodynamics of the two-dimensional XY model from functional renormalization," Phys. Rev. E **93**, 062145 (2016).

[40] Y. D. Hsieh, Y. J. Kao, and A. W. Sandvik, "Finite-size scaling method for the Berezinskii–Kosterlitz–Thouless transition," J. Stat. Mech.: Theory Exp. **2013**, P09001 (2013).

[41] W. Zhang, J. Liu, and T. C. Wei, "Machine learning of phase transitions in the percolation and XY models," Phys. Rev. E **99**, 032142 (2019).

[42] X. Leoncini, A. D. Verga, and S. Ruffo, "Hamiltonian dynamics and the phase transition of the XY model," Phys. Rev. E **57**, 6377 (1998).

[43] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, "Theory of protein folding: the energy landscape perspective," Annu. Rev. Phys. Chem. **48**, 545–600 (1997).

[44] P. G. Debenedetti and F. H. Stillinger, "Supercooled liquids and the glass transition," Nature **410**, 259–267 (2001).

[45] D. J. Wales, "Exploring energy landscapes," Annu. Rev. Phys. Chem. **69**, 401–425 (2018).

[46] C. Wiegand, "The meaning of Mondrian," J. Aesthet. Art Crit. **2**, 62–70 (1943).

[47] A. Chandler, *The Aesthetics of Piet Mondrian* (MSS Information Corporation New York, 1972).

[48] I. C. McManus, B. Cheema, and J. Stoker, "The aesthetics of composition: A study of Mondrian," Empirical Studies of the Arts **11**, 83–94 (1993).

[49] A. Fallahzadeh and G. Gamache, "Equilibrium and rhythm in Piet Mondrian's Neo-Plastic compositions," Cogent Arts Humanit. **5**, 1525858 (2018).

[50] H. Bacher and S. Suryavanshi, *Vision: Color and Composition for Film* (Laurence King Publishing, 2018).

[51] J. A. Parker, R. V. Kenyon, and D. E. Troxel, "Comparison of interpolating methods for image resampling," IEEE Trans. Med. Imaging **2**, 31–39 (1983).

[52] L. M. Martyushev and V. D. Seleznev, "Maximum entropy production principle in physics, chemistry and biology," Phys. Rep. **426**, 1–45 (2006).

[53] G. Paschos, "Perceptually uniform color spaces for color texture analysis: an empirical evaluation," IEEE Trans. Image Process. **10**, 932–937 (2001).

[54] K. T. Litherland and A. I. Mørch, "Instruction vs. emergence on r/place: Understanding the growth and control of evolving artifacts in mass collaboration," Comput. Human Behav. **122**, 106845 (2021).

[55] A. M. Adams, J. Fernandez, and O. Witkowski, "Two ways of understanding social dynamics: Analyzing the predictability of emergent of objects in Reddit r/place dependent on locality in space and time," arXiv preprint arXiv:2206.03563 (2022).

[56] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," Rev. Mod. Phys. **81**, 591 (2009).

[57] A. J. Koch and H. Meinhardt, "Biological pattern formation: from basic mechanisms to complex structures," Rev. Mod. Phys. **66**, 1481 (1994).

[58] A. Nakamasu, G. Takahashi, A. Kanbe, and S. Kondo, "Interactions between zebrafish pigment cells responsible for the generation of Turing patterns," Proc. Natl. Acad. Sci. U. S. A. **106**, 8429–8434 (2009).

[59] S. Kondo and T. Miura, "Reaction-diffusion model as a framework for understanding biological pattern formation," science **329**, 1616–1620 (2010).

[60] B. M. Alessio and A. Gupta, "Diffusiophoresis-enhanced Turing patterns," Sci. Adv. **9**, eadj2457 (2023).

[61] K. B. Singh and M. S. Tirumkudulu, "Cracking in drying colloidal films," Phys. Rev. Lett. **98**, 218302 (2007).

[62] A. F. Routh, "Drying of thin colloidal films," Rep. Prog. Phys. **76**, 046603 (2013).

[63] M. Leang, F. Giorgiutti-Dauphine, L. T. Lee, and L. Pauchard, "Crack opening: from colloidal systems to paintings," Soft Matter **13**, 5802–5808 (2017).

[64] J. P. Gollub and J. S. Langer, "Pattern formation in nonequilibrium physics," Rev. Mod. Phys. **71**, S396 (1999).

[65] F. Gallaire and P. T. Brun, "Fluid dynamic instabilities: theory and application to pattern forming in complex media," Philos. Trans. R. Soc. A **375**, 20160155 (2017).

[66] S. Zetina, F. A. Godínez, and R. Zenit, "A hydrodynamic instability is used to create aesthetically appealing patterns in painting," PLoS One **10**, e0126135 (2015).

[67] B. Palacios, A. Rosario, M. M. Wilhelmus, S. Zetina, and R. Zenit, "Pollock avoided hydrodynamic instabilities to paint with his dripping technique," PLoS One **14**, e0223706 (2019).

[68] R. Zenit, "Some fluid mechanical aspects of artistic painting," Phys. Rev. Fluids **4**, 110507 (2019).

[69] A. J. Reagan, L. Mitchell, D. Kiley, C. M. Danforth, and P. S. Dodds, "The emotional arcs of stories are dominated by six basic shapes," EPJ Data Sci. **5**, 1–12 (2016).

[70] M. Perc, "Beauty in artistic expressions through the eyes of networks and physics," J. R. Soc. Interface **17**, 20190686 (2020).

[71] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell. **20**, 1254–1259 (1998).

[72] L. Itti and C. Koch, "Computational modelling of visual attention," Nat. Rev. Neurosci. **2**, 194–203 (2001).

[73] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, Vol. 19, edited by B. Schölkopf, J. Platt, and T. Hoffman (MIT Press, 2006).

[74] D. Kim, S. W. Son, and H. Jeong, "Large-scale quantitative analysis of painting arts," Sci. Rep. **4**, 7370 (2014).

[75] B. Lee, D. Kim, S. Sun, H. Jeong, and J. Park, "Heterogeneity in chromatic distance in images and characterization of massive painting data set," PLoS One **13**, e0204430 (2018).

[76] B. Lee, M. K. Seo, D. Kim, I. Shin, M. Schich, H. Jeong, and S. K. Han, "Dissecting landscape art history with information theory," Proc. Natl. Acad. Sci. U. S. A. **117**, 26580–26590 (2020).

[77] A. Karjus, M. Canet Solà, T. Ohm, S. E. Ahnert, and M. Schich, "Compression ensembles quantify aesthetic complexity and the evolution of visual art," EPJ Data Sci. **12**, 21 (2023).

[78] H. H. Jiang, L. Brown, J. Cheng, M. Khan, A. Gupta, D. Workman, A. Hanna, J. Flowers, and T. Gebru, "AI art and its impact on artists," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (2023) pp. 363–374.

[79] A. Y. J. Ha, J. Passananti, R. Bhaskar, S. Shan, R. Southen, H. Zheng, and B. Y. Zhao, "Organic or diffused: Can we distinguish human art from AI-generated images?" arXiv preprint arXiv:2402.03214 (2024).