




# Compression ensembles quantify aesthetic complexity and the evolution of visual art

Andres Karjus<sup>1\*</sup> , Mar Canet Solà<sup>1</sup>, Tillmann Ohm<sup>1</sup>, Sebastian E. Ahnert<sup>2,3†</sup> and Maximilian Schich<sup>1†</sup>

\*Correspondence:

[andres.karjus@tlu.ee](mailto:andres.karjus@tlu.ee)

<sup>1</sup>ERA Chair for Cultural Data Analytics, Tallinn University, Narva mnt 25, 10120, Tallinn, Estonia  
Full list of author information is available at the end of the article  
<sup>†</sup>Equal contributors

## Abstract

To the human eye, different images appear more or less complex, but capturing this intuition in a single aesthetic measure is considered hard. Here, we propose a computationally simple, transparent method for modeling aesthetic complexity as a multidimensional algorithmic phenomenon, which enables the systematic analysis of large image datasets. The approach captures visual family resemblance via a multitude of image transformations and subsequent compressions, yielding explainable embeddings. It aligns well with human judgments of visual complexity, and performs well in authorship and style recognition tasks. Showcasing the functionality, we apply the method to 125,000 artworks, recovering trends and revealing new insights regarding historical art, artistic careers over centuries, and emerging aesthetics in a contemporary NFT art market. Our approach, here applied to images but applicable more broadly, provides a new perspective to quantitative aesthetics, connoisseurship, multidimensional meaning spaces, and the study of cultural complexity.

**Keywords:** Aesthetic complexity; Kolmogorov complexity; Image compression; Art history; Family resemblance; Artistic careers; NFT art; Temporal resemblance

## 1 Introduction

The quantification of visual aesthetics and artistic expression goes back to Birkhoff [1] and Bense [2], inspiring several computational approaches [3–16]. Research drawing on information theory has shown repeatedly and in parallel that visual complexity can be estimated with some accuracy using compression algorithms such as ZIP or GIF [17–29]. While some previous proposals were tested against perceptual human judgments, results diverge as to which compression algorithm or approach would be optimal. Elsewhere in cultural research, measures of compression length have been used at face value to compare the complexity of visual inputs [30–32]. Related information-theoretic approaches have also used entropy to quantify artistic styles and conceptual groupings [8, 12, 33].

Considering quantitative analysis of visual art, it makes sense to adopt algorithmic approaches, as creative processes themselves also follow a set of procedures — or algorithms, in the broadest sense — in which artists may be similar or differ [2, 3, 34]. Algorithmic complexity is best understood via information theory [35, 36], defining the Kolmogorov complexity of a string or dataset as the length of the shortest algorithm reproducing the

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

data. Kolmogorov complexity is uncomputable, but can be approximated using the compression size of a dataset as its upper bound. In the case of images, the more compressible one is — i.e. the larger the difference between a bitmap and a compressed version e.g. GIF — the visually simpler it is. This notion in turn can be extended to measures of algorithmic distance, one example being normalized compression distance (NCD) [37]. This however requires a separate compression for each comparison event, and has been rarely applied to visual materials [38]. Pairwise image comparison frameworks [19, 22], suffer from the same computational bottleneck. Furthermore, visual media such as art or photography can, intuitively, differ in various dimensions of complexity, such as color, composition and detail. A single value of complexity would likely fail to capture this multiplicity.

### 1.1 A method proposal

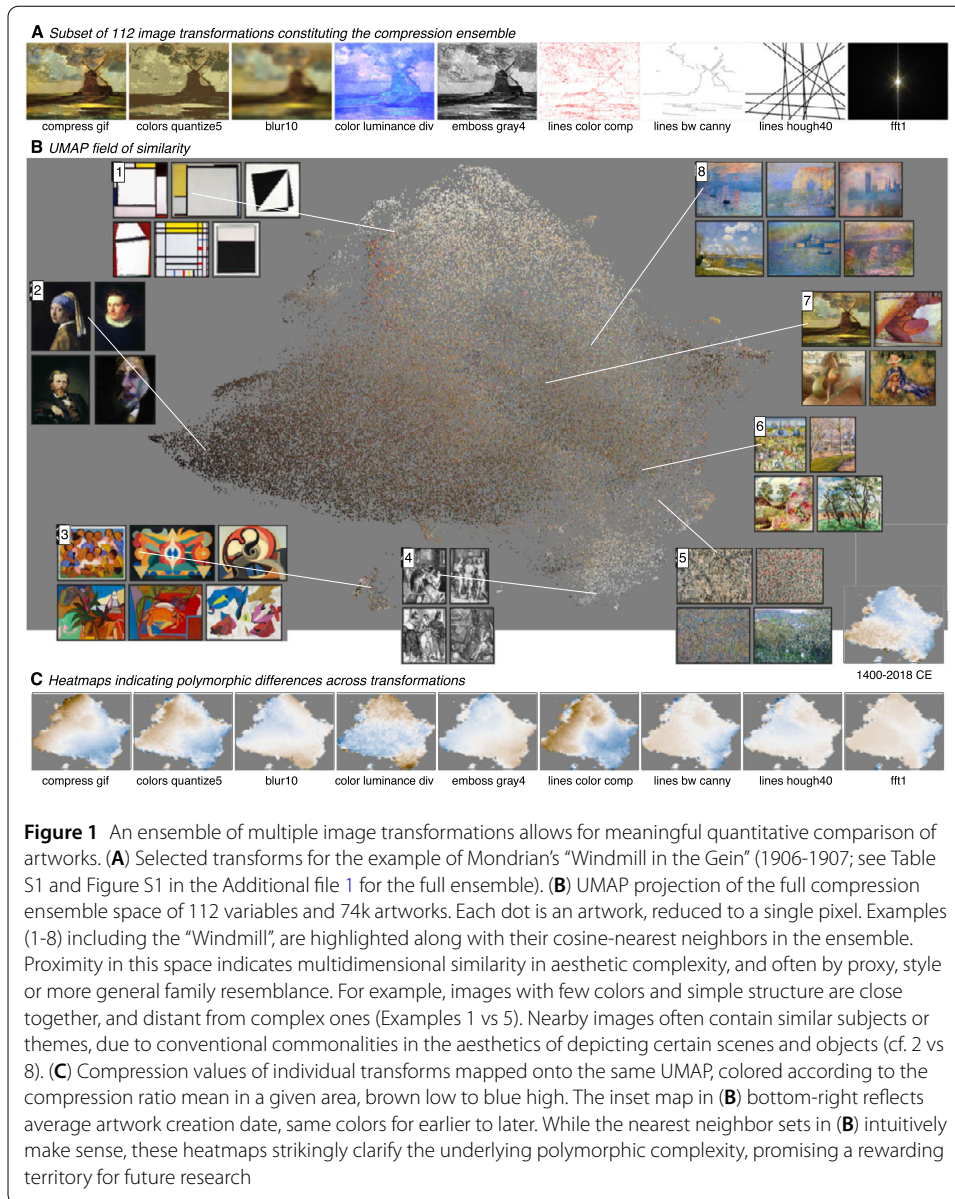
We introduce “compression ensembles”, an efficient and explainable algorithmic comparison framework for images. Instead of attempting to find a single algorithm or metric to best match human judgments or perform on downstream tasks, we argue for an ensemble approach of concurrently using multiple measures. This consists of two steps: multiple transformations and subsequent compression. For a given input image, a set of alternative images is produced by applying a number of different image processing transformations, including various filters, blurs, distortions and color manipulations (see Fig. 1.A, Methods section, and Additional file 1, for details).

These new images are compressed (possibly with multiple compression algorithms), and the resulting file sizes are divided by the compression size of the original bitmap; the unfiltered original input is also compressed but divided by the size of the bitmap input. This yields a vector of compression ratios. Further statistical transformations such as colorfulness metrics and fractal dimension [5, 16, 39, 40] (compressions in a broader sense) can also be added to these vectors, and rescaled if the magnitudes differ, which we do (see Methods). Given the perspective of artists as (or as executing) algorithms — in the very broadest sense — our approach aims to capture the residual signal of the generating process, the “algorithmic fingerprints” of an artwork, through the operationalization of various aspects of visual aesthetic complexity, as estimated by compression ratios of the visual transformations. This approach can, however, currently only assess the visual complexity apparent in the image. This may or may not correlate with production complexity, i.e. the time, effort, expertise, etc. that goes into producing an artwork.

As demonstrated below, compression ensembles can be used for estimating relative visual complexity as a multidimensional phenomenon; inferring specific aesthetic aspects of complexity where any two images or artist oeuvres differ; and as input to any predictive downstream tasks where image complexity (and by proxy aesthetic or stylistic) profiles matter. We also show its use as a “computational art historian” or “curator” algorithm to quantify and systematically explore the dynamics of art artistic careers in very large datasets. While we focus our showcase here on art, it is suitable for comparing any sort of images, and as described in the Discussion, straightforwardly extendable to other types of media to estimate their inherent aesthetic complexity.

### 1.2 Relation to other image embedding approaches

Compression vectors can be rapidly compared, clustered, or used in image recognition and comparison tasks, as demonstrated below. Our method is comparable, but largely orthogonal to deep learning models of computer vision that also embed images in numeric vector



spaces. Typically pre-trained on very large image databases, the latter excel at inferring image similarity (in the feature space), and can be tuned to predict discrete classes such as objects on the image, historical style, or authorship [9, 41, 42]. Compression ensembles can also be used to estimate similarity (in the complexity space), and nearby images often depict similar subjects — albeit incidentally, due to shared stylistic commonalities in how certain subjects are depicted (Fig. 1.B). However, the main function of the ensembles is to operationalize visual complexity as such. Our approach infers meaningful values for an input without requiring any costly pre-training, and therefore also entirely bypasses the “training set bias” problem inherent to pre-trained machine learning models. Fitting new images to an already generated vector space here does not require retraining or realignment of the space either, just a matching set of transformations.

We are less interested in discrete classification (although we do run tests on these tasks, see Methods), and more in efficient exploration and curation of large image sets in the

continuous complexity space. We show that this is also a good proxy for aesthetic complexity, and by extension, indeed notions of artistic style (see Figs. 1.B, 3). The issue with discrete style classification [5, 42] is that it inevitably requires training on some “ground truth”, which in the case of historical style periods is highly debatable. Our approach goes beyond limited discrete categories to operate in a continuous aesthetics space.

To contrast another related body of literature, some previous research has tried to build statistical and machine learning models to explicitly predict what humans participants in psychology experiments perceive as intuitively complex, “aesthetic”, or “beautiful” [14, 20, 25, 43–45]. We did test our approach on a set of human complexity judgments and show that it performs very well (see Methods), but our goal, to provide a general framework, is broader. We use complexity as a practical comparative algorithmic measure, and, to be very clear, “aesthetic” not as a value judgment. Besides “beauty”, here we also do not attempt to quantify complexity in terms of iconography, semantics nor number of depicted subjects (but see Discussion for possible extensions).

Our approach is also conceptually related to “ensemble methods” in statistics and machine learning [46], but instead of a single aggregated prediction, the useful output here is the full vector of complexity estimates. Approaches using ensembles of various metrics have also been proposed in linguistics for language complexity [47] and economics for tax system complexity [48].

### 1.3 Explainable vector spaces

Importantly, unlike all of the aforementioned machine learning based image embeddings, the dimensions or variables in a compression ensemble remain interpretable: a compression ratio difference between two images for a given transformation indicates that they differ in this aspect. Applying a black-and-white transformation to a colorful image increases compressibility relative to the original (compression ratio  $\ll 1$ ), but has no effect on an already black-and-white image (ratio  $\approx 1$ ). Applying coarse pixelation to Piet Mondrian’s abstract paintings (Fig. 1.B.1) barely changes their compressibility, while it greatly increases compressibility for the highly detailed works of Hieronymus Bosch (Fig. 1.B.6). Therefore, images similar in multiple aspects end up close together in multidimensional compression space, while dissimilar ones stay far apart. Though our approach does not include any visual similarity or object recognition features in the machine learning sense, certain genres do appear more popular within certain regions of the space, depicting features yielding a similar complexity profile (e.g. human half-figures on dark background; Fig. 1.B.2).

The model used in this contribution consists of 112 transformations (see Fig. 1.A for examples; Table S1 and Figure S1 in the Additional file 1 for the full list; we use lossless GIF, and also PNG and lossy JPEG on a smaller subset). The exact nature and number of transformations is unimportant yet subject to optimization. More features increase computation time, but provide more information, i.e. the approximation honing in on the true uncomputable Kolmogorov complexity (see Methods). As demonstrated in the classification experiments in the Methods section, different transforms are informative for different tasks, and, given a specific task, a handful of well-chosen features can yield accuracy close to using a large ensemble. Some transformations, like blurs of various magnitudes, may well correlate. In applications where multicollinearity must be avoided or a lower number of dimensions is desired, methods such as Principal Component Analysis (PCA)

or UMAP [49] can easily be applied. In this work we use both, as PCA is directly interpretable due to linear relationship with original variables, while UMAP arguably offers better low-dimensional representations [49].

Dimension reduction like UMAP can be used to produce a complementary “field of similarity” [50] where similar images cluster together intuitively, while remaining subject to the “curse of dimensionality” (Fig. 1.B). We can break the curse by repeatedly mapping individual transformations onto the common UMAP, effectively using it as a reference topography (Figs. 1.C and S2). The resulting “small multiple” of visualizations provides intuition why an ensemble of multiple transformations is necessary towards a fuller understanding of visual aesthetic complexity. It also relates to the explainability aspect: images being in different ends of a given transformation variable (colored brown to blue in 1.C) indicates they differ in the aspect of complexity represented by the given distortion.

In this contribution, we define and test the compression ensemble approach in a number of experiments from human complexity judgments to example downstream classification tasks (see Methods section). In the Results section, we showcase the utility of the approach for the exploration of art collections, quantifying global trends in historical art over the past six centuries in a large dataset, and the first half a year of a non-fungible token (NFT) art marketplace. On historical timescales, we introduce a temporal resemblance model to quantify artistic career trajectories, grouping them into qualitatively distinct types. We reveal artists that were well embedded in the historical tradition of their time, those who simultaneously experimented with different areas of the aesthetic complexity space, artists with transitory success, and those who were later seen as ahead of their time. Finally, we discuss the broader relevance of the approach to digital art history, cultural evolution, and extensions to other media and modalities.

## 2 Results

We make use of two large art corpora to proof the application of the compression ensemble approach for visual data, while exemplifying the exploration of historical and contemporary dynamics of visual art. The first dataset which we denote as “Historical” (henceforth capitalized when being referred to) is illustrated in Fig. 1.B. It is sourced from the art500k project [42], filtered to only include two-dimensional art with a retrievable year of creation. Our subset contains 74028 (primarily Western) artworks representing 6555 artists from the years 1400-2018 (older art exists in the dataset but is sparsely distributed). This filtered dataset ends up consisting mostly of items art500k had in turn sourced from Wikiart.org. The latter is an online, user-editable, encyclopedic collection of mostly Western art images, also frequently used in computer vision research.

In the case of art collections or databases like Wikiart and art500k, it is important to be clear that these consist of small, curated, often biased samples of art of some place and period. As such, they represent the historiography of art first and the actual history of art second [12] (see Data Limitations in the Methods section for further discussion). When we make claims here about the history or dynamics of visual art, we are only referring to information derived from the sample — but we make the assumption that the sample is reasonably representative and as such informative of the population of Western art in the time periods we cover. This means that the figures depicting historical changes may look different if more data would be available. However, our quantitative method could also be used for systematic study of data set bias.

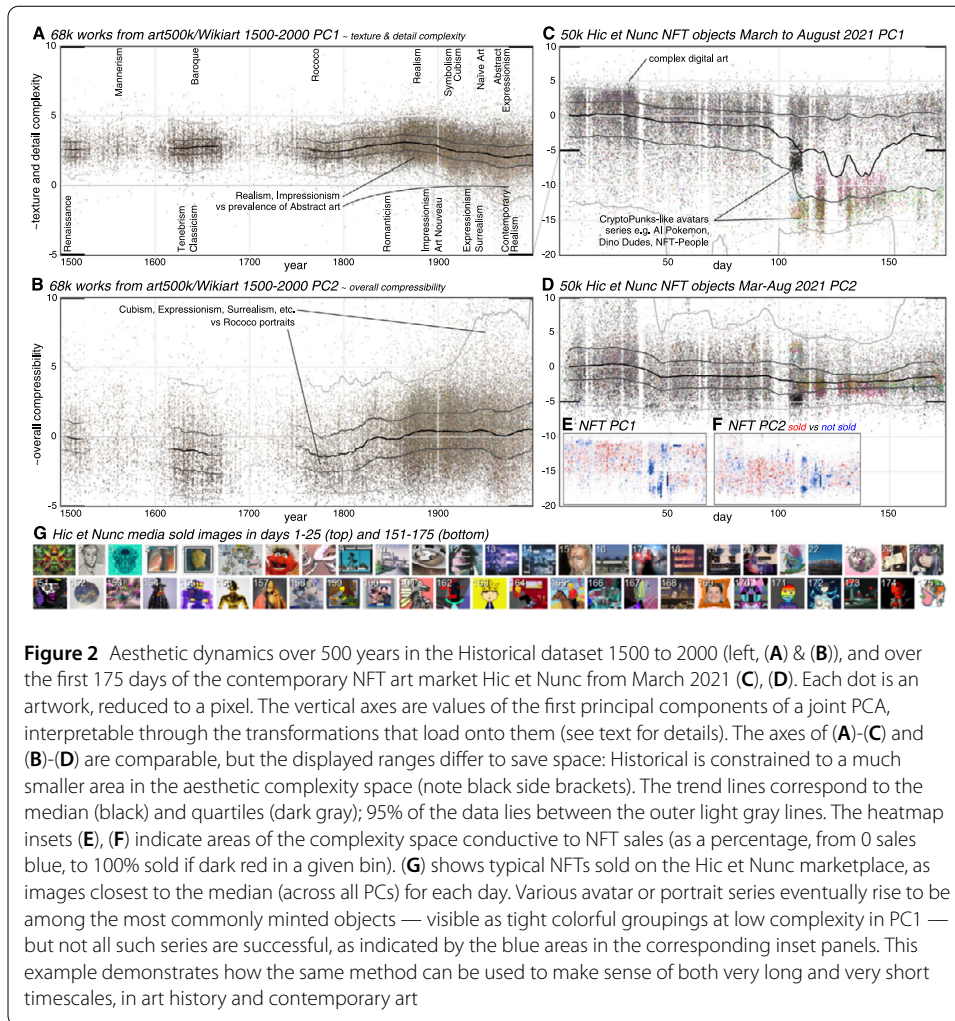
The second dataset, denoted “Contemporary”, is mined from Hic et Nunc, a Tezos blockchain-based NFT art marketplace, representing the first 175 days of its existence (March to August 2021; 51640 artworks, 7284 artists). It contains 31% of all the objects added to the marketplace during our observation period. We only include static images (JPEG, PNG), as the currently presented approach does not yet extend to multi-frame objects such as animated GIFs and videos. We also exclude low resolution images (such as icons), and a subset for which the data collection process failed to retrieve the image. Unlike the Historical dataset which consists of digitized art, the vast majority in the Contemporary set are born-digital images. For an overview of the NFT-driven “crypto art” market, see references [51, 52].

The obvious question is whether simple vectors of file size ratios are provably informative about visual complexity or artistic aesthetics. We carried out two sets of experiments to verify that the compression ensembles are fit for this purpose, and show that they allow us to meaningfully track the evolution of complexity in historical art. As detailed in the Evaluation subsection under Methods, we first tested the approach against two sets of human visual complexity judgments, where it performed very well, aligning with what people would judge to be visually simple or complex. This indicates the method is cognitively plausible as a visual complexity estimator. Secondly, we devised a set of classification experiments to determine if there is enough information in a compression ensemble to meaningfully delineate aspects of interest in visual art such as style, genre, authorship and medium. The model yields reasonable accuracy in all cases (and its mis-classifications make sense from an art historical point of view). This indicates the method is fit for purpose for the exploratory tasks showcased in the next section.

## 2.1 Tracking historical and contemporary art dynamics

Given the explainable nature of compression ensembles, and its demonstrable cognitive and technical plausibility, we proceed to use the method to investigate and interpret aesthetic trends over time. We do this for both the Historical and the Contemporary NFT datasets. To simplify this task, we apply Principal Component Analysis: compression vectors for both the Historical and the Contemporary datasets are fitted in the same PCA space for comparability. An alternative would be to map change over each individual transformation variable, but PCA conveniently allows for focusing on decorrelated latent aspects with most variance, while remaining interpretable through the transformations that load onto each component. Figure 2 depicts change over time in the two first most informative components. The Historical dataset is limited to 1500-2000 on the graph, as both ends outside of that range are quite sparse. The “trend lines” are estimates from a rolling window of  $\pm 10$  (years of Historical data, days of Contemporary Hic et Nunc). Where there is insufficient data, the window is stretched up to size 50 to include at least 1000 artworks where possible; these broader estimates are reflected by decreased line opacity. We do not engage in statistical testing here, as this exercise is explorative (for dating and style classification testing, see Evaluation in the Methods section).

Changes in the trends of the half-millennium Historical dataset correspond broadly to art historical style classifications. PC1 in this model corresponds to texture and detail complexity (loading onto blurs, despeckle filters, and the Canny edge transform). A set of the more frequent style period labels are shown in Fig. 2.A, arranged by the median year of the respective artworks. Visible in the right half of (A), there is a visible median complexity decrease between the period of detailed paintings of Realism and Impressionism and



**Figure 2** Aesthetic dynamics over 500 years in the Historical dataset 1500 to 2000 (left, **(A)** & **(B)**), and over the first 175 days of the contemporary NFT art market Hic et Nunc from March 2021 (**(C)**, **(D)**). Each dot is an artwork, reduced to a pixel. The vertical axes are values of the first principal components of a joint PCA, interpretable through the transformations that load onto them (see text for details). The axes of **(A)**–**(C)** and **(B)**–**(D)** are comparable, but the displayed ranges differ to save space: Historical is constrained to a much smaller area in the aesthetic complexity space (note black side brackets). The trend lines correspond to the median (black) and quartiles (dark gray); 95% of the data lies between the outer light gray lines. The heatmap insets **(E)**, **(F)** indicate areas of the complexity space conducive to NFT sales (as a percentage, from 0 sales blue, to 100% sold if dark red in a given bin). **(G)** shows typical NFTs sold on the Hic et Nunc marketplace, as images closest to the median (across all PCs) for each day. Various avatar or portrait series eventually rise to be among the most commonly minted objects — visible as tight colorful groupings at low complexity in PC1 — but not all such series are successful, as indicated by the blue areas in the corresponding inset panels. This example demonstrates how the same method can be used to make sense of both very long and very short timescales, in art history and contemporary art

the second half of the 20th century where (in this dimension less complex) styles such as Abstract Expressionism and Pop Art become more prevalent.

PC2 corresponds to overall compressibility (loading onto compression of the original unfiltered image but with different compression algorithms). The median in the Historical dataset is somewhat lower where the dataset contains many Rococo style portraits (in the middle of Fig. 2.B), which typically contain plain (easily compressible) backgrounds — not unlike the pixel-art portraits of Hic et Nunc (cf. days 100-150 in Fig. 2.C-D). PC2 values in Historical (Fig. 2.B) go up around the onset of Impressionism, and the bounds are pushed once more with Cubism, Expressionism, Surrealism, and the general diversification of classic modern “-isms”. As demonstrated in the Evaluation section in Methods, given a sufficient number of transformations, such differences are consistent and diverse enough to predict style periods with reasonable accuracy.

The Historical and Contemporary sets are combined in the same space, but the ranges of the vertical axes representing the components in Figs. 2.A-B versus C-D are intentionally different, as the two datasets occupy markedly different ranges in the complexity space, with much higher variance in the Contemporary Hic et Nunc dataset compared to the more conventional Historical dataset. This does not necessarily mean that art in the last 500 years has been less creative or explorative. The relative boundedness instead is more

plausibly rooted in a combination of material affordances and limits of curation and scholarship. The latter is a function of cultural selection, as collectors, audiences, and art historians put a bound on what has been and is considered worthwhile of adding to collections from the time of creation to current retrospectives.

In contrast, anybody who is able to pay the fairly low “minting” fee can upload an artwork to blockchain art market places such as Hic et Nunc, making their creations public in an attempt to get attention and sell. The Historical broadening of the parameter space goes in lockstep with the fraction of noted creatives growing faster than world population in the last five centuries [53]. It is broadly established knowledge in art history that new technologies and concepts, from pigments to theories of perception [54], were harnessed by said creatives. Examples include the emergence of more affordable blue pigment alternatives to the rare and expensive azurite and lapis lazuli, or (color) photography, which put traditional pictorial conventions of depiction into question. Another striking difference between the Historical and the Contemporary NFT dataset becomes visible in Figs. 2.A-B versus C-D when we focus on the range of colors in the single-pixel reductions of the artworks. The digital NFT images appear darker and more saturated, as they are using the full RGB color space, while the dominant color of Historical artworks tends to remain in the range of “natural” pigments, which one could buy in a physical art supply store.

Since we have information on transactions in the Hic et Nunc dataset (as of the data collection time, 22 August 2021), successful sales are shown as inset heatmaps (E, F) in Fig. 2.D. The heatmaps show the fraction of sales across the first and second principal components respectively. About half the objects in the Hic et Nunc sample in total were sold off by their authors during our observation period, with some areas — dark red in the insets — being clearly more conducive to sales, while others do not sell at all.

Even qualitatively, one can see revealing patterns, such as the mass-minting of initially non-selling NFT portrait images starting around day 110 in mid 2021. These can be described as simple, typically procedurally generated, mugshot or portrait-style images depicting various human, humanoid or cartoon characters. At the height of the NFT boom, the perhaps more widely known examples of this trend on other platforms included the “CryptoPunks” and “Bored Ape Yacht Club” series [51]. On Hic et Nunc, such series are titled for example “AI Pokemon”, “Dino Dudes”, and “NFT-People”. An emergent quality of these mass-produced images is that their texture and detail complexity (PC1, Fig. 2C/E, day 100-150) is substantially lower than the all preceding art, putting them more in the realm of icons or brand logos. At the same time their overall compressibility (PC2, Fig. 2D/F) is not only systematically lower, but also subject to less variance. This could potentially indicate low effort attempts at production of these simple images with hopes of fast monetary gain in the marketplace. The narrowness of the mass-produced NFT series also expresses itself in their skew towards highly saturated primary dominant colors. In at least one case, this indeed seemed to work, where sales follow in the wake of a strong minting burst, mostly consisting of the “NFT-People” and “NFT Kids” series (cf. the rightmost vertical blue line in Fig. 2.E, followed by a light red wake).

These initial observations could of course be augmented with more systematic statistical or predictive modeling in future research. As an example, we trained a simple classifier, Linear Discriminant Analysis (see Evaluation section under Methods) on the sales data, predicting whether an NFT art piece was sold or not, based on the values in the compression vectors. Using training sets of size 20k per class and separate test sets of 5k (and



replicating the model 500 times), the model predicts sales at an average accuracy of 58% (or a 17% kappa, given the 50-50 baseline). This is despite containing no information on the prestige or reach of the artists, past sales, the depicted content, nor trends of the market of the respective time. A linear regression model fitted to 23370 sold items predicting log price (excluding zero-price giveaways) by all the compression variables describes about 6% of variance (adjusted  $R^2$ ); allowing for interaction with the time variable improves this to 8%. While these are all fairly low scores in absolute terms, we consider this a promising result for future research, likely improved by combining our aesthetics model with the aforementioned variables of author properties, sales history and past trends [44, 52], to predict future trends in evolving art markets.

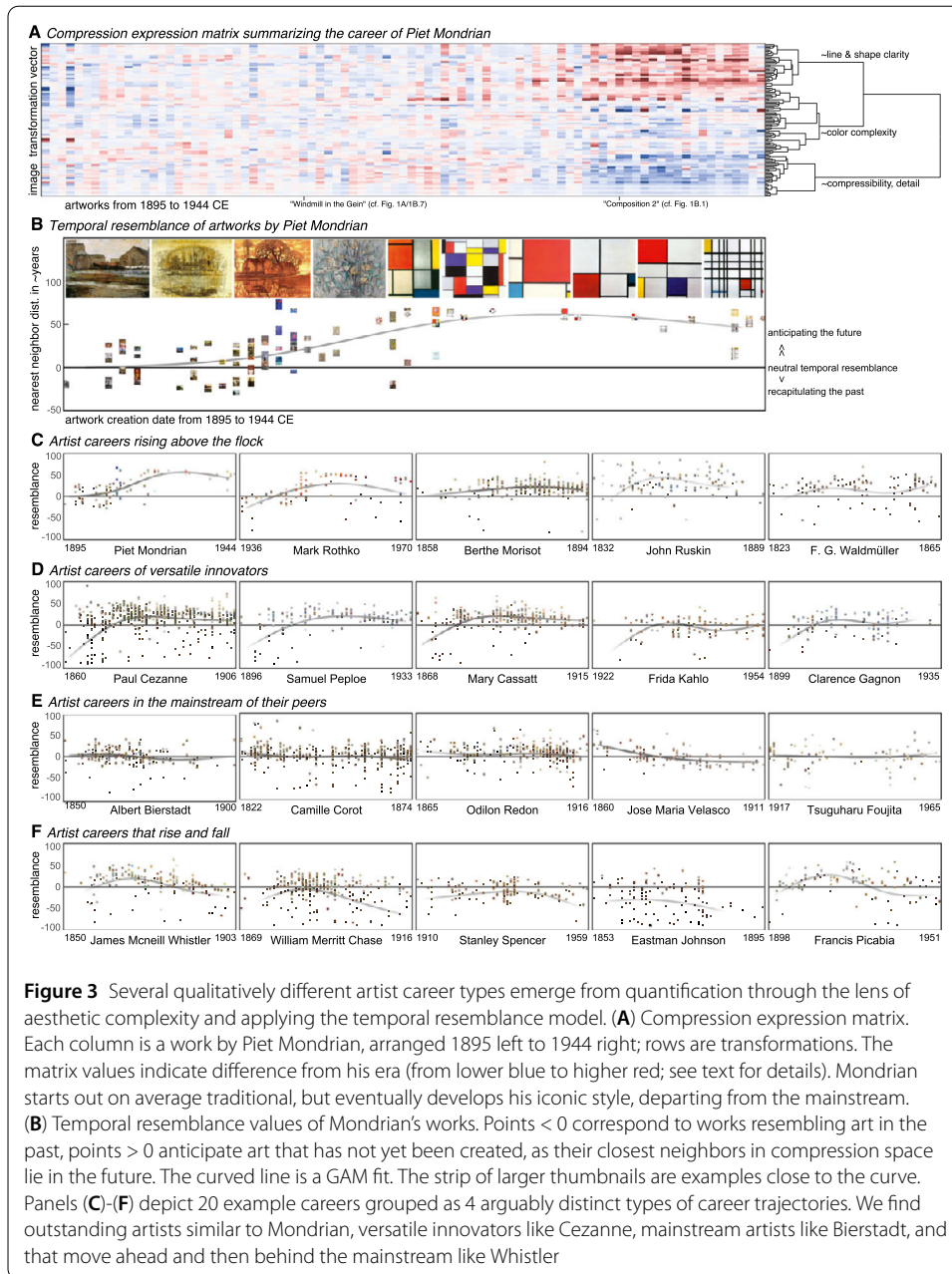
## 2.2 Quantifying temporal resemblance in artistic careers

In the previous section, we took a look into art history as a whole. We can also use compression ensembles to investigate how the oeuvres of individual artists progress and are situated in their eras. Tracing “the lives of the artists” has been a foundational and central direction in the historiography of art since the 1550 book by Giorgio Vasari which initiated the genre [55], followed by a great number of artist monographs and critical catalogs. More recently, multidisciplinary science has tackled the issue using methods of network science and quantitative measures of success, making use of data such as demographic and migration records [53, 56, 57], museum, exhibition and art market price information [58, 59], but also visual aesthetic aspects using information theory or machine learning [7, 12, 60].

We introduce “temporal resemblance”, a simple metric to summarize and compare artistic careers (Fig. 3). This could be applied to any numeric space (including deep learning embeddings) that includes temporal metadata, but its interpretation of course depends on the space. Here, it represents resemblance in the compression ensemble approximated aesthetic complexity space.

Figure 3.A provides an illustration, using the oeuvre of Piet Mondrian: the rows are transformations; the columns his works, arranged diachronically. For this example, the compression ratio values are z-scored using the mean and standard deviation of his era. This means that the blue to red scale is interpretable as “higher or lower than contemporaries”, and white as being close to the mean. Mondrian’s departure from the era’s mainstream is quite clear from the increase of darker reds and blues. The transformations provide insight as to which areas are most contrasting (labels on the dendrogram in 3.A).

The temporal resemblance method generalizes this comparison. Given the vector space of the Historical set (decorrelated using PCA), we can calculate the nearest neighbors for each artwork vector (like the columns in 3.A). We use cosine similarity and the top closest 100 neighbors (excluding works by the same artist). The median temporal distance between these neighbors and the target work indicates if it resembles the past or anticipates some yet unseen future. This allows us to group artists who are traditionalist or historicist, those who stay current, and those ahead of their time. It is necessary to adjust the median time distances to account for the boundedness and density bias of the dataset: the metric reported here is derived from the residuals of a generalized additive regression model (GAM), still on the same yearly timescale (see Additional file 1 for technical details of the adjustment). Figure 3.B again depicts Mondrian’s works, with temporal resemblance now on the vertical axis.



**Figure 3** Several qualitatively different artist career types emerge from quantification through the lens of aesthetic complexity and applying the temporal resemblance model. **(A)** Compression expression matrix. Each column is a work by Piet Mondrian, arranged 1895 left to 1944 right; rows are transformations. The matrix values indicate difference from his era (from lower blue to higher red; see text for details). Mondrian starts out on average traditional, but eventually develops his iconic style, departing from the mainstream. **(B)** Temporal resemblance values of Mondrian's works. Points < 0 correspond to works resembling art in the past, points > 0 anticipate art that has not yet been created, as their closest neighbors in compression space lie in the future. The curved line is a GAM fit. The strip of larger thumbnails are examples close to the curve. Panels **(C)-(F)** depict 20 example careers grouped as 4 arguably distinct types of career trajectories. We find outstanding artists similar to Mondrian, versatile innovators like Cezanne, mainstream artists like Bierstadt, and that move ahead and then behind the mainstream like Whistler

This metric is relative to the point in time of each work, and all measures are relative to all other works. Therefore, curves that stay close to the zero line in Figs. 3.B/C should be interpreted as artists who produce works that are similar to other artworks made in the same years, in terms of aesthetic complexity (and thus aspects of their style — but we are not quantifying the subjects they depict). That does not preclude changes in their style, if the changes in the artist and their era correlate. Staying around the zero may also indicate that a given artist is surrounded by a handful of prolific contemporaries with very similar output, who as a group may not be representative of the mainstream. Descending curves can indicate an artist who becomes more traditional, the world catching up to an artist's style, or the world adopting other new styles.

Figure 3.C depicts the careers of 20 artists, grouped by career trend similarity, revealing different modes of artistic existence, similar yet not identical to the narrative types of Vonnegut [61]. To trained art historians (as represented among the authors of this paper) these results are intuitively correct — but of course the career comparisons discussed here only refer to artworks that are present and dated in our dataset. Some such as Piet Mondrian or Mark Rothko, “rise above the flock”, starting out in the mainstream, growing into their own distinct style, with works that could be considered ahead of their time. Paul Cezanne and Mary Cassatt instead become “constant innovators”, starting out by producing conventional, retrospective works, but growing and remaining innovative throughout the rest of his careers. Albert Bierstadt and Camille Corot represent “mainstream artists”, appearing more narrow in their practice, and remaining consistent with the current of their peers. There are also those who “rise and fall” (see James McNeill Whistler or William Merritt Chase), growing to their moment in history, then becoming more conventional again over the course of their careers. The metric highlights Eastman Johnson, who was predominantly drawing inspiration from the past, even at the height of his career. But even innovative careers may include revivals on occasion; e.g. Paul Cezanne also has works resembling the complexity profiles of art preceding his own by 100-200 years.

As a static graph, Fig. 3 just exemplifies how compression ensembles (or other embeddings) can be used to filter and cluster artistic trajectories. Figure S5 provides an alternative version of Fig. 3.B/C with larger thumbnails. Interactive versions of such plots could function as a research instrument for qualitative experts, to investigate the quantitative model and dataset biases, and compare artists between different datasets. In a related project [62], we developed an interactive web interface titled Collection Space Navigator, which allows for on-the-fly visualization of large collections, and operations such as zooming, filtering, and hovering for more information. Its online demo also features a large subset of the Historical dataset used here, including functions to operate with compression ensembles (link in the Availability of data and materials section below).

### 3 Discussion

Products of human culture, such as art, language and music are all subject to ongoing change, complex dynamics, and cumulative evolution [63–66]. And even though complexity could emerge from a simple generating mechanism in principle, a single unidimensional measure would likely prove insufficient to capture the polymorphic complexity of human cultural interaction and cultural products [67]. Here, we have demonstrated the utility of explainable compression ensembles to quantify polymorphic visual aesthetic complexity. We showed how the approach can recover and reveal meaningful patterns in datasets of historical and contemporary art, and, in the methods section, we evaluated the cognitive plausibility of our approach, tested its viability at author, date, style, genre, and medium detection tasks. Given the increasing availability of cultural datasets in machine-readable form, this operationalization opens up new avenues to study the dynamics of visual art aesthetics at scale, over long time spans and almost in real time. As such, the approach may help to transcend the still considerable specialization and bifurcation (by artist, period, style, etc.) of qualitative art historical scholarship. Our approach could be used to fill a similar niche in art history as computational corpus linguistics does in relation to the qualitative study of literature.

### 3.1 Cultural evolution and aesthetic value

While we focus on complexity, the same approach could be used on other related phenomena. For example, Sinclair et al. [68] raise the concept of “aesthetic value” or “attractiveness” of a given cultural product, to discuss whether art and music indeed could be considered as products of cumulative cultural evolution [66] or not, as “cumulative” would allude to objective improvement over time [69]. From the art historical perspective, a style that builds on or grows out of another style may not be necessarily objectively better, but may better meet the preferences of its consumers in a given time, place, or ecological niche. This is not unlike the concept of communicative need in language: a structure or lexical configuration might not be better in some absolute evolutionary terms, but can be more optimal or efficient given the usage tendencies or needs of a given language community [70, 71].

Our method goes beyond discrete categories and facilitates studying visual culture using a continuous form of aesthetic representation to examine such questions, as exemplified in the temporal resemblance section above. Naturally, the extent that need or preference can be studied depends on available data. E.g. the Historical set represents only a rough estimate of a (primarily Western, biased, somewhat dated) preference consensus, while the Hic et Nunc data includes artist, collector, trade and price info, which could be (carefully) interpreted as preference, as well as linked with the (social) media activity of the sellers and buyers.

### 3.2 Family resemblance and connoisseurship

The vector space of the ensemble allows for flexible operationalization and visualization, e.g. a single figure (e.g. 3) can summarize careers of several artists. This is not entirely dissimilar to the intuition of an human connoisseur trained on a given corpus of art. As each transformation in an ensemble represents a tangible visual aspect e.g. abundance of detail or colorfulness, as a whole, it constitutes an estimate of the philosophical and cognitive concept of polymorphic family resemblance, originally used to characterize similarity of games such as chess and soccer, later extended to polymorphic visual perception [72–74]. The recognition of visual family resemblance is arguably foundational and intrinsically mastered by trained human art connoisseurs, yet “exists at an unarticulated level, easy to invoke but difficult to explicate” [3]. Such recognition skills are also required of other visual experts e.g. radiologists, detecting health issues in medical imaging. Deep learning models have made headway in solving the latter among other object recognition tasks, but are not very good at explaining why or how they recognize something either (though cf. [29]).

Our explorations of the Historical dataset yield results that meaningfully reflect the art historical scholarship underlying the dataset (Figs. 2, 3). The model captures enough family resemblance to cluster together similar styles or works by the same artist (as verified in Methods), demonstrating the explanatory power of using an ensemble of multiple transparent transformations, effectively addressing what Friedländer in his foundational art connoisseurship book [75] called “the visible in its manifoldness and unity, bristling against conceptual segmentation, so that the boundaries between the species of images get into flow” or get “blurred” (p. 60; our translation).

As an interesting byproduct, the continuous multidimensional ensemble space of compression ratios also allows for mathematical vector operations (not unlike in word embeddings [76]) and explainable latent space exploration. Adding the vectors of Example 4 and 6

of Fig. 1.B, an etching or woodcut print plus the “Garden of earthly delights”, yields a vector where the closest neighbors are fittingly prints of trees and nature (cf. Figure S3). Multiplying a Mondrian vector of Example 1 with an averaged vector of all landscape paintings nets an abstract landscape. Exploring these operations could be an interesting venue of future research.

### 3.3 A window to cultural meaning space

We used three kinds of vector spaces: the full multidimensional ensemble space of compression ratios, the decorrelated multidimensional space of associated PCA components, and a dimension-reduced UMAP space providing a proxy topography. These can also be understood as subspaces of more general cultural meaning spaces, and interpreted though or used in various other approaches to culture, briefly exemplified below, with potential to be developed further in future work on quantitative aesthetics. In Cassirer’s most general reference framework they would be “spaces of geometric intuition” [77–79], also later referred to in art history as “iconologic” aspects, complementing the associated contextual information including “iconographic” aspects [80]. They resonate with the conceptual spaces theory of Gärdenfors [81, 82], as well as with the notion of information space [83]. Future work could also look into resonance between our and the recent deep learning driven “distributed information bottleneck” proposal [29] which also involves visual transformations in the case of images.

### 3.4 Extensions to other visual media, audio, video and text

Here we focused on static, 2D art such as paintings and digital drawings. However, there is no reason the same methodology could not be applied to quantify other visual media such as photographs, maps [11, 17], websites [84] or natural patterns [27] to assess their aesthetic complexity (and by proxy, style) in a transparent, explainable framework. Multi-frame visual media such as film and animation could be split up by frame or shot, and represented in a compression ensemble as ordered sets of vectors, or alternatively, transformed using video filters and compressed directly using video compression algorithms. Similarly, the aesthetic complexity of sound and music [85, 86] could be inferred by either applying the same visual transformations to spectrograms, or by using audio filters as the transformations followed by audio compression. 3D objects such as sculptures, architecture, or clothing can be operationalized by systematically scanning them from multiple angles, or using 3D versions of transformations and compressions (voxels instead of pixels). For written text, an ensemble of byte pair encoding [87] models (with variable parameters or trained on different genres) could be used as the transformations.

### 3.5 Combining compression ensembles with other vector spaces to study culture at scale

For multi-modal media, multiple ensembles or embeddings can be concatenated, provided a principled way to weigh or normalize their contribution [88]. If both feature similarity and aesthetic complexity matter in a given application, a compression ensemble could be horizontally aligned and concatenated with a suitable deep learning embedding [42], an approach shown to be fruitful in NLP [89]. A scene in a film or a recorded theater play could be represented by the concatenation of a visual compression ensemble, an audio compression ensemble, an image embedding, and a language model embedding

[90] of the spoken dialogue. A compression ensemble could also be combined with the larger apparatus of art history, in the form of socio-cultural context information from relevant databases and knowledge graphs [91] or any other visual feature vectors [13, 16, 92]. Furthermore, features or objects (e.g. humans) could be extracted using a deep learning classifier, followed by a compression ensemble of these sub-images, producing e.g. an aesthetic vector space of human pose, effectively a further operationalization of Aby Warburg's Mnemosyne [93, 94].

Understanding cultural products at scale is not only relevant just because of the growing body of born-digital culture and digitization efforts of non-digitized culture. We stand at the threshold of an AI revolution, where the fully automated generation of photo-realistic, artistic and otherwise previously primarily human-produced visual content (and soon likely multimedia too) has suddenly become feasible, accessible and affordable, using e.g. models like Dall-E or Stable Diffusion [95]. This is likely to transform multiple entire industries, but the functionally near-infinite content quantity will require curation and understanding to be made efficient use of. Pretrained deep learning embeddings can be used to calculate and cluster similar items in a feature space, or be trained to predict preset objects, styles or human preferences. Our explainable compression ensembles are however well-positioned to *make sense* of such spaces and navigate aesthetics, without requiring any “training” — indeed, consisting of already meaningful values, they are not constrained by the need to predict anything specific to be functional.

## 4 Methods

### 4.1 Constructing a vector space of algorithmic distance

As discussed in the Introduction, compression as such has been used to estimate visual and aesthetic complexity before. In some applications, it has also consisted of or included combination with limited visual transformations [5, 24, 25, 27–29, 40, 44]. However, the fairly large number of transformations is key to our approach, with the following rationale.

Consider two algorithmically similar uncompressed images  $A$  and  $B$ , for example two versions of the same famous view of Rouen cathedral by Claude Monet (of which the artist painted more than 30 in 1892/93). These two images will yield similar compressed sizes for the same compression algorithm because the “algorithm” that generated them (being a function of Monet's perspective, style, and execution) is similar. Another artwork  $C$ , e.g. a late, abstract work by Piet Mondrian will, due its lack of detail, likely have a much smaller compression size. However it is entirely conceivable that a work  $D$  that is stylistically very different to Monet's Rouen cathedral, e.g. a surrealist painting by Salvador Dali, might *by chance* have a very similar compression size. The “algorithms” used by Monet and Dali differ greatly, and an equal compression size does not imply that they are of equal algorithmic complexity either, as the efficiency of the compression algorithm itself will differ depending on the detailed characteristics of the images.

However, now consider an image transformation  $T$  (e.g. Gaussian blur), which we apply to the uncompressed versions of our four images  $A$ ,  $B$ ,  $C$ , and  $D$  before compressing them. The compressed sizes of  $T(A)$  and  $T(B)$  are still likely to be very similar, as the algorithms that generated the original images are very similar, and the transformation and compression algorithms are identical.  $T(C)$  is very likely to still be very different to  $T(A)$  and  $T(B)$ . While the compressed size of  $D$  was similar to  $A$  and  $B$  by chance, it is much more unlikely that  $T(D)$  is also similar to  $T(A)$  and  $T(B)$ , as the interaction between the transformation

$T$  and the generative algorithm of  $D$  would have to change the compressibility in the same way as the interaction of  $T$  and  $A/B$ . Put more intuitively, a Gaussian blur is very likely to affect the compressibility of a Monet very differently from the compressibility of a Dali. Thus, more generally speaking, two images with similar compressed sizes are much less likely to still yield similar compressed sizes by chance after a transformation *unless* they are algorithmically similar to start with in which case the combined algorithms of generation and transformation (and their interaction with the compression algorithm) remain similar. If we now consider the application of  $N$  different transformations of an uncompressed image  $A$ , each applied before a subsequent compression, the compressed sizes (including of the untransformed image)  $c(A), c(T_1(A)), c(T_2(A)), \dots, c(T_N(A))$  form a vector  $\mathbf{v}(A)$  of length  $N + 1$ . It follows from the above argument about coincidental proximity that it becomes increasingly unlikely for two algorithmically dissimilar images to remain close together as  $N$  increases. Thus the resulting vector space of compressed sizes provides an indication of algorithmic distance between images.

## 4.2 Data processing

In practice, we use normalized compression lengths. The compression size of the original image without transformations is divided by the size of the original bitmap image. Compressions of transformations are divided by the size of the original compression. In most applications discussed in this paper, it also makes sense to rescale the vector space components (we use z-scoring), to put the compression ratios and the additional statistical transformations (fractal dimension, colorfulness metrics, etc) on a comparable scale. For more technical details on the processing pipeline, list of the transformations and implementation, see Additional file 1.

Both the Historical and Contemporary Hic et Nunc dataset are preprocessed the same way, downscaling images to 160,000 pixel bitmaps ( $400 \times 400$  in the case of a perfect square) while retaining aspect ratio. Smaller images up to 50% of that size are allowed (but not upscaled), smaller images are discarded. Another option would be to resize all images to identical squares, but that would distort the composition of wide or tall artworks. The aspect differences, size differences resulting from integer division of the 160,000 and the inclusion of smaller images, are all controlled for in the next step. The assigned file size of a compressed image (or its transformation) is actually the mean of two compressions, of the original and its 90 degree rotation. The compression ratios are calculated in terms of the respective downscaled bitmaps. Furthermore, one of our visual transformations is the Fast Fourier Transform; given its square-shaped output components, the transform is applied twice, on the original and its rotation, and the resulting components are also additionally rotated for compression.

## 4.3 Data limitations and biases

This approach to homogenizing the images is far from perfect, as the size of the originals that these photographs and scans represent may well range from the size of a postcard to that of an altar piece. Not only that, but the latter may well be represented by a lower resolution image than the former, with better or worse color grading, etc. The dataset contains sparse metadata on original size and we have no way to systematically quantify this issue at scale, which remains a limitation of the current study. However, in a sense, our approach is not very different from making art historical inferences by going through and looking at

large visual resource collections, much like students of art history examining art historical survey literature, or an art connoisseur training their eye using a large comprehensive 35 mm slide collection of a library of photos, which historically served exactly this very purpose.

Since we are interested in making comparisons over time, the Historical dataset was also filtered for items with an identifiable creation date. We carried out some preprocessing of the date metadata, retrieving four-digit years from descriptions that included them. However, much of earlier art is tagged with heterogeneous and approximate descriptions such as “early XVI century”. Discarding these made the earlier end of this dataset even smaller, which is why we limit some analyses to the 19-20th century.

Our Historical dataset is also likely biased in a number of ways. It features primarily Western art, most of the data is concentrated in the 20th century, the metadata quality varies and is of unidentifiable origin, the sampling mechanisms are unknown but likely biased by archival and selection practices of the various museums and collections in which these reproductions originate and the websites that house them. Still, from an art historical standpoint the dataset provides a reasonable and sufficient proxy benchmark to show the feasibility of our approach. Known biases of the Historical dataset include reliance on partially dated literature, including a corresponding gap of 18th century art, and very likely some variation in terms of reproduction quality due to the broad variety of the crowdsourced images, either found in the public domain or taken from a great variety of literature and online sources on the basis of fair use. Digitizing larger amounts of visual cultural heritage in high resolution, consistent quality, and minimal bias is a generational challenge. While the Historical dataset is sufficient for our proof of concept, as more data in better quality becomes available, descriptions based on our method are expected to also become more precise and representative.

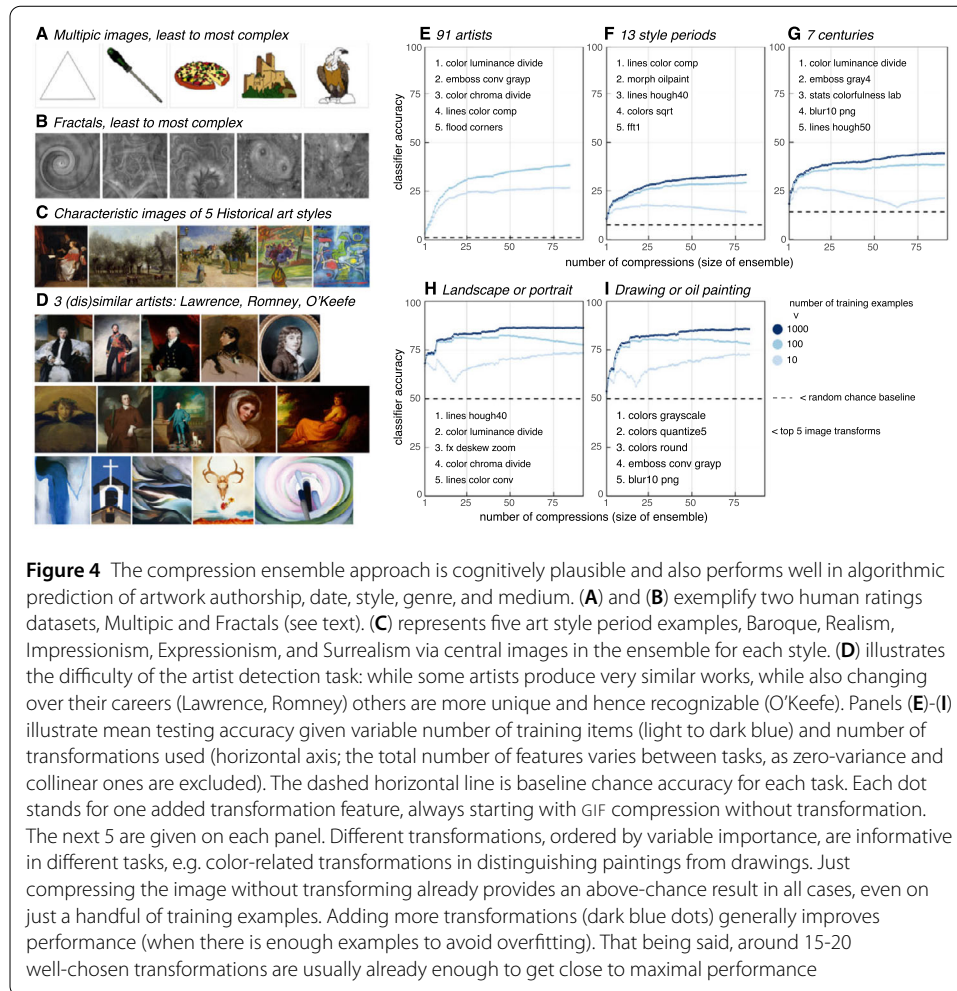
#### 4.4 Evaluation

We evaluate the compression ensemble approach extensively using three datasets and two methodologies, (1) examining correlations of our model predictions with human judgments of visual complexity, (2) using the model to perform authorship and style attribution. We show that our model performs very well on the first task and with fair accuracy on the second task (despite not being trained for the specific purpose). The second experiment also demonstrates explicit connections between specific dimensions in the vector space of compression ratios and particular aspects of the corresponding artworks. For example, the compression ratio of edge-filter transformations are informative regarding the genre of the work (portraiture versus landscape), while color-affecting transforms can help predict the medium (drawing vs oil painting).

##### 4.4.1 Human complexity norms

We assess the cognitive plausibility of the compression ensemble approach by comparing its predictions of visual complexity with human judgment norms from two datasets. The first dataset, MultiPic [96] consists of 750 colored pictures of concrete concepts, and human judgments on various aspects of visual perception, including complexity, based on experiments with 620 participants from six language communities (British English, Spanish, French, Dutch, Italian, German; see Fig. 4.A). The dataset includes means for each image for a given language sample. The second dataset, Fractals [26] consists of 400 abstract fractals and related norms, again as mean judgments of visual complexity, by 512





**Figure 4** The compression ensemble approach is cognitively plausible and also performs well in algorithmic prediction of artwork authorship, date, style, genre, and medium. (A) and (B) exemplify two human ratings datasets, Multiple and Fractals (see text). (C) represents five art style period examples, Baroque, Realism, Impressionism, Expressionism, and Surrealism via central images in the ensemble for each style. (D) illustrates the difficulty of the artist detection task: while some artists produce very similar works, while also changing over their careers (Lawrence, Romney) others are more unique and hence recognizable (O'Keefe). Panels (E)-(I) illustrate mean testing accuracy given variable number of training items (light to dark blue) and number of transformations used (horizontal axis; the total number of features varies between tasks, as zero-variance and collinear ones are excluded). The dashed horizontal line is baseline chance accuracy for each task. Each dot stands for one added transformation feature, always starting with GIF compression without transformation. The next 5 are given on each panel. Different transformations, ordered by variable importance, are informative in different tasks, e.g. color-related transformations in distinguishing paintings from drawings. Just compressing the image without transforming already provides an above-chance result in all cases, even on just a handful of training examples. Adding more transformations (dark blue dots) generally improves performance (when there is enough examples to avoid overfitting). That being said, around 15-20 well-chosen transformations are usually already enough to get close to maximal performance

German-speaking participants (Fig. 4.B). Previous research has engaged in analogous exercises of evaluations against human complexity judgements [24, 28]. We use the datasets described here as they are both publicly available while representing fairly large pools of participants.

We generate the compression ensemble vectors separately for each of the two datasets, then carry out repeated out-of-sample evaluation where we train a linear regression model on a set of vectors to predict human scores, then test its accuracy on a separate test set. The results are very good, with median absolute error ranging from 0.19 (Multiple English) to 0.23 (Multiple Flemish) on a scale of 0 to 5. To put this in perspective, this is smaller than the differences between languages in this dataset (the median standard deviation of complexity scores per image across languages is 0.24). In Fractals, median absolute error is 0.46 on the same scale of 0 to 5. The linear regression model with compression ratios as predictors describes most of the variance (measured as adjusted  $R^2$ ) in human visual complexity ratings: 73% (Multiple Italian) to 83% (Multiple Flemish), and 32% in Fractals. By comparison, using GIF compression alone describes just 37 – 44% (Multiple) and 10% (Fractals). These results provide us with confidence that the approach is cognitively valid, correlating with what the human eye would consider visually complex.

#### 4.4.2 Artist, date, style, genre, and medium classification

The second evaluation involves the Historical dataset, in the form of a number of retrieval or classification experiments. We generate the compression ensemble vectors for the entire dataset, and extract the following subsets, where each included class has at least 1100 unique examples: 13 style periods as per metadata (5 of which are exemplified in see Fig. 4.C), 7 centuries, drawings vs oil paintings, landscape paintings vs human portraits, and 91 artists with at least 110 artworks each (3 of which are exemplified in see Fig. 4.D).

We perform out-of-sample evaluation where we repeatedly train a classifier for each subset, on a randomly sampled set of vectors from each class in the subset to predict the relevant class labels such as style period ( $n = 1000$  per class, except 100 for authors due to limited data), then test its accuracy on a separate test set ( $n = 100$  per class, except  $n = 10$  per author). We use Linear Discriminant Analysis — a simple, computationally lightweight supervised machine learning model that straightforwardly generalizes to multi-label classification. To probe how well the ensembles work on this task given different amounts of data and number of transforms, we carry this out in a step-wise manner, as depicted in Fig. 4.E-I. Each classifier is trained on 10, 100 and 1000 examples of each class, and employing an increasing number of transforms, starting from the baseline of GIF compression (ratio to raw bitmap file size). The rest of the features are ordered by a rough estimate of variable importance (derived from repeatedly training binomial logistic regression classifiers on all possible pairs of classes and averaging the t-statistics of the variables).

Even with a handful of examples and a couple of the most informative transforms, the simple classifier is able to detect above chance the creator, the date, style, genre, and medium of a given artwork. With a 100 examples and the full ensemble of transforms, author ( $n = 91$ ) detection accuracy is 38%, which is much higher than the accuracy of 1.1% that random attribution would achieve by chance. Provided 1000 examples per class, oil paintings are distinguished from drawings about 86% of the time, same for landscapes vs human portraits (both have 50% random chance baseline), style period 34% (baseline  $\sim 8\%$ ) and century 44% (baseline  $\sim 14\%$ ).

The ranking of the transforms beyond the compression baseline (as depicted in Fig. 4.E-I) is also informative. Some of the aspects represented by the transformations are more useful than others for the prediction task; for example, gray-scaling distinguishes pencil drawings from colorful oil paintings, because this is one of the primary aspects they differ in. Turning this around, the explainable features of the compression ensemble can be used to describe how any two images (or sets of images) differ, by looking into which transformation dimensions describe the most variance.

Inspecting the relevant confusion matrices reveals the errors are fairly systematic and intuitive, as classification errors are more likely between adjacent style periods and artists. In the set of 91 artists, Thomas Lawrence and George Romney are most often confused with each other by the model — and indeed, both are portrait artists from roughly the same period (see Fig. 4.D). Conversely, artists in a distinguishable style or genre are easy to identify, for example the 19th century engraver Charles Turner is detected at 97%. Rococo, also known as Late Baroque, is correctly labeled in 47% of the tests, while 16% of it is misclassified as Baroque. Impressionism is easiest to identify (53% correct) — but confused with Post-Impressionism (14%). Expressionism is by far the hardest to put a finger

on at 12% (cf. Figure S4 for the confusion matrix of style periods). These error structures are interesting in themselves, and could be investigated in future research.

It is important to be clear about the purpose of the experiments here: these are to verify that the compression ensembles approach, given a sufficient number of transformations, is capable of delineating art historically interesting aspects that may differ in aesthetic complexity. The results indicate this to be the case, and lends confidence to explorative findings in the Results section.

Deep learning models, which typically involve lengthy pre-training on large image databases, can also be tuned to perform similar tasks [9, 42, 97]. Our method is primarily a general zero-shot complexity estimation algorithm and does not involve pre-training; but indeed its output is informative enough to perform some such downstream tasks with reasonable accuracy. The purpose of this exercise here however is not to compete with these approaches, but to show that a compression ensemble — despite consisting of no features other than file size ratios and statistical transformations, and containing no pre-trained baseline — still captures and disambiguates enough family resemblance to place stylistically similar artworks close together and dissimilar ones apart, with a non-random error structure.

Nevertheless, below are some analogous results, to give a sense of machine learning accuracy in similar classification tasks using similar art corpora. These are however not directly comparable to ours due to training and test set differences. Mao et al. [42] report a 39% accuracy for style period and 30% for author retrieval; Tan et al. [41] report 55% for style and 76% for artist (but that is between just 23 artists with the most training data). If for example authorship attribution was the goal of a given application, we envision that its accuracy could likely be improved by concatenating image embeddings with compression ensembles, as mentioned in the Discussion.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1140/epjds/s13688-023-00397-3>.

**Additional file 1.** The attached supplementary information file includes further technical details on the compression ensembles algorithm, its implementation, and additional supporting results as referred to in the main text. (PDF 3.4 MB)

## Acknowledgements

We would like to thank Dr. Mikhail Tamm for helpful discussions. Thumbnail previews of artworks depicted for informative purposes as fair use.

## Funding

A.K., M.C.S., T.O., and M.S. are supported by the CUDAN ERA Chair project, funded through the European Union's Horizon 2020 research and innovation program (Grant No. 810961). S.E.A. was funded by the Royal Society as a University Research Fellow during some of the time of his work on this.

## Abbreviations

NFT, non-fungible token; NCD, normalized compression distance; PCA, Principal Component Analysis; UMAP, Uniform Manifold Approximation and Projection; GAM, generalized additive regression model.

## Availability of data and materials

The datasets supporting the conclusions of this article are available through publicly available databases, as detailed in Results and Methods sections, in particular these sources: <https://deepart.hkust.edu.hk/ART500K/art500k.html> (art500k), <https://www.bcbi.eu/databases/multipic> (Multipic stimuli) and <https://osf.io/ckfmv> (fractals stimuli). The R code to run the compression ensemble models as presented here is available at [https://github.com/andreskarjus/compression\\_ensembles](https://github.com/andreskarjus/compression_ensembles). See the code repository for details on running the code. The code is based on R and ImageMagick. An optimized, more efficient and modular implementation of the compression ensemble algorithm is available here: [https://github.com/Collection-Space-Navigator/compression\\_ensembles](https://github.com/Collection-Space-Navigator/compression_ensembles). It is

based on Python, and uses OpenCV and PIL instead of ImageMagick. An interactive web demo that includes a compression ensemble embedding of a subset of the Historical art dataset is available here: <https://collection-space-navigator.github.io>.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

### Author contributions

AK, MS, and SEA co-designed the research. AK also prepared data, designed and performed data analysis, co-wrote the text, and created the figures. MS also co-designed the data analysis and figures, performed preliminary analysis, co-wrote the text, and provided conceptual guidance. SEA designed the compression ensemble algorithm as documented in a preliminary manuscript, performed preliminary analysis, co-wrote the text and provided conceptual guidance for the analysis and figures. MCS and TO contributed to the design of the analysis and figures, and performed data mining. All authors read and approved the final manuscript.

### Author details

<sup>1</sup>ERA Chair for Cultural Data Analytics, Tallinn University, Narva mnt 25, 10120, Tallinn, Estonia. <sup>2</sup>Department of Chemical Engineering and Biotechnology, University of Cambridge, hilippa Fawcett Drive, CB30AS, Cambridge, UK. <sup>3</sup>The Alan Turing Institute, 96 Euston Road, NW12DB, London, UK.

Received: 26 January 2023 Accepted: 24 May 2023 Published online: 14 June 2023

## References

1. Birkhoff GD (1933) *Aesthetic measure*. Harvard University Press, Cambridge
2. Bense M (1969) Einführung in Die Informationstheoretische Ästhetik Grundlegung Und Anwendung in Der Texttheorie. Rowolt Verlag, Reinbek
3. Kirsch JL, Kirsch RA (1988) The anatomy of painting style: description with computer rules. *Leonardo* 21(4):437. <https://doi.org/10.2307/1578708>
4. Galanter P (2003) What is generative art? Complexity theory as a context for art theory. In: GA2003–6th generative art conference. Citeseer
5. Shamir L, Macura T, Orlov N, Eckley DM, Goldberg IG (2010) Impressionism, expressionism, surrealism: automated recognition of painters and schools of art. *ACM Trans Appl Percept* 7(2):1–17. <https://doi.org/10.1145/1670671.1670672>
6. Kim D, Son S-W, Jeong H (2014) Large-scale quantitative analysis of painting arts. *Sci Rep* 4(1):7370. <https://doi.org/10.1038/srep07370>
7. Elgammal A, Saleh B (2015) Quantifying creativity in art networks. arXiv preprint. [arXiv:1506.00711](https://arxiv.org/abs/1506.00711)
8. Sigaki HYD, Perc M, Ribeiro HV (2018) History of art paintings through the lens of entropy and complexity. *Proc Natl Acad Sci* 115(37):8585–8594. <https://doi.org/10.1073/pnas.1800083115>. Chap. PNAS Plus
9. Elgammal A, Liu B, Kim D, Elhoseiny M, Mazzone M (2018) The shape of art history in the eyes of the machine. In: 32nd AAAI conference on artificial intelligence, AAAI 2018. AAAI Press, Menlo Park, pp 2183–2191
10. Müller TF, Winters J (2018) Compression in cultural evolution: homogeneity and structure in the emergence and evolution of a large-scale online collaborative art project. *PLoS ONE* 13(9):0202019. <https://doi.org/10.1371/journal.pone.0202019>
11. Zanette DH (2018) Quantifying the complexity of black-and-white images. *PLoS ONE* 13(11):0207879. <https://doi.org/10.1371/journal.pone.0207879>
12. Lee B, Seo MK, Kim D, Shin I-s, Schich M, Jeong H, Han SK (2020) Dissecting landscape art history with information theory. *Proc Natl Acad Sci* 117(43):26580–26590. <https://doi.org/10.1073/pnas.2011927117>. <https://www.pnas.org/content/117/43/26580.full.pdf>
13. Manovich L (2020) *Cultural analytics*. MIT Press, Cambridge
14. Perc M (2020) Beauty in artistic expressions through the eyes of networks and physics. *J R Soc Interface* 17(164):20190686. <https://doi.org/10.1098/rsif.2019.0686>
15. Efthymiou A, Rudinac S, Kackovic M, Worring M, Wijnberg N (2021) Graph neural networks for knowledge enhanced visual representation of paintings. In: Proceedings of the 29th ACM international conference on multimedia. Assoc. Comput. Mach., New York, pp 3710–3719
16. Srinivasa Desikan B, Shimao H, Miton H (2022) WikiArtVectors: style and color representations of artworks for cultural analysis via information theoretic measures. *Entropy* 24(9):1175. <https://doi.org/10.3390/e24091175>
17. Fairbairn D (2006) Measuring map complexity. *Cartographic J* 43(3):224–238. <https://doi.org/10.1179/000870406X169883>
18. Rigau J, Feixas M, Sbert M (2007) Conceptualizing Birkhoff's aesthetic measure using Shannon entropy and Kolmogorov complexity. In: Proceedings of the third eurographics conference on computational aesthetics in graphics, visualization and imaging. Computational Aesthetics'07. Eurographics Assoc., Goslar, pp 105–112
19. Campana BJL, Keogh EJ (2010) A compression-based distance measure for texture. *Sci J* 3(6):381–398. <https://doi.org/10.1002/sam.10093>
20. Forsythe A, Nadal M, Sheehy N, Cela-Conde CJ, Sawey M (2011) Predicting beauty: fractal dimension and visual complexity in art. *Br J Psychol* 102(1):49–70. <https://doi.org/10.1348/000712610X498958>
21. Palumbo L, Ogden R, Makin ADJ, Bertamini M (2014) Examining visual complexity and its influence on perceived duration. *J Vis* 14(14):3. <https://doi.org/10.1167/14.14.3>
22. Guha T, Ward RK (2014) Image similarity using sparse representation and compression distance. *IEEE Trans Multimed* 16(4):980–987. <https://doi.org/10.1109/TMM.2014.2306175>

23. Chamorro-Posada P (2016) A simple method for estimating the fractal dimension from digital images: the compression dimension. *Chaos Solitons Fractals* 91:562–572. <https://doi.org/10.1016/j.chaos.2016.08.002>
24. Machado P, Romero J, Nadal M, Santos A, Correia J, Carballal A (2015) Computerized measures of visual complexity. *Acta Psychol* 160:43–57. <https://doi.org/10.1016/j.actpsy.2015.06.005>
25. Fernandez-Lozano C, Carballal A, Machado P, Santos-del-Riego A, Romero J (2019) Visual complexity modelling based on image features fusion of multiple kernels. *PeerJ* 7:e7075. <https://doi.org/10.7717/peerj.7075>
26. Ovalle-Fresa R, Di Pietro SV, Reber TP, Balbi E, Rothen N (2022) Standardized database of 400 complex abstract fractals. *Behav Res Methods* 54:2302–2317. <https://doi.org/10.3758/s13428-021-01726-y>
27. Bagrov AA, Iakovlev IA, Iliasov AA, Katsnelson MI, Mazurenko VV (2020) Multiscale structural complexity of natural patterns. *Proc Natl Acad Sci* 117(48):30241–30251. <https://doi.org/10.1073/pnas.2004976117>. Chap. Physical Sciences
28. McCormack J, Gambardella CC (2022) Complexity and aesthetics in generative and evolutionary art. arXiv preprint. [arXiv:2201.01470](https://arxiv.org/abs/2201.01470)
29. Murphy KA, Bassett DS (2022) The distributed information bottleneck reveals the explanatory structure of complex systems. [arXiv:2204.07576](https://arxiv.org/abs/2204.07576) [cond-mat]
30. Tamariz M, Kirby S (2015) Culture: copying, compression, and conventionality. *Cogn Sci* 39(1):171–183. <https://doi.org/10.1111/cogs.12144>
31. Miton H, Morin O (2021) Graphic complexity in writing systems. *Cognition* 214:104771. <https://doi.org/10.1016/j.cognition.2021.104771>
32. Han SJ, Kelly P, Winters J, Kemp C (2021) Chinese characters have increased in visual complexity over three millennia. *PsyArXiv preprint*
33. Tran N-H, Waring T, Atmaca S, Beheim BA (2021) Entropy trade-offs in artistic design: a case study of Tamil kolam. *Evolut Human Sci* 3:23. <https://doi.org/10.1017/ehs.2021.14>
34. Ecker DW (1963) The artistic process as qualitative problem solving. *J Aesthet Art Crit* 21(3):283–290. <https://doi.org/10.2307/427437>
35. Kolmogorov A (1968) Logical basis for information theory and probability theory. *IEEE Trans Inf Theory* 14(5):662–664. <https://doi.org/10.1109/TIT.1968.1054210>
36. Chaitin GJ (1977) Algorithmic information theory. *IBM J Res Dev* 21(4):350–359
37. Li M, Chen X, Li X, Ma B, Vitányi PM (2004) The similarity metric. *IEEE Trans Inf Theory* 50(12):3250–3264
38. Cilibrasi R, Vitányi PM (2005) Clustering by compression. *IEEE Trans Inf Theory* 51(4):1523–1545
39. Taylor R (2004) Pollock, Mondrian and the nature: recent scientific investigations. *Chaos Complex Letters* 1(3):265–277
40. Alghamdi EA, Velloso E, Gruba P (2021) AUVANA: an Automated video analysis tool for visual complexity. *OSF Preprints*. <https://doi.org/10.31219/osf.io/kj9hx>
41. Tan WR, Chan CS, Aguirre HE, Tanaka K (2016) Ceci n'est pas une pipe: a deep convolutional network for fine-art paintings classification. In: 2016 IEEE international conference on image processing (ICIP). IEEE, Phoenix, pp 3703–3707. <https://doi.org/10.1109/ICIP.2016.7533051>
42. Mao H, Cheung M, She J (2017) DeepArt: learning joint representations of visual arts. In: Proceedings of the 25th ACM international conference on multimedia. MM '17. Assoc. Comput. Mach., New York, pp 1183–1191. <https://doi.org/10.1145/3123266.3123405>
43. Cela-Conde CJ, Ayala FJ, Munar E, Maestú F, Nadal M, Capó MA, del Río D, López-Ibor JJ, Ortiz T, Mirasso C, Marty G (2009) Sex-related similarities and differences in the neural correlates of beauty. *Proc Natl Acad Sci* 106(10):3847–3852. <https://doi.org/10.1073/pnas.0900304106>
44. Lakhali S, Darmon A, Bouchaud J-P, Benzaquen M (2020) Beauty and structural complexity. *Phys Rev Res* 2(2):022058. <https://doi.org/10.1103/PhysRevResearch.2.022058>
45. Nakauchi S, Tamura H (2022) Regularity of colour statistics in explaining colour composition preferences in art paintings. *Sci Rep* 12(1):14585. <https://doi.org/10.1038/s41598-022-18847-9>
46. Opitz D, Maclin R (1999) Popular ensemble methods: an empirical study. *J Artif Intell Res* 11:169–198. <https://doi.org/10.1613/jair.614>
47. Bentz C, Gutierrez-Vasques X, Sozinova O, Samardžić T (2022) Complexity trade-offs and equi-complexity in natural languages: a meta-analysis. *Linguist Vanguard*. <https://doi.org/10.1515/lingvan-2021-0054>
48. Tran-Nam B, Evans C (2014) Towards the development of a tax system complexity index. *Fisc Stud* 35(3):341–370. <https://doi.org/10.1111/j.1475-5890.2014.12033.x>
49. McInnes L, Healy J, Saul N, Großberger L (2018) UMAP: uniform manifold approximation and projection. *J Open Sour Softw* 3(29):861. <https://doi.org/10.21105/joss.00861>
50. Riedl R (2019) Structures of complexity: a morphology of recognition and explanation. Springer, Cham. <https://doi.org/10.1007/978-3-030-13064-0>
51. Nadini M, Alessandretti L, Di Giacinto F, Martino M, Aiello LM, Baronchelli A (2021) Mapping the NFT revolution: market trends, trade networks, and visual features. *Sci Rep* 11(1):20902. <https://doi.org/10.1038/s41598-021-00053-8>
52. Vasan K, Janosov M, Barabási A-L (2022) Quantifying NFT-driven networks in crypto art. *Sci Rep* 12(1):2769. <https://doi.org/10.1038/s41598-022-05146-6>
53. Schich M, Song C, Ahn Y-Y, Mirsky A, Martino M, Barabási A-L, Helbing D (2014) A network framework of cultural history. *Science* 345(6196):558–562. <https://doi.org/10.1126/science.1240064>
54. Gombrich EH (1960) *Art and illusion: a study in the psychology of pictorial representation*. Pantheon, New York
55. Vasari G, Bondanella JC, Bondanella P (1998) *The lives of the artists*. Oxford University Press, Oxford
56. Galenson DW (2004) The life cycles of modern artists. *Hist Methods J Quant Interdiscip Hist* 37(3):123–136. <https://doi.org/10.3200/HMTS.37.3.123-136>
57. Ginsburgh V, Weyers S (2006) Creativity and life cycles of artists. *J Cult Econ* 30(2):91–107. <https://doi.org/10.1007/s10824-006-9011-x>
58. Fraiberger SP, Sinatra R, Resch M, Riedl C, Barabási A-L (2018) Quantifying reputation and success in art. *Science* 362(6416):825–829. <https://doi.org/10.1126/science.aau7224>
59. Solà MC, Korepanova A, Mukhina K, Schich M (2023) Quantifying collection lag in European modern and contemporary art museums. <https://doi.org/10.48550/arXiv.2305.14159>. arXiv preprint. [arXiv:2305.14159](https://arxiv.org/abs/2305.14159)
60. Liu L, Dehmamy N, Chown J, Giles CL, Wang D (2021) Understanding the onset of hot streaks across artistic, cultural, and scientific careers. *Nat Commun* 12(1):5392. <https://doi.org/10.1038/s41467-021-25477-8>

61. Reagan AJ, Mitchell L, Kiley D, Danforth CM, Dodds PS (2016) The emotional arcs of stories are dominated by six basic shapes. *EPJ Data Sci* 5(1):1. <https://doi.org/10.1140/epjds/s13688-016-0093-1>
62. Ohm T, Solà MC, Karjus A, Schich M (2023) Collection space navigator: an interactive visualization interface for multidimensional datasets. <https://doi.org/10.48550/arXiv.2305.06809>. arXiv preprint. [arXiv:2305.06809](https://arxiv.org/abs/2305.06809)
63. Boyd R, Richerson PJ (1996) Why culture is common, but cultural evolution is rare. In: Runciman WG, Smith JM, Dunbar RIM (eds) *Evolution of social behaviour patterns in primates and man*, vol 88. Oxford University Press, London, pp 77–93
64. Tomasello M (2009) *The cultural origins of human cognition*. Harvard University Press, Cambridge
65. Beckner C, Blythe R, Bybee J, Christiansen MH, Croft W, Ellis NC, Holland J, Ke J, Larsen-Freeman D, Schoenemann T (2009) Language is a complex adaptive system: position paper. *Lang Learn* 59(s1):1–26. <https://doi.org/10.1111/j.1467-9922.2009.00533.x>
66. Mesoudi A, Thornton A (2018) What is cumulative cultural evolution? *Proc R Soc B, Biol Sci* 285(1880):20180712. <https://doi.org/10.1098/rspb.2018.0712>
67. Ebeling W, Freund J, Schweitzer F (1998) *Komplexe strukturen: entropie und information*. Teubner, Leipzig
68. Sinclair NC, Ursell J, South A, Rendell L (2022) From Beethoven to Beyoncé: do changing aesthetic cultures amount to “Cumulative cultural evolution?” *Front Psychol* 12
69. Gombrich EH (1971) *The ideas of progress and their impact on art*, 1st edn. Cooper Union School of Art and Architecture
70. Kemp C, Xu Y, Regier T (2018) Semantic typology and efficient communication. *Annu Rev Linguist* 4(1):109–128. <https://doi.org/10.1146/annurev-linguistics-011817-045406>
71. Karjus A, Blythe RA, Kirby S, Wang T, Smith K (2021) Conceptual similarity and communicative need shape colexification: an experimental study. *Cogn Sci* 45(9):13035. <https://doi.org/10.1111/cogs.13035>
72. Wittgenstein L (1953) *Philosophical investigations*. Philosophische untersuchungen. Macmillan & Co., Oxford, p 232
73. Weitz M (1956) The role of theory in aesthetics. *J Aesthet Art Crit* 15(1):27–35
74. Rosch E, Mervis CB (1975) Family resemblances: studies in the internal structure of categories. *Cogn Psychol* 7(4):573–605
75. Friedlander MJ (1946) *Von Kunst Und Kennerschaft*. Reclam Verlag, Leipzig
76. Vylomova E, Rimell L, Cohn T, Baldwin T (2016) Take and took, gaggle and goose, book and read: evaluating the utility of vector differences for lexical relation learning. In: *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)*. Assoc. Comput. Linguistics, Berlin, pp 1671–1682. <https://doi.org/10.18653/v1/P16-1158>
77. Cassirer E (1927) *Philosophie der Symbolischen Formen: Zweiter Teil – Das Mythische Denken*, 1st edn. Meiner, F, Hamburg
78. Cassirer E (1927) *Das Symbolproblem Und Seine Stellung Im System der Philosophie*. *Z Ästhet Allg Kunstwiss* 21:295–322
79. Schich M (2019) Cultural analysis situs. ART-Dok eprint. <https://doi.org/10.11588/artdok.00006347>
80. Panofsky E (1939) *Studies in iconology: humanistic themes in the art of the renaissance*. Oxford University Press, New York
81. Gärdenfors P (2000) *Conceptual spaces: the geometry of thought*. <https://doi.org/10.7551/mitpress/2076.001.0001>
82. Gärdenfors P (2014) *The geometry of meaning: semantics based on conceptual spaces*. <https://doi.org/10.7551/mitpress/9629.001.0001>
83. Eigen M (2013) *From strange simplicity to complex familiarity: a treatise on matter, information, life and thought*. Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780198570219.001.0001>
84. Dou Q, Zheng XS, Sun T, Heng P-A (2019) Webthetics: quantifying webpage aesthetics with deep learning. *Int J Hum-Comput Stud* 124:56–66. <https://doi.org/10.1016/j.ijhcs.2018.11.006>
85. Beauvois MW (2007) Quantifying aesthetic preference and perceived complexity for fractal melodies. *Music Percept* 24(3):247–264. <https://doi.org/10.1525/mp.2007.24.3.247>
86. Clemente A, Pearce MT, Nadal M (2022) Musical aesthetic sensitivity. *Psychol Aesthet Creat Arts* 16(1):58–73. <https://doi.org/10.1037/aca0000381>
87. Sennrich R, Haddow B, Birch A (2016) Neural machine translation of rare words with subword units. In: *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)*. Assoc. Comput. Linguistics, Berlin, pp 1715–1725. <https://doi.org/10.18653/v1/P16-1162>
88. Srinivasa Desikan B, Evans J (2022) Aggregate, integrate and align to embed everything: a multi-modal framework for measuring cultural dynamics. In: *Cultures in AI/AI in culture*. A NeurIPS 2022 workshop
89. Wang X, Jiang Y, Bach N, Wang T, Huang Z, Huang F, Tu K (2021) Automated concatenation of embeddings for structured prediction. In: *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers)*. Assoc. Comput. Linguistics, Berlin, pp 2643–2660. <https://doi.org/10.18653/v1/2021.ac-long.206>
90. Devlin J, Chang M-W, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*. Assoc. Comput. Linguistics, Minneapolis, pp 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
91. Schich M (2010) Revealing matrices. In: Steele J, Iliinsky N (eds) *Beautiful visualization: looking at data through the eyes of experts*. O'Reilly Media, Sebastopol, pp 227–254
92. Sabetsarvestani Z, Sober B, Higgitt C, Daubechies I, Rodrigues MRD (2019) Artificial intelligence for art investigation: meeting the challenge of separating x-ray images of the Ghent altarpiece. *Sci Adv* 5(8):7416. <https://doi.org/10.1126/sciadv.aaw7416>
93. Warburg A (2008) *Der bilderatlas mnemosyne*. Akademie Verlag, Berlin
94. Impett L, Süssstrunk S (2016) Pose and pathosformel in aby Warburg's bilderatlas. In: *European conference on computer vision*. Springer, Berlin, pp 888–902. [https://doi.org/10.1007/978-3-319-46604-0\\_61](https://doi.org/10.1007/978-3-319-46604-0_61)
95. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B (2022) High-resolution image synthesis with latent diffusion models. <https://doi.org/10.48550/arXiv.2112.10752>. [arXiv:2112.10752](https://arxiv.org/abs/2112.10752)

96. Duñabeitia JA, Crepaldi D, Meyer AS, New B, Pliatsikas C, Smolka E, Brysbaert M (2018) MultiPic: a standardized set of 750 drawings with norms for six European languages. *Q J Exp Psychol* 71(4):808–816.  
<https://doi.org/10.1080/17470218.2017.1310261>
97. Strezoski G, Worring M (2017) OmniArt: multi-task deep learning for artistic data analysis. arXiv preprint, 1708.00684.  
[arXiv:1708.00684](https://arxiv.org/abs/1708.00684)

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---