Exploring the definition of art through deep net visualization

J.P. Lewis[1,2] **I Chen Yeh**[3] **Agata Migalska**[4] **Samuel B. Johnson**[3] **Ruth West**[3]
[1]SEED Electronic Arts [2]Victoria University
[3]xREZ Art + Science Lab, University of North Texas
[4]Wroclaw University of Science and Technology
noisebrain@gmail.com ruth.west@unt.edu

Abstract

The definition of art has been debated for more than 2000 years and continues to be a puzzle. While scientific investigations offer hope that we might resolve this puzzle, machine learning classifiers that discriminate art from non-art images generally do not provide an explicit definition, and brain imaging and psychological theories are at present too coarse to provide a formal characterization. In this work, rather than approaching the problem using a machine learning approach trained on existing artworks, we hypothesize that art can be defined in terms of preexisting properties of the visual cortex. Specifically, we propose that a broad subset of visual art can be defined as *patterns that are exciting to a visual brain.* Resting on the finding that artificial neural networks trained on visual tasks can provide predictive models of processing in the visual cortex, our definition is operationalized by using a trained deep net as a surrogate "visual brain", where "exciting" is defined as the activation energy of particular layers of this net. We find that this definition easily discriminates a variety of art from non-art, and further provides a ranking of art genres that is consistent with our subjective notion of "visually exciting". By applying a deep net visualization technique, we can also validate the definition by generating example images that would be classified as art. The images synthesized under our definition resemble visually exciting art such as Op Art and other human-created artistic patterns.

1 Introduction

We propose an objective definition of a subset of visual art. Specifically, we propose that a broad subset of visual art consists of images (or objects) that are exciting to a visual brain, where "exciting" can be quantified in terms of the activations of neurons while processing such images. The proposed definition is simulated and evaluated using a deep network trained on an image recognition task, and by using a deep network visualization technique to synthesize sample images that are "exciting" for this network.

Our methodology rests on the recent discovery that artificial deep nets trained on visual tasks are surprisingly accurate predictive models of both cortical spiking and population aggregate responses of primate visual brains [KRK14, Kri15, GvG15, CKP+16, YD16, WSZ+17]. By making use of this correspondence, we obtain a computational realization of the proposed definition that would not be possible using alternative methods such as brain imaging.

Our definition is not intended to cover all forms of art that have visual expression. For example, it does not apply to conceptual art as exemplified by (for example) Duchamp's *Fountain* (a urinal). Nevertheless, we believe that this definition describes many genres of visual art that coincide with the laypersons' typical understanding of the word, and it includes both prominent styles such as

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

impressionism and expressionism, and also the traditional art forms of many cultures, such as Islamic tiles, aboriginal art, etc. (Figure 1).

Art is a nearly universal part of human experience. Visual artforms are common across both cultures and history, and appear in our everyday life not just through paintings and artistic photographs, but also in the form of decorative patterns on rugs, clothing, etc. Yet despite the prevalence of art, its definition has been a subject of philosophical discussion and debate since the time of Plato (Section 2). Many of the proposed definitions are vague or refer to other concepts that are themselves subjectively defined [Ste97, Dav91, Dut09, Car00, Dav06, Ada07]. For example, [Bel14] describes "lines and colors combined in a particular way, certain forms and relations of forms [that] stir our aesthetic emotions". Few, if any, existing definitions permit a formal procedure that would allow disagreeing observers (or a computer program) to determine if a particular image or object is "art". From a scientific viewpoint, however, an objective categorization should come with some objective or operational test (i.e. algorithm) that can define which objects belong in a category.

Modern scientific research in neuroaesthetics and machine learning has adopted objective measures in studies of art. FMRI studies have found brain regions that are particularly activated by viewing of art images [VS14, BBP+16]. However the level of analysis currently provided by FMRI is relatively coarse and does not provide a description with sufficient detail to permit computational simulation.

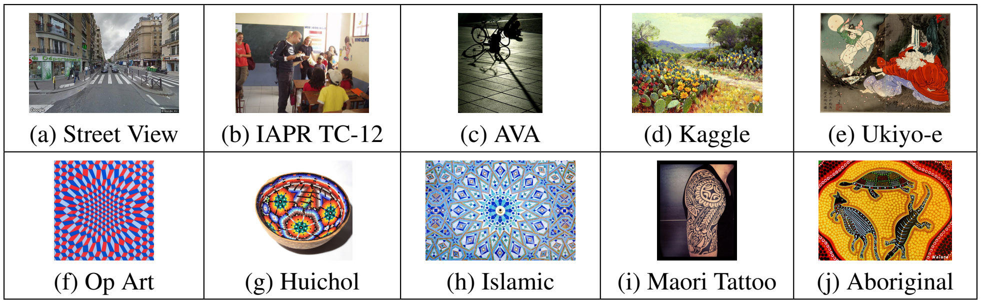
Art can be implicitly defined by training statistical learning algorithms to distinguish art from non-art, and researchers in machine learning and computer science have found relatively simple discriminative features that successfully classify many art and non-art images. For example, [TMJ99] used a box-counting definition of the fractal dimension to analyze Pollock paintings. For our purpose these approaches have a clear limitation: simple discriminative models provide a necessary but not sufficient criterion and they can be applied to objects outside the intended domain, with nonsensical results. For example, [TMJ99] was famously challenged on the grounds that crude line drawings result in similar values of the chosen measure [JSM06]. Similarly, it has been shown that many artworks have an amplitude spectrum $f^{-\alpha}$ with exponent α near one, however the Heaviside step function also has a spectrum of this form [Bra00]. We feel that this is not a flaw of these particular studies, but rather an expected consequence of using a single discriminative feature.[1]

Complementing brain imaging and approaches using statistical classifiers, a third approach is computational experiments. Such experiments by definition require computational detail that is lacking in current brain imaging, and can provide generative predictions of phenomena that are lacking in discriminative approaches. As one example of this approach, [OF96] has been quite influential in computational neuroscience discussions of early visual processing. Our approach is a computational experiment in this spirit.

We operationalize our proposed definition by defining "exciting" in terms of the activations of a feedforward network trained on a visual classification task. Specifically, we use VGG [SZ14], trained on ImageNet [RDS+15]. This approach fundamentally relies on the observation that deep nets trained on visual tasks form representations that resemble those in the visual cortex [Kri15, CKP+16, YD16]. The resemblance includes both receptive fields [LEN08] (c.f. [OF96, RB99]) and cortical activity [GvG15, YD16, WSZ+17]. Further, the nature of discovered receptive fields is surprisingly independent of the task on which the network is trained [YCBL14]. With this operational definition, the "art-ness" of any image can be objectively assessed by measuring the activations of particular layers in the deep net (Section 3). We find that this measure strongly distinguishes art from utilitarian images (Section 4). Further, it produces a ranking of visual art genres that is consistent with our expectations, with visually striking genres such as Op Art ranked as the most energetic.

By adapting deep net visualization techniques [MV15, YCN+15, MV16], we are also able to "invert" our definition to hallucinate examples of images that would be classified as art under the definition. Specifically, art-like images are generated by optimizing an image, starting from random noise, such that it maximally excites particular layers of a deep image classification net (Section 3). The particular layers are preferentially chosen from those that are most excited by art images. The generated images resemble Op Art and other human-generated abstract images (Section 4.1), providing an additional degree of plausibility and validation that the definition is not capturing irrelevant categories of objects.

[1] As a less domain-specific example, height can be used to predict the gender of adult humans with some degree of accuracy, but a particular height may equally describe non-human objects.



(a) Street View (b) IAPR TC-12 (c) AVA (d) Kaggle (e) Ukiyo-e
(f) Op Art (g) Huichol (h) Islamic (i) Maori Tattoo (j) Aboriginal

Figure 1: Examples of art and non-art image data used in this study: (a) Google Street View, (b) IAPR TC-12 Benchmark, (c) AVA (artistic photos), (d) Kaggle subset of Wikiart (Impressionism category), (e) Ukiyo-e (Japan), (f) Op Art, (g) Huichol (Mexico), (h) Islamic tile, (i) Maori tattoo (New Zealand), (j) Aboriginal (Australia). Please enlarge to see details.

2 Related work

The nature and definition of art has been considered in philosophy, art theory, and psychology, and more recently is being explored in neuroscience and machine learning. The literature on this topic is too large to fully survey and we will only mention some representative ideas and research here.

A survey of philosophical and art-theory definitions of art from a scientific point of view can be found in [Cha15]. Proposals range from Plato's simple "art is an imitation of life" to the anti-essentialist viewpoint which holds that art can never be defined. Another view proposes that one cannot determine if an object is art by looking at it; instead art must be understood by assessing its relation to culture and history. It is safe to say that no consensus has resulted from two millennia of philosophical discussions.

Researchers in psychology have considered the relationship between the complexity of an image and its arousal in the viewer. [Ber71] proposed that there is an inverted-U shaped relation between the complexity of a pattern and aesthetic preference for that pattern. This relationship has been confirmed in some but not all studies. The discrepancy may be due to the simple patterns and rudimentary complexity measures used. [Mar90] proposed that art maximizes "arousal potential" in the viewer while attempting to avoid radical changes that would provoke negative reactions. These ideas may be consistent with our proposal, though they are sufficiently high-level that it is difficult to quantitatively evaluate them.

Art relies on aesthetic perception, which is generally assumed to be a universal ability in humans [Dis95]. The question of why an aesthetic sense exists is considered in the field of evolutionary aesthetics [VG03]. One proposal is that aesthetically appealing objects resemble hospitable habitats. However, it is easy to think of counter examples – images of coral reefs and tigers are aesthetically pleasing to many people. In evolutionary aesthetics the discussion of aesthetics is also linked with the display of patterns in animals. The evolutionary justification for the peacock's tail was a puzzle for Darwin, and in fact he proposed that animals may have an aesthetic sense [Gay10]. The Fisherian runaway (FR) hypothesis addresses the peacock tail paradox. It proposes that some, possibly arbitrary, physical characteristic comes to be associated with fitness, perhaps by chance, and then these characteristics are exaggerated in subsequent generations. However, an important question (not addressed by FR) is why the patterns of peacock feathers, tropical fish, and butterflies should appear attractive to humans as well as the respective birds, fish, and insects.

Art and aesthetics have increasingly become subjects of scientific study, and these subjects are now being investigated with tools from neuroscience, computer science, and machine learning as well as psychology. FMRI studies have found areas of the brain that respond to art images [VS14, BBP$^+$16]. These studies also show that a large number of brain regions not associated with visual processing are activated. This may be the result of processing the subject matter of representational paintings, which involve people, places, and events, as well as the need to perform judgements requested in the studies. On the theory side our proposed definition is most directly anticipated by a statement in [RH99], "that artists either consciously or unconsciously deploy certain rules or principles (we call them laws) to titillate the visual areas of the brain" and may relate to the peak shift principle defined in that essay, though it appears to be unrelated to some of the other proposed laws. The discrepancy may be due to our focus on patterns in traditional visual art, whereas the discussion in [RH99] applies to all art forms.

Statistical and machine learning approaches successfully discriminate art from non-art images. [GR10] show that many paintings have a power-law amplitude spectrum close to f^{-1}, similar to the spectra of natural images. [SWF09] showed that different types of art can be clustered based on simple spatial and color statistics. [BBR17] discriminate art from non-art images with high accuracy using three simple statistics on the first convolutional layer of Alexnet. Aesthetic rankings of photographs by humans have been predicted using SVMs operating on color and SIFT features [MMP12].

Deep learning approaches have recently resulted in generative algorithms for artistic images, by implicitly imitating statistics of given training images. These approaches have produced remarkable results. [GEB15] transforms images to the style of a given art image by optimizing an initial random image to have VGG activations similar to those of the provided style and content images. Specifically, the notion of artistic style is captured by matching the Gram matrix of activations of one or more layers to those of a given style image. Although this is matching a second-order statistic, and thus might seem to be analogous (by the Wiener-Khinchine theorem) to the amplitude or power-spectrum statistic mentioned above, the fact that the activations of a deep net are used rather than pixels means that high-order statistics are captured.

While the results of [GEB15] are remarkable, the approach produces imitations of a given, existing style. This is common to a number of previous approaches (e.g. texture synthesis methods) that mimic given exemplars, and so might be argued to be not truly creative. [ELEM17] introduced an algorithmic definition of creativity. Their approach is a modified GAN [GPAM$^+$14] in which the discriminator provides two losses. The first is an art/non-art loss. The second is a stylistic ambiguity term, expressed as the cross-entropy between the style class posterior and a uniform distribution. While the first term encourages the generator to produce art images from the same distribution as the training set, the second causes it to attempt to produce images that are difficult to easily classify in any known style, and thus can be considered as stylistically novel. There is also a long tradition of computer art such as [McC91] produced using traditional computer programming rather than machine learning. From our perspective these approaches embody the particular programmer's notion of art, and it is difficult to extrapolate them to a more general definition.

Deep net visualization techniques [MV15, YCN$^+$15, MV16] have provided one of the primary means of understanding the computations in deep nets. Going beyond preimages, [MLB$^+$17] computes impressive heatmaps of the pixels in a specific image that contribute to a classification. The technique of *activation maximization* [Lew88, MV15, YCN$^+$15, MV16] is central to our work. This technique is a "dual" of the supervised training of a deep classifier: instead of modifying initially random network weights so as to reduce a classification loss, the weights from a trained network are used without modification, and the *input* is modified subject to being constrained to displayable code values. The loss is simply the activation of a particular neuron whose function is being investigated, resulting in an example of an image patch that causes a strong response in that neuron. In visualizing these pre-images of particular neurons it is often necessary to regularize the synthesized image to preserve characteristics of natural images (such as their piecewise-smooth character), since the preimages otherwise contain spikes and other atypical features [MV15, YCN$^+$15, DB16]. In our work we invert entire layers (generally more than one). In this scenario we found that regularization is not necessary to produce appealing, highly detailed images. However we also sometimes reduce high frequencies during the optimization [YCN$^+$15] as a method to explore lower frequency structure (Section 3 describes an alternate multi-resolution construction). [MV16] provides a comprehensive guide to recent visualization and associated regularization strategies.

The definition proposed and explored in this paper appears to be distinct from those in art theory and philosophy. It is anticipated by thinking in psychology and neuroscience however, and adds an potential operational form and some experimental validation to those ideas. The synthesis experiments in our work are motivated as an exploration of the definition of art and are not intended as a serious attempt to generate computer art. In contrast to many generative art and texture synthesis approaches, ours is not trained to mimic statistics of provided example images, but instead relies only on the generic visual processing provided by VGG. It is therefore more of a "from first principles" approach to our problem.

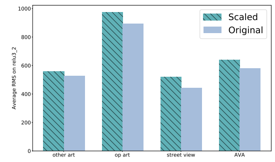


Figure 2: Whole image activations versus scaled subregions with the highest activations.

3 Method

3.1 Linear formulation

A linear version of our "exciting image" definition is easy to formulate as an eigenvector problem. We start with the common observation that simple cells in the visual cortex resemble Gabor wavelets [OF96]. The action of an early stage of the visual cortex then resembles the application of a filter bank $\mathbf{G}$ to the image, where each element of the filter bank consists of a Gabor wavelet of some frequency and orientation to be convolved with the image. The most exciting image can be formulated by maximizing the total energy of neurons, e.g. the sum squares filter responses, while keeping the L_2-norm of the resulting image constant. The convolution can be expressed in linear algebra terms by vectorizing the (m, n) image into a vector $\mathbf{x}$ of size $(mn, 1)$, and expressing the filter bank as a matrix $\mathbf{G}$ with k Toeplitz blocks, each representing one filter. The goal of finding an exciting image can then be expressed as

$$\arg\max \|\mathbf{G}\mathbf{x}\|^2 \quad \text{subject to} \quad \|\mathbf{x}\|^2 = \text{const.}$$

which can be solved as the eigenvector problem $\mathbf{G}^T\mathbf{G}\mathbf{x} = \lambda\mathbf{x}$. This approach is restricted to modeling the visual cortex as a single linear operator, however, which is an overly simplistic assumption.

3.2 Deep net formulation

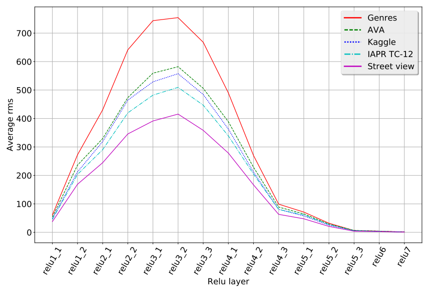
To address this, we instead adopted VGG [SZ14] as a representative nonlinear "visual cortex". The 'visually exciting' definition is implemented by measuring the activation across several of the early (non-semantic) relu layers.[2] We experimented with three measures of activation across a layer: RMS, L_1, and entropy. Entropy was measured by normalizing the histogram of activations, and thus is a (lack of) sparsity measure, whereas L_1 reflects lack of sparsity as well as energy. All three measures gave similar rankings, though the RMS measure provided a broader spread of values. RMS was adopted as a default choice for our experiments. Figure 3 shows the resulting activations across different layers for several categories of art and non-art, described in Section 3.3.

For the image synthesis tests in Section 4.1, we performed an activation maximization optimization [MV16], operating across selected layers of VGG rather than on individual neurons. Human visual perception is quite scale invariant. To take this into account we also constructed a multiscale variant of VGG. In this architecture, lower resolution processing is effected by passing the input image through an avgpool. The result is processed by a separate VGG forward sweep, and the activations of the different resolutions are weighted and summed, typically with the lower resolution having smaller weights. Results are shown in Section 4.1.

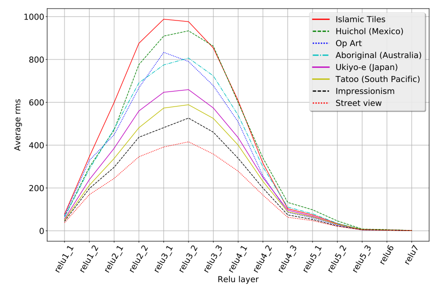
3.3 Image Datasets

Testing our hypothesis requires obtaining VGG activations on sufficient sets of art and non-art images. Despite the lack of an accepted independent definition of art, there are readily available databases of works that are generally agreed to be art. Finding non-art images requires some care. Generic photographs from [GCMD06] are a source of images that are not regarded as nor intended to be art. One possible objection is however that human-composed photographs may contain an unconscious bias ("the hidden artist in all of us"). To address this, we also obtained images from Google Street View [ADF$^+$10] consisting of half urban and half rural locations, selected as outlined in the supplementary material. A second objection is that a successful discrimination of utilitarian

[2] We used VGG-19, though only the early layers are required so VGG-16 would presumably give equivalent results.



(a) Overall categories



(b) Individual genres

Figure 3: RMS activation at several relu layers of VGG, averaged across images.

Table 1: One-way permutation analysis

Comparison	Stat	p.value	p.adjust
AVA/IAPR = 0	13.97	0	0.00
AVA/Kaggle = 0	4.317	1.581e-05	0.00
AVA/Genres = 0	-21.13	4.382e-99	0.00
AVA/Street View = 0	30.6	0	0.00
IAPR/Kaggle = 0	-9.28	1.697e-20	0.00
IAPR/Genres = 0	-29.05	1.464e-185	0.00
IAPR/Street View = 0	23.08	0	0.00
Kaggle/Genres = 0	-23.44	1.829e-121	0.00
Kaggle/Street View = 0	26.95	0	0.00
Genres/Street View = 0	36.69	0	0.00

photos from images of paintings and other artwork may actually be discriminating photo versus painting rather than art vs. non-art. To address this, we included a set of intentionally artistic photos from [MMP12].

On the art side, a Kaggle competition provided a large index of art images from WikiArt [noa]. These are primarily from Western art of all eras, with a small number of other cultures represented. The coverage of other cultures was broadened by selecting specific genres (Impressionist paintings, Op Art, Islamic tiles, Aboriginal art, Japanese Ukiyo-e ink paintings, Huichol bowls from Mexico, and Maori tattoos from New Zealand; see Figure 1). Two hundred images from each genre were obtained, using selection methodology as outlined in the supplementary material. We originally gathered five such categories with 200 images each, to be included in a balanced design against randomly chosen subsets of 1000 images from the larger databases (Street View, IAPR TC-12, Kaggle). The number of cultures and genres represented was subsequently increased to seven (Figure 1), resulting in balanced groups of 1400 images.

3.4 Image Resolution

One initial concern was that VGG would not “see” the detail in the images after reducing to the relatively low 224-pixel input resolution. To address this we did a search across five scales, taking the max across different crops at that scale. The results (averaged across images) were not markedly different from simply using the whole image, as shown in Figure 2, so the unit scale (whole image) was used. (This test was performed on a preliminary dataset with 40 images in the smallest category.)

4 Results

Figure 3a shows the RMS activations of a number of VGG layers averaged across images in each of our major data categories. Genres is the lumped collection of the seven cultures/genres indicated in Section 3.3, while each of the other groups contains 1400 randomly selected images, with a balance of urban and rural in the case of Street View. The relu3_2 layer responds most strongly. The art categories have higher activations than the two non-art categories, as expected under our hypothesis. Interestingly, the human-composed photographs from IAPR TC-12 have somewhat higher activations

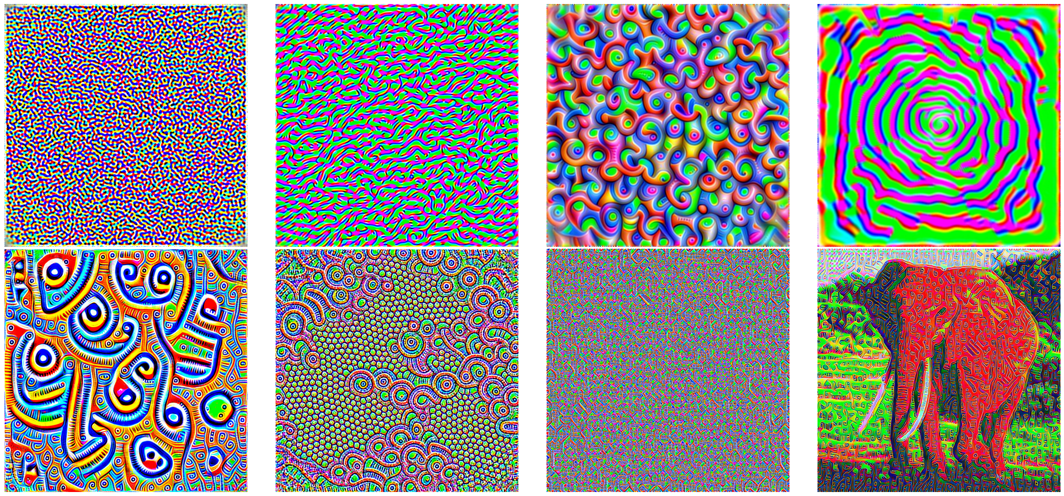


Figure 4: Image synthesis results. The image on the top right used a foveal warp simulating the retino-cortical map [BCG+02]. The elephant image resulted from optimizing an existing image rather than random values (image obtained from [pix]). Please enlarge to full screen for best effect.

than those of Street View, although they are not taken by artists nor intended as art. It would seem that even casual photographers choose modestly "exciting" locations and subjects.

As a first test of our hypothesis, we grouped the non-art images, 1400 from Street View including a balance of urban and rural locations, and 1400 randomly selected images from IAPR TC-12. These were compared against 2800 randomly selected images from Kaggle using a Welch's (unequal variances) two-sided t-test of the mean RMS activation across images of relu3_2. The normality condition for this test is satisfied due to the CLT effect of the mean. There is a significant difference in means (non-art = 462.44, art = 559.81), $t = 27.54, df = 5026.3, p$ <2.2e-16.

One conceivable objection might be that perhaps the presence of urban images heavily skews the results, and that images of nature might not be distinguishable from art under the proposed definition. Similarly, one could question whether the activations are discriminating human-made from natural scenes, rather than art from non-art. In fact, the average RMS of the rural subset of Street View images is similar to and slightly lower than the urban images (398.57 vs. 435.39). These images were taken in regions of national parks and predominantly contain natural objects.

Figure 3b shows the average RMS activations from each of the seven genre categories, with a random subset of Street View images included for comparison. The ranking of genres in this figure generally corresponds to our subjective impression of which images are "visually exciting", with categories such as Op Art and Huichol bowls ranked higher than Impressionism.

To explore further, a one-way ANOVA was performed to test the effect of non/art categories on the RMS VGG activations in layer relu3_2. There was a significant effect, F(4,6995) = 896.88, p=0.0000. Our experiment meets guidance on the use of a parametric test [McD09] – balanced design, more than 10 samples per group, and standard deviations across groups do not vary by a large factor. However we also performed a one-way permutation analysis (R package coin) (Table 1). The results show that all the pairwise difference of RMS activations in Figure 3a are significant. The deep net activation criterion not only easily separates art from non-art, but also separates different genres of art.

4.1 Synthesis results

Figure 4 shows images synthesized using activation maximization across selected early (non-semantic) relu layers of VGG. Foveal warps are implemented using [JSZK15]. The images resemble Op Art and other human-generated patterns (see supplementary material). With one exception the images have RMS activations above 1000, though of course lower activation levels can be obtained if desired.

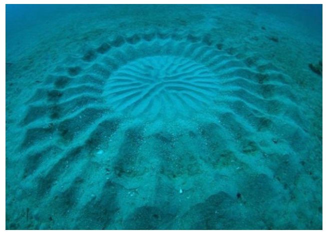


Figure 5: Pattern created by a male pufferfish [KOI13]. Permission to reproduce Figure 3 of [KOI13] obtained from Rightslink copyright clearance center.

5 Discussion

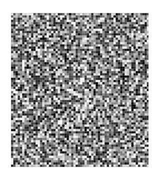
The proposed definition has several potentially controversial consequences. One is that art can be appreciated somewhat independently of culture, since it is fundamentally a reflection of processing in the human visual cortex (this statement is qualified by the extent to which the visual cortex itself differs across cultures due to environment-specific learning). The fact that we may appreciate cave drawings and the diverse images in Figure 1 supports this idea. However it is in contradiction to one major viewpoint in art theory (Section 2).

A second surprising consequence is that the definition should be relevant to animals since they, too, have brains capable of sophisticated visual processing.[3] We will not attempt to address this controversial idea, but will give a few considerations that suggest that it cannot be immediately rejected.

Relatively few animals create patterns. The bowerbird is a well known exception; some bowerbirds create intricate symmetric patterns, and their skill increases over their lifetime [Din08]. The pattern in Figure 5, created by a pufferfish [KOI13], has a VGG relu3_2 RMS activation of 590.65, placing it solidly in the "art" category and slightly above the mean activation on the Kaggle dataset.

While few animals create patterns, birds, butterflies, and other animals that rely on visual signaling often display patterns that are appealing or even beautiful. While these patterns have significance for the particular animals, it is not obvious why they should be appealing for humans. Following our hypothesis where it leads, we could imagine that the visual systems of these animals extract similar features (oriented edges, blobs, etc.) as are found in higher mammals, and that evolution has exploited this by evolving feathers and wings that are particularly exciting. Our proposal thus suggests why humans also find these patterns attractive – by a convergent evolution principle, the early visual brains of various creatures sharing the same environment may have similarities, and thus the preimages that excite these brains would be similar as well.

Although this is just an hypothesis with little support, consider the following alternative hypothesis: the patterns displayed by animals are simply arbitrary examples from the space of possible patterns. In fact, we know that this alternative hypothesis is not true as stated: by a standard counting argument [LV97], the overall space of visual patterns (as well as other types of signals) overwhelmingly consists of nearly incompressible (random) patterns such as the adjacent figure. The patterns found in nature are quite different from the typical pattern, and thus some non-arbitrary selection principle is involved.



6 Conclusion

We propose a definition of (some) visual art: that art is patterns that are exciting to a visual brain. The definition is intuitively plausible, and broadly consistent with "arousal" based theories in psychology. It is also specific enough to test, by operationalizing it using a well known visual recognition deep

[3]For example, pigeons can be taught to accurately discriminate between paintings by Monet and Picasso [WSW95], between "good" and "bad" children's drawings [Wat10], and they are also adept at detecting tumors in both histology images and mammograms [LKNW15]. Remarkably, honeybees can also discriminate both human faces, and between Monet and Picasso paintings, and in the case of the paintings they do not appear to rely on simple measures such as overall luminance or spatial frequency information [WMTR13].

net as a surrogate "visual brain". The operationalized definition succeeds in distinguishing art from utilitarian images, and it applies across a number of time periods and cultures.

Applying a deep net visualization technique allows a generative test of the proposed definition. The generated images resemble those of Op Art and other exciting human-generated art. The activation maximization procedure resembles a theory of drug-induced visual hallucination [GST03], in which a drug uniformly stimulates the visual cortex resulting in the appearance of patterns. More generally, activation maximization superficially resembles the process of painting itself, in which the artist starts with an uninteresting canvas and iteratively places colors while observing the resulting visual effect [Lew88, lat08].

On the other hand, while "visually exciting" appears to describe much and perhaps most traditional visual art, the definition does not apply to many forms of modern art in which a concept is primary and its visual form plays only a supporting role.

Acknowledgments

We would like to acknowledge David Balduzzi's contribution to early discussions, several literature and concept pointers from Javier von der Pahlen, and the assistance of xREZ Art + Science lab research assistants in data collection: Claire Adams, Luke Hillard, Zain Khoja, and Nicholas Ligon.

References

[Ada07] Thomas Adajian. The Definition of Art. Online: Stanford Encyclopedia of Philosophy, October 2007.

[ADF+10] Dragomir Anguelov, Carole Dulong, Daniel Filip, Christian Frueh, Stéphane Lafon, Richard Lyon, Abhijit Ogale, Luc Vincent, and Josh Weaver. Google Street View: Capturing the world at street level. *Computer*, 43(6):32–38, 2010.

[Bar17] Terry Barrett. *Why is That Art?: Aesthetics and Criticism of Contemporary Art*. Oxford, 2017.

[BBP+16] M. Boccia, S. Barbetti, L. Piccardi, C. Guariglia, F. Ferlazzo, A. M. Giannini, and D. W. Zaidel. Where does brain neural activation in aesthetic responses to visual art occur? meta-analytic evidence from neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, 60:65–71, 2016.

[BBR17] Anselm Brachmann, Erhardt Barth, and Christoph Redies. Using CNN features to better understand what makes visual artworks special. *Frontiers in Psychology*, 8, May 2017.

[BCG+02] Paul C. Bressloff, Jack D. Cowan, Martin Golubitsky, Peter J. Thomas, and Matthew C. Wiener. What geometric visual hallucinations tell us about the visual cortex. *Neural Computation*, 14(3):473–491, March 2002.

[Bel14] Clive Bell. *Art*. Stokes, 1914.

[Ber71] Daniel E Berlyne. *Aesthetics and psychobiology*. Appleton-Century-Crofts, 1971.

[Bra00] R.N. Bracewell. *The Fourier Transform and Its Applications*. McGraw Hill, 2000.

[Car00] Noël Carroll, editor. *Theories of Art Today*. University of Wisconsin Press, Madison, Wis. ; London, 2000.

[Cha15] Anjan Chatterjee. *The Aesthetic Brain: How We Evolved to Desire Beauty and Enjoy Art*. Oxford University Press, New York, reprint edition, September 2015.

[CKP+16] Radoslaw Martin Cichy, Aditya Khosla, Dimitrios Pantazis, Antonio Torralba, and Aude Oliva. Deep neural networks predict hierarchical spatio-temporal cortical dynamics of human visual object recognition. *CoRR*, abs/1601.02970, 2016.

[Dav91] Stephen Davies. *Definitions of Art*. Cornell University Press, Ithaca, N.Y, 1991.

[Dav06] Stephen Davies. *The Philosophy of Art*. Wiley-Blackwell, Malden, MA ; Oxford, 1 edition edition, February 2006.

[DB16] Alexey Dosovitskiy and Thomas Brox. Generating images with perceptual similarity metrics based on deep networks. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 658–666. Curran Associates, Inc., 2016.

[Din08] Sadie Dingfelder. My elephant could have painted that: The creative drive–as well as some aesthetic sense–may cross species lines. *Science Watch*, 39(11):30, 2008.

[Dis95] Ellen Dissanayake. *Homo Aestheticus: Where Art Comes From and Why*. University of Washington Press, Seattle, October 1995.

[Dut09] Denis Dutton. *The Art Instinct: Beauty, Pleasure, & Human Evolution*. Oxford University Press, 2009.

[Eat83] M.M. Eaton. *Art and Nonart: Reflections on an Orange Crate and a Moose Call*. Fairleigh Dickinson University Press, 1983.

[ELEM17] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. CAN: Creative Adversarial Networks, Generating "Art" by learning about styles and deviating from style norms. *arXiv:1706.07068 [cs]*, June 2017.

[Gay10] Jean Gayon. Sexual selection: Another Darwinian process. *Comptes Rendus Biologies*, 333(2):134–144, February 2010.

[GCMD06] Michael Grubinger, Paul Clough, Henning Müller, and Thomas Deselaers. The IAPR TC-12 benchmark: A new evaluation resource for visual information systems. In *International Workshop OntoImage*, volume 5, page 10, 2006.

[GEB15] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv:1508.06576 [cs, q-bio]*, August 2015.

[Goo17] Google. Google Street View – Where We've Been & Where We're Headed Next, 2017.

[GPAM+14] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*, June 2014.

[GR10] Daniel J. Graham and Christoph Redies. Statistical regularities in art: Relations with visual coding and perception. *Vision Research*, 50(16):1503–1509, July 2010.

[GST03] Martin Golubitsky, Lie June Shiau, and Andrei Török. Bifurcation on the visual cortex with weakly anisotropic lateral coupling. *SIAM Journal on Applied Dynamical Systems*, 2(2):97–143, January 2003.

[GvG15] Umut Güçlü and Marcel A. J. van Gerven. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, 35(27):10005–10014, 2015.

[JSM06] Katherine Jones-Smith and Harsh Mathur. Fractal analysis: revisiting Pollock's drip paintings. *Nature*, 444(7119):E9–E10, November 2006.

[JSZK15] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial Transformer Networks. *arXiv:1506.02025 [cs]*, June 2015.

[Ken58] William E. Kennick. Does traditional aesthetics rest on a mistake? *Mind*, 67(267):317–334, 1958.

[KOI13] Hiroshi Kawase, Yoji Okata, and Kimiaki Ito. Role of huge geometric circular structures in the reproduction of a marine pufferfish. *Nature Scientific Reports*, 2013.

[Kri15] Nikolaus Kriegeskorte. Deep Neural Networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, 1(1):417–446, 2015.

[KRK14] Seyed-Mahdi Khalig-Razavi and Nikolaus Kriegeskorte. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, 10(11), November 2014.

[lat08] Evolution of the visual system is key to abstract art. Science Daily, 22 November, 2008.

[LEN08] Honglak Lee, Chaitanya Ekanadham, and Andrew Y. Ng. Sparse deep belief net model for visual area V2. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 873–880. Curran Associates, Inc., 2008.

[Lew88] J.P. Lewis. Creation by refinement: A creativity paradigm for gradient descent learning networks. In *International Conference on Neural Networks*, volume 2, pages 229–233, New York, 1988. IEEE.

[LKNW15] Richard M. Levenson, Elizabeth A. Krupinski, Victor M. Navarro, and Edward A. Wasserman. Pigeons (Columba livia) as trainable observers of pathology and radiology breast cancer images. *PLoS ONE*, 10(11), November 2015.

[LV97] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer Verlag, New York, 1997.

[Mar90] Colin Martindale. *The Clockwork Muse: The Predictability of Artistic Change*. Basic Books, New York, NY, US, 1990.

[McC91] Pamela McCorduck. *Aaron's Code*. W. H. Freeman & Co., New York, NY, USA, 1991.

[McD09] John H. McDonald. *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, Maryland, USA, second edition, 2009.

[MLB+17] Gregoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, and Klaus-Robert Muller. Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognition*, 65(Supplement C):211–222, May 2017.

[MMP12] Naila Murray, Luca Marchesotti, and Florent Perronnin. AVA: A large-scale database for aesthetic visual analysis. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2408–2415. IEEE, 2012.

[MV15] Aravindh Mahendran and Andrea Vedaldi. Understanding deep image representations by inverting them. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[MV16] Aravindh Mahendran and Andrea Vedaldi. Visualizing deep convolutional neural networks using natural pre-images. *International Journal of Computer Vision*, 120(3):233–255, Dec 2016.

[noa] Kaggle Art Image Dataset for Painter by Numbers Challenge.

[OF96] Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, June 1996.

[pix] Image from pixabay.com, CC0 license. http://www.pixabay.com.

[RB99] Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, January 1999.

[RDS+15] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vision*, 115(3):211–252, December 2015.

[RH99] Vilayanur S. Ramachandran and William Hirstein. The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6(6-7):15–51, 1999.

[Ste97] Robert Stecker. *Artworks: Definition, Meaning, Value*. Pennsylvania State University Press, 1997.

[SWF09] M. Spehr, C. Wallraven, and R. W. Fleming. Image statistics for clustering paintings according to their visual appearance. In *Proceedings of the Fifth Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging*, Computational Aesthetics'09, pages 57–64, Aire-la-Ville, Switzerland, 2009. Eurographics Association.

[SZ14] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[TMJ99] R.P. Taylor, A.P. Micolich, and D. Jonas. Fractal analysis of Pollock's drip paintings. *Nature*, 399(6735):422, 1999.

[VG03] E. Voland and K. Grammer. *Evolutionary Aesthetics*. Springer, 2003.

[VS14] Oshin Vartanian and Martin Skov. Neural correlates of viewing paintings: Evidence from a quantitative meta-analysis of functional magnetic resonance imaging data. *Brain and Cognition*, 87:52 – 56, 2014.

[Wat10] Shigeru Watanabe. Pigeons can discriminate "good" and "bad" paintings by children. *Animal Cognition*, 13(1):75–85, January 2010.

[Wei70] Morris Weitz. *The Role of Theory in Aesthetics*. Bobbs-Merrill, 1970.

[WMTR13] Wen Wu, Antonio M. Moreno, Jason M. Tangen, and Judith Reinhard. Honeybees can discriminate between Monet and Picasso paintings. *Journal of Comparative Physiology A*, 199(1):45–55, January 2013.

[WSW95] S. Watanabe, J. Sakamoto, and M. Wakita. Pigeons' discrimination of paintings by Monet and Picasso. *Journal of the Experimental Analysis of Behavior*, 63(2):165–174, March 1995.

[WSZ+17] Haiguang Wen, Junxing Shi, Yizhen Zhang, Kun-Han Lu, Jiayue Cao, and Zhongming Liu. Neural encoding and decoding with deep learning for dynamic natural vision. *Cerebral Cortex*, pages 1–25, 2017.

[YCBL14] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *arXiv:1411.1792 [cs]*, November 2014.

[YCN+15] Jason Yosinski, Jeff Clune, Anh Nguyen, Thomas Fuchs, and Hod Lipson. Understanding neural networks through deep visualization. *arXiv preprint arXiv:1506.06579*, 2015.

[YD16] Daniel L. K. Yamins and James J. DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3):nn.4244, February 2016.

Supplementary Materials

Philosophical definitions of art

Bell [Bel14] provided this definition in the art theory community in 1914:

> [Art is] "significant form [in which] lines and colors combined in a particular way, certain forms and relations of forms, stir our aesthetic emotions. These relations and combinations of lines and colours, these aesthetically moving forms, I call 'Significant Form'; and 'Significant Form' is the one quality common to all works of visual art."

A more recent philosopher of art, Eaton [Eat83] defines:

> x is a work of art if and only if (1) x is an artifact and (2) x is discussed in such a way that information concerning the history of production of x directs the viewer's attention to properties that are worthy of attention.

Major proposals in 20th century philosophy include the anti-essentialist view that art cannot be defined [Wei70, Ken58], and the institutional view that art must be defined by its relation to culture and the institutions of art. Dickie [Bar17, p.6] represents the institutional view:

> 1. An artist is a person who participates with understanding in making a work of art. 2. The work of art is an artifact of a kind created to be presented to an artworld public. 3. The public is a set of persons whose members are prepared in some degree to understand an object that is presented to them. 4. The artworld is the totality of all artworld systems. 5. An artworld system is a framework for the presentation of a work of art by an artist to an artworld public.

As can be seen, definitions such as these are not sufficient to allow an uninformed person to identify what is art. Bell's and Eaton's definitions rely on vague statements such as "'lines and colors combined an a particular way" and "properties that are worthy". The institutional view statement that "The work of art is an artifact of a kind created to be presented to an artworld public" appears to sidestep the issue of an independent definition, instead relying on the art world to identify art – i.e., 'art is what you're told it is (by the art world)'.

Image datasets

A total of 7,000 images representing five different image categories are compared in our study. Two categories are of non-art imagery and three of art imagery. The non-art image data is comprised of 1400 randomly selected Google Street View non-art, algorithmically composed images [ADF$^+$10], and 1400 randomly selected IAPR TC-12 non-art human-composed images [GCMD06]. The art imagery is comprised of: 1) 1400 randomly selected images from a collection of publicly available photographs created by an online community of digital photographers with artistic intent in response to specific aesthetic and artistic challenge criteria [MMP12], 2) 1400 randomly selected images from a publicly available collection of images of both western and non-western artwork from multiple time periods, genres, media and varied subject matter [noa] and 3) 1400 images of artwork comprised of an equally balanced number of images from seven genres in varied media from multiple time periods created to expand our analysis to multiple cultures.

Non-art, not composed image dataset: Google Street View.

The algorithmically selected camera pose of Google Street View images [ADF$^+$10] distinguishes them from images where the camera pose is chosen by an individual to capture specific content, irrespective of non-artistic or artistic intent. We therefore utilize Street View imagery as an example of non-art, not composed images for our study. Since its inception in 2007, the number of locations worldwide with Street View coverage has steadily increased [Goo17]. For the purposes of this study we utilize the Google Street View API to collect imagery from both cities and national parks on all major continents to create a randomized set of 1400 images containing a balance of 50% urban and 50% rural Street View imagery. A list of image location (country, city, or national park name), lat/lon coordinates, and heading (degrees) can be obtained from the authors. Through inspection of image metadata we excluded all images that were user-contributed rather than sourced from Google Street View's automatic data collection process.

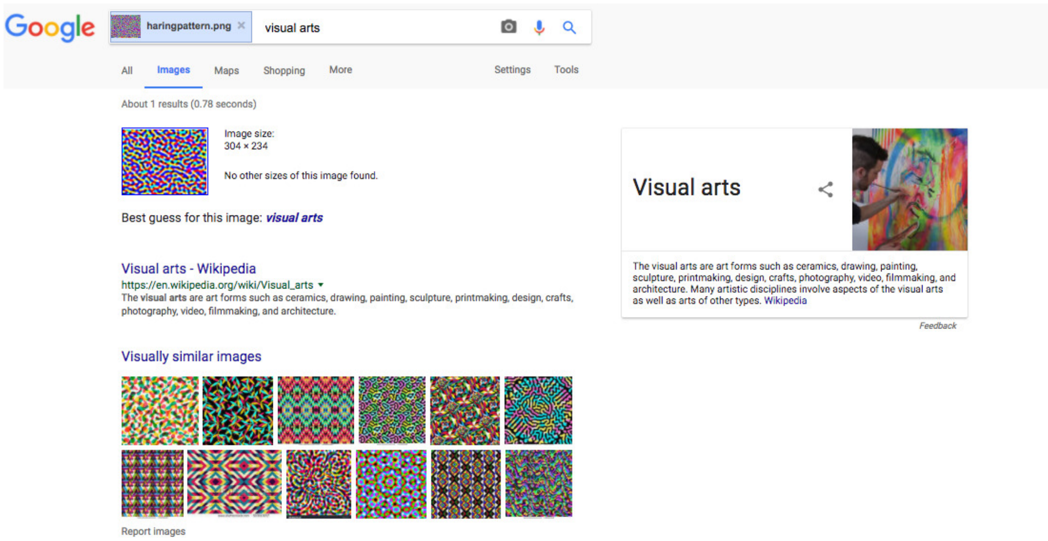
Non-art, composed image dataset: IAPR TC-12 Benchmark. The IAPR TC-12 benchmark is a publicly available image dataset [GCMD06] comprised of 20,000 still images composed and captured by individuals, for non-artistic purposes, during the course of activities common to contemporary life in geospatially diverse locations in over 30 countries. [GCMD06] Images are annotated in three languages for semantic and visual content. Images containing similar visual content vary in "illumination, viewing angle and background" [GCMD06, p. 13]. While initially constructed for evaluating both text and image-based retrieval methods, its high quality color photographs of natural scenes from multiple aspects of contemporary life (e.g. sports, cityscapes, landscape, animals, action shots, people etc.), each with multiple objects per scene, make it suitable as an example of non-art, yet composed, imagery. We randomly select 1400 images from the IAPR TC-12 dataset to create a non-art, composed image dataset for our study.

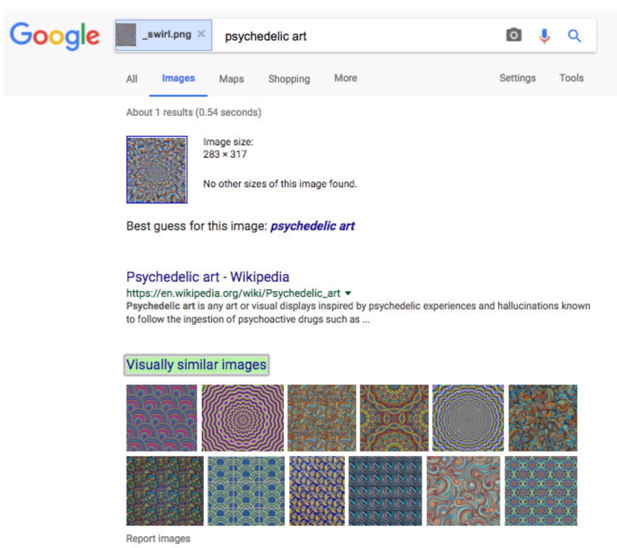
Art, composed image dataset: AVA subset. The AVA, is a publicly available large-scale database for aesthetic visual analysis comprised of approximately 255,000 photographs created by an online community of digital photographers in response to specific aesthetic and artistic challenge criteria [MMP12]. Images are annotated semantically, aesthetically, and with information for the specific challenge they are created in response to. We randomly select 1400 images from the AVA dataset to create an art, composed image dataset for our study.

Art, Western and Non-Western Artwork: Kaggle Challenge Dataset. The Kaggle "Painter by Numbers" Challenge Dataset [noa] is comprised of over 100,000 (Testing and Training) images of Western and Non-Western artworks from multiple time periods, genres, media and varied subject matter collected from WikiArt.org and WikiPedia. Image annotations include artist, title, genre, style, and date. The data is annotated with over 50 artistic genres representing both western and non-western artistic movements and styles throughout history. We randomly select 1400 images from the "Painter by Numbers" dataset to create an art, western and non-western aggregated image dataset for our study.

Art, Individual Genres: Multiple Cultures. To extend our analysis to art from multiple cultures we include a dataset comprised of an equal number of images from seven genres in varied media from multiple time periods: Australian Aboriginal art (200), Huichol Mandala bowls (200), Impressionism (200), Islamic tile art (200), Maori tattoos (200), Op Art (200) and Ukiyo-e (200). The Impressionism images were selected from the Impressionism category in the Kaggle dataset, with some categorization errors manually removed. Other images were located via Google Image Search and individually downloaded. A list of individual image urls can be obtained from the authors.

Examples of similar human art and patterns

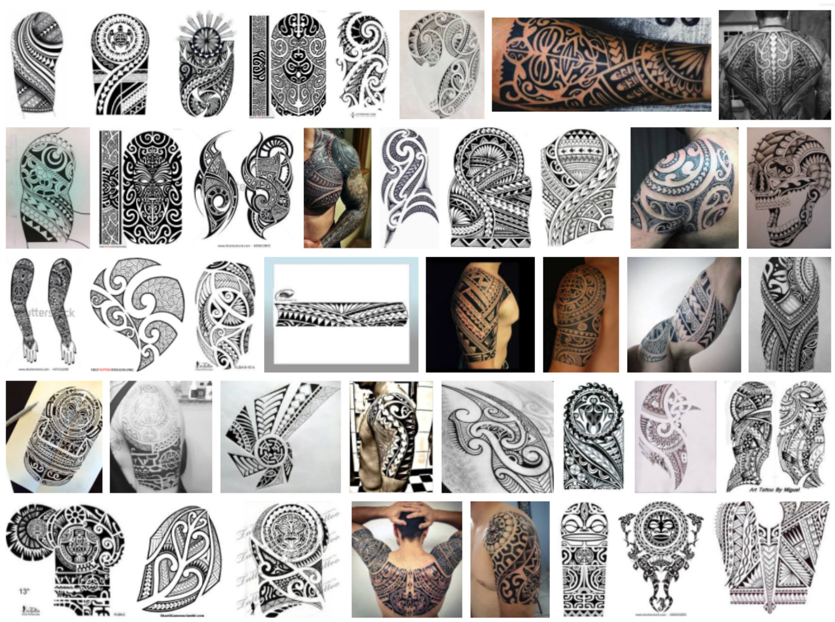




Examples of image categories: google image search “Islamic tile art”

Examples of image categories: google image search “aboriginal art”

Examples of image categories: google image search "Maori tattoo design"



Examples of image categories: google image search "Huichol mandala bowl"

Examples of image categories: google image search “Op art”