

Optimal construction of montages from mathematical functions on a spectrum of order-disorder preference

Kate Smith-Miles & Mario Andrés Muñoz

To cite this article: Kate Smith-Miles & Mario Andrés Muñoz (2022) Optimal construction of montages from mathematical functions on a spectrum of order-disorder preference, *Journal of Mathematics and the Arts*, 16:4, 347-373, DOI: [10.1080/17513472.2022.2139663](https://doi.org/10.1080/17513472.2022.2139663)

To link to this article: <https://doi.org/10.1080/17513472.2022.2139663>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 11 Nov 2022.



Submit your article to this journal



Article views: 124



View related articles



View Crossmark data

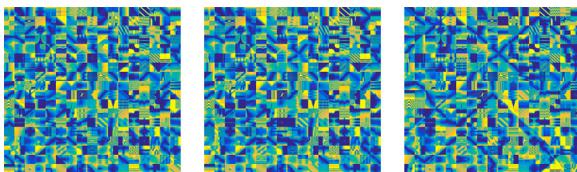
Optimal construction of montages from mathematical functions on a spectrum of order–disorder preference

Kate Smith-Miles  and Mario Andrés Muñoz 

School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC, Australia

ABSTRACT

We previously generated diverse mathematical functions that are difficult for optimization algorithms. Represented as 2D contour plots, each image depicts a ‘blue river’ running through an intricate landscape. This paper describes the challenge of constructing an aesthetic montage of these images. A survey revealed a spectrum of tastes, divergent in preference from order to disorder, considering the structure created by connecting these ‘blue rivers’. A new artwork, *Negentropy Triptych*, was created to depict this spectrum by manually swapping images from a random arrangement, guided by human eye to enhance or destroy the structure. An optimization algorithm automates the process, with the results of its efforts to emulate the artistic vision presented and discussed. The challenges faced by the algorithm, despite exploring several objective functions, highlight the difficulties of capturing the goals that a human decision-maker can easily achieve. Therefore, machine learning of these goals is a promising future direction.



ARTICLE HISTORY

Received 30 August 2021

Accepted 14 July 2022

KEYWORDS

Evolutionary programming;
optimization problems;
exploratory landscape
analysis; genetic algorithms;
evolved mathematical
functions; aesthetics; mosaic
arrangement; global
structure

1. Introduction

This paper describes the creation of an artwork based on mathematical functions, and the challenge to create a computer algorithm to automate the construction process to strengthen its aesthetic appeal. The key requirement to achieve this goal is to mathematically define the aesthetic qualities that guided the human artistic vision, and embed the same objective into the algorithm. Since aesthetic judgement of art tends to be subjective, the constructed artwork reflects a spectrum of aesthetic tastes based on considerations

CONTACT Kate Smith-Miles  smith-miles@unimelb.edu.au  School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

such as individual preferences for order and complexity. Before discussing the construction process of our mathematical artwork, we first review the key factors from the relevant literature influencing aesthetic preferences in both mathematics and the visual arts.

In mathematics, the concept of aesthetics has been studied by surveying mathematicians about the properties of their favourite mathematical proof (Inglis & Aberdein, 2014), revealing that mathematics can include a significant amount of complexity without detracting from its perceived beauty. The perceived role of complexity in aesthetic appreciation of visual art however is not without controversy. The Fechner-Berlyne tradition (Cupchik, 1986) posits that beauty requires a balance between order and complexity, with ‘unity in variety’ being a mechanism for achieving such a balance. If such balance is not achieved, the aesthetic value is compromised, resulting in emotional responses such as confusion or boredom that impact perception of beauty (Berlyne, 1970). The belief that complexity and order must be balanced to achieve the required emotional response for aesthetic appreciation is well explained by Arnheim (1966): ‘Complexity without order produces confusion. Order without complexity causes boredom’. In contrast, the Birkhoff (1933) tradition posits that aesthetic value is a monotonic function of order and complexity, and that individual preferences reflect their knowledge and cognitive state, so that aesthetic appreciation can be learned with training and is not merely an emotional response. We refer the interested reader to a recent review by Geert and Wagemans (2020) for a discussion of the divergent theories about the role of complexity and its interplay with order in aesthetic appreciation.

Many studies have supported the view that people’s perception of aesthetic value of visual art is also influenced by a range of factors beyond complexity and order, to include other considerations such as symmetry, colour combinations, balance points, spatial properties of objects and shapes, and arrangements of shapes (Palmer et al., 2013). Indeed, when these other considerations are considered, it appears possible to unite divergent theories if different types of complexity are defined (Nadal et al., 2010): one related with the amount and variety of elements, another related with the way those elements are organized, and a third type of perceived complexity due to asymmetry.

Recent efforts in the field of computational aesthetics have seen machine learning methods used to predict human aesthetic responses (Carballal, Fernandez-Lozano, Rodriguez-Fernandez et al., 2019), and to explain aesthetic taste (Wang et al., 2019) based on large collections of annotated images. Other studies have explored predictive modelling that considers the knowledge and emotional state of the human subject – combining the hedonic analysis (Fechner-Berlyne) and cognitive analysis (Birkoff) perspectives – along with different types of complexity and order, showing that aesthetic appreciation can be understood as a combination of both aesthetic emotion and aesthetic judgement (Leder et al., 2004).

Recent directions (Geert & Wagemans, 2020) have started to explore whether an individual’s personality has influence on their perceptions of complexity and the amount of order they require for aesthetic appreciation to be attained. While some researchers have argued that personality effects cannot adequately be modelled as an aesthetic science (Markovic, 2010), others report that much can and has been done through empirical studies of aesthetics using behavioural methods (Palmer et al., 2013; Swami & Furnham, 2020), to advance understanding of human aesthetic response to and preferences for visual stimulation. Research into the interplay between an individual’s personality traits, emotional or



cognitive driven perceptions of complexity and order, and their requisite balance point to achieve aesthetic satisfaction is still in its infancy (Chamorro-Premuzic et al., 2010; Rawlings & Furnham, 2000).

With these background concepts reviewed, this paper considers the role of order and complexity in creating aesthetically pleasing montages of computer-generated mathematical functions. In this context, it is important to understand how the complexity of the mathematical functions are likely to be perceived by a general (non-mathematical) audience, and how a montage structure can be used to create the required ‘unity in variety’ for high aesthetic value. In recent work (Muñoz & Smith-Miles, 2020a), we have used Genetic Programming to evolve a novel set of mathematical functions. The *scientific motivation* for creating these functions was to stress test optimization algorithms that seek to find the global minima of functions. Challenges are created for optimization algorithms when the landscape is very rugged with numerous local minima, or contains flat plateaus, creating convergence traps in the search for the lowest point in the landscape. Our new collection of 2D and 10D test functions, has contributed greater diversity to the benchmark test suites used in the literature, exhibiting unique structures and complexities that have not previously been tested on optimization algorithms.

It is the 2D test functions that form the basis of the artwork described in this paper, since these functions of two variables (x_1 and x_2) can be viewed as contour plots in the 2D plane (x_1, x_2), with a colour scale used to indicate the function value $y = F(x_1, x_2)$. Figure 1(a) displays two examples of simple black box optimization (BBO) test functions from existing benchmarks for illustration of such contour plots. The minimal points of the function value are represented with the colour blue, and more interesting functions have resulting landscapes that tend to present ‘blue rivers’ or pockets of blue surrounded by steep and rugged slopes that disguise the location of the minimal points. In addition to presenting challenges for optimization algorithms, our newly evolved functions create images that individually possess the constituent ingredients of visual beauty, at least for mathematicians: intricacy, symmetry, and an elegance when appreciating the surprising complexity generated by a simple process of combining two variables in a novel functional form. The *artistic motivation* for combining these 306 images into an 18×17 element montage was to highlight the ‘unity in variety’ that can be generated when the full diversity of the collection is displayed.

The key question tackled in this paper, both artistic and scientific, is *how to arrange the individual images in a montage that is aesthetically pleasing?* The objectives of this study are therefore to:

- (1) understand how the interplay between order and complexity of the elements of any proposed montage arrangement influence the aesthetic value, as perceived by different human subjects;
- (2) construct, by human eye, montages that are aesthetically appealing based on order-disorder preferences;
- (3) construct, by an optimization algorithm, montages using various objective functions that seek to capture the same sense of aesthetic value.

This paper reports on the process of creating the resulting artwork that explores this key question, and its solution in the form of a triptych of images, acknowledging that aesthetic taste is subjective and spans a spectrum of order-disorder preferences.

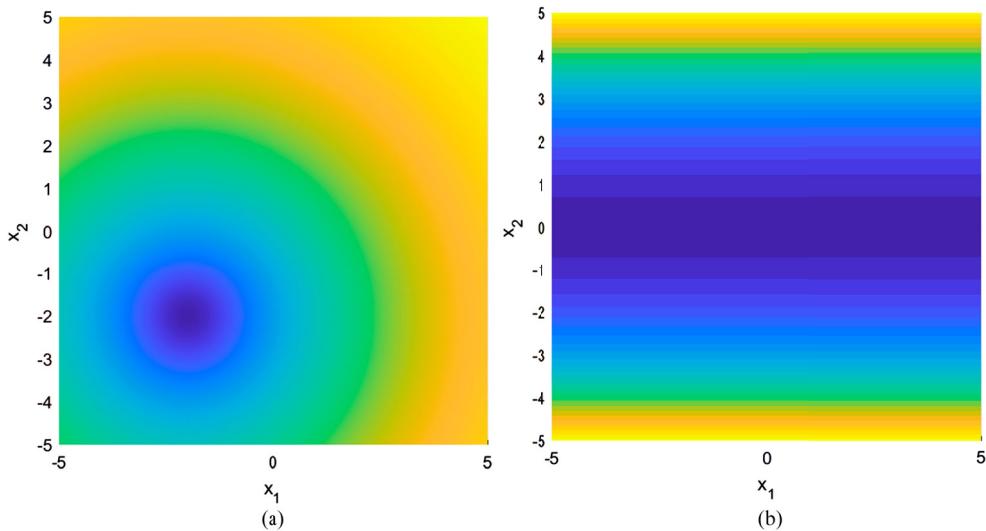


Figure 1. 2D contour plots of example functions: (a) a sphere function which is easy for most BBO algorithms due to its convexity, and (b) a truncated ellipse function, which is more challenging due to many flat plateaus that cause BBO algorithms to become trapped in local minima. Dark blue represents the global minimum of the functions given by the Equations (1) and (2).

An informal survey of around 50 friends and colleagues of the authors (majority non-mathematicians) quickly revealed that aesthetic judgement about the montage arrangement reflected a strong preference for order or disorder that was noticeably aligned to personality traits. In particular, those who expressed strong preference for order and structure much preferred montages which created ‘unity in variety’ through the continuity of ‘blue rivers’ that run through and between images, creating a meandering global background structure to unify the montage. Others expressed a strong dislike of the imposition of such order. Our aim was not to test the hypothesis of strong correlations between these divergent aesthetic taste and personality traits via psychological or behavioural methods – although that is certainly an interesting future use of our mathematical images. Instead, our aim in this paper is to acknowledge the alignment of this untested hypothesis with the aesthetics literature (Geert & Wagemans, 2020), and proceed to tackle the artistic and scientific challenge of creating an artwork that acknowledges this spectrum of aesthetic taste, from disorder to order, supported by the literature. Manual swaps of a random arrangement guided by the authors’ visual judgements produced montages that enhanced or destroyed ‘blue river’ connectivity; but automating this process posed some computational challenges to maximize aesthetic appreciation at both ends of the spectrum. While it is relatively straightforward to destroy ‘blue river’ structure by maximizing entropy in the arrangement of images in the montage, it is more challenging to automate the creation of a satisfying connectivity of blue rivers, reportedly valued as aesthetic by those preferring order.

In general, we can think of our algorithmic challenge as an *image mosaicing* task, where multiple images, often overlapping, are aligned into a large composition representing a larger field-of-view, without compromising the spatial resolution (Ghosh



& Kaabouch, 2016; Pandey & Pati, 2019). Interest in mosaicing has increased in recent years due to the growing number of applications, such as satellite and medical imaging, virtual reality, image editing, visual effects, and surveillance (Pandey & Pati, 2019). Mosaicing is considered a special case of image reconstruction, where the images are related by planar holography, and involves four processing steps:

- (1) registration, or identifying geometric correspondence between a part of images;
- (2) reprojection, or the alignment of the images into a common coordinate system using the computed geometric transformations;
- (3) stitching, or the overlay of the aligned images on a larger canvas by merging pixel values of the overlapping portions and retaining pixels where no overlap occurs; and
- (4) blending, or minimizing the discontinuities in the global appearance of the mosaic.

Because of the lack of exploitable overlap in our images and the fact that we do not attempt to minimize discontinuities, registration is the only step from mosaicing of interest. Moreover, techniques that are used for *image compositing* of non-overlapping images, e.g. blending the objects from one image into the background of another (Pandey & Pati, 2019), are not relevant. Image registration can be classified into ‘area’ and ‘feature’ based techniques (Ghosh & Kaabouch, 2016). The former class relies on comparing the pixel values between two images, as to identify how much they match. Techniques in this class are those based on minimizing the intensity difference, or maximizing the correlation or the mutual information. The latter class relies on extracting salient features, such as isolated points, continuous curves or connected regions, common to the images. State-of-the-art feature extractors are the Harris corner detector, Features from Accelerated Segmented Test (FAST) corner detector, Scale Invariant Feature Transform (SIFT), and Speeded Up Robust Features (SURF). Generally, ‘feature’ based algorithms are more accurate than ‘area’ based ones, as long as the features are extracted correctly (Ghosh & Kaabouch, 2016; Pandey & Pati, 2019). For our work, we propose two ‘area’ based approaches and one ‘feature’ based approach.

The remainder of this paper is organized as follows: In Section 2, we briefly describe the evolutionary programming approach used to evolve the novel set of diverse mathematical functions, and present a selection of the 306 two-dimensional contour plots to illustrate their complexity, order, symmetry, and other properties that influence aesthetic value. In Section 3, we discuss the aesthetic considerations arising in the construction of a montage of the 306 individual functions into an image array, and the feedback from human subjects about aesthetic value of various arrangement choices, which revealed two divergent opinions about the value placed on order. The section concludes with the presentation of the resulting artwork, *Negentropy Triptych*, which acknowledges the spectrum of aesthetic taste reported in the literature (Geert & Wagemans, 2020) and observed in our informal survey, by depicting the emergence of order from disorder. Section 4 proposes an optimization algorithm to evolve the optimal arrangements of the montage for each end of the spectrum of order preferences, using a variety of objective functions to guide the search for optimally aesthetic montages. A discussion about the human versus algorithm efforts and the considerations to improve further the algorithm’s understanding of human aesthetic judgement concludes the paper in Section 5.

2. Evolving images of diverse mathematical functions

In this section, we briefly review the scientific motivation for evolving novel mathematical functions in our previous work (Muñoz & Smith-Miles, 2020a), stemming from the need to test the performance of optimization algorithms on test functions with diverse characteristics. The methodology used to establish the lack of diversity of existing test functions, and then to evolve new test functions to add greater diversity and challenge for optimization algorithms, is briefly summarized. From this scientific motivation, a set of diverse and novel 2D contour plots have been generated that form the basis of the artwork discussed in subsequent sections. A small selection of such images is presented in this section and the properties of these images are discussed.

2.1. The need for diverse test functions for black-box optimization

Black box optimization (BBO) is the mathematical challenge of finding the minimum or maximum of an unknown function based only on a limited sample of function evaluations. Unlike other kinds of optimization problems, the defining characteristic in BBO is the absence of an algebraic or analytical function that indicates how the decision variables influence the objective function to be optimized. Consequently, no gradient information is available, and the decision space must be searched only by guiding the sampling process based on an efficient use of function evaluations (Audet & Hare, 2017). Such BBO problems arise in practical applications where the function evaluation may involve conducting physical experiments, or in computation settings such as machine learning, where the function evaluation involves running a prediction or simulation model. Due to its practical importance, many optimization algorithms have been proposed for solving BBO problems, and their performance is assessed based on how close their best objective function is compared to the globally optimal objective value after a budgeted number of function evaluations (Muñoz & Smith-Miles, 2017). Most noteworthy in this experimental evaluation of algorithms is that the function is known analytically for the purposes of evaluating the solution quality, but the algorithm is only given a sample of points on which to evaluate the function, and does not know the functional form, which may be continuous or discontinuous in nature.

Two simple examples are provided in Figure 1 above where the well-known sphere function and a truncated ellipse function, defined by Equations (1) and (2) respectively, establish the truth of a functional form; but an optimization algorithm is only given a small sample of n triplets (x_1, x_2, y) to decide where the global minimum of the function lies. These examples also illustrate the concept of a known mathematical function being represented as an image in the form of a contour plot: with horizontal and vertical axes representing the x_1 and x_2 variables respectively, and the colour scale used to represent the function value y , in this case with minimal values in blue and maximal values in yellow. The two examples illustrate different landscapes that are created by functions: the sphere function depicts a ‘blue puddle’ at the location of the global minimum $(-2, 2)$ with symmetrical increases in the values of this convex function; while the truncated ellipse function (where $\lceil \cdot \rceil_p$ indicates the rounding of a value to p decimal places) depicts a ‘blue river’ structure of global minima that run horizontally across the contour plot, with long and narrow plateaus



or ridges that step up in the vertical direction.

$$y = (x_1 + 2)^2 + (x_2 + 2)^2 \quad (1)$$

$$y = \lceil x_1^2 \rceil_3 + 10^6 \lceil x_2^2 \rceil_0 \quad (2)$$

Depending on the shape and complexity of the landscape of a function, e.g. ruggedness, large plateaus, sharp funnels etc., different optimization algorithms may show strengths or weaknesses. Critically, if we do not test algorithms on a suite of test functions with diverse characteristics, there is a danger that the weaknesses of algorithms may never be exposed. Our scientific motivation commenced therefore from the quest to understand the sufficiency of existing test functions for BBO, and to generate new test functions with more diverse characteristics.

2.2. Assessing diversity of existing test functions

Exploratory Landscape Analysis (ELA) metrics (Mersmann et al., 2011) have been proposed to characterize critical properties of test functions that create challenges for optimization, such as ruggedness seen in multimodal functions with many local minima, large flat plateaus that can cause algorithms to converge prematurely, and deceptive landscapes with steep funnels that are out of place with the surrounding landscape.

In our previous work (Muñoz & Smith-Miles, 2020a), we constructed a 2D visualization of all existing test functions within the well-studied COCO (Comparing Continuous Optimizers) benchmark test suites for BBO, using a recently developed methodology known as Instance Space Analysis (Muñoz et al., 2018; Smith-Miles et al., 2014). The instance space visualization relies on various ELA metrics (summarized in the Appendix) to quantify the similarities and differences between characteristics of different test functions, representing each test function as a unique feature vector in a high dimensional feature space, before projecting all test functions to a 2D instance space using an optimized projection equation (see Appendix for a summary and (Muñoz & Smith-Miles, 2020a) for full details).

The resulting instance space visualization of the existing COCO benchmark test functions is shown as grey points in the background of Figure 2. It is clear that the existing COCO benchmark test functions occupy a small region within the instance space, indicating their lack of diversity despite originating from 24 different basis functions intended to represent an array of challenges for BBO algorithms. In order to expand the diversity of the benchmark test functions, the blue ‘cross’ marks (\times) and the red ‘plus’ marks ($+$) respectively indicate the location of target points and new functions evolved to lie as close to the target points as possible, as described in the following section.

2.3. Evolving new test functions to increase diversity

Previously, we proposed a Genetic Programming (GP) approach to generate new test functions whose feature vector, when projected to the 2D instance space via Equation (A1), lies as closely as possible to a user-defined target as shown in Figure 2. We briefly summarize the approach here, and refer the reader to Muñoz and Smith-Miles (2020a) for full implementation details.

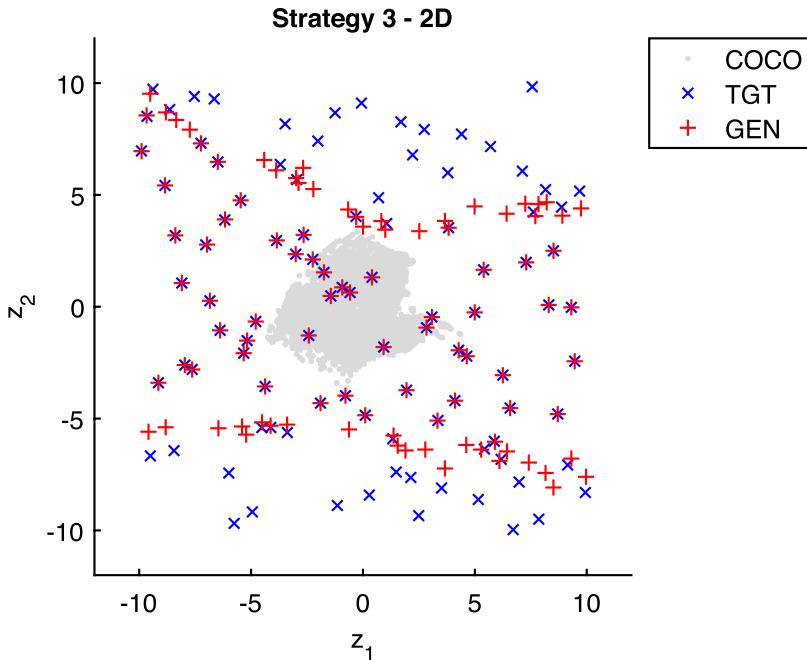


Figure 2. Instance Space showing location of existing COCO benchmark test functions as grey dots, with their locations given as coordinates (z_1, z_2) of axes determined given by the projection Equation (A1). The \times marks indicate the target location of new functions evolved to add diversity, and the red+ marks indicate the location of such generated functions as reported in Muñoz and Smith-Miles (2020a).

We represent a BBO test function by a binary tree whose leaves are variables or constants, and nodes represent a mathematical operator applied to the variables, chosen from a permitted vocabulary: $\{\times, +, -, x^2, \sin, \cos, \tanh, e^{-x}, e^x, \lfloor x \rfloor\}$, where x represents a variable at an adjoining leaf. For example, the bivariate function $y = 2 \sin(x_1) + \lfloor x_2 \rfloor \cos(x_1 x_2)$ is represented by the binary tree shown in Figure 3. We then used GPTIPS (Genetic Programming Toolbox for the Identification of Physical Systems) v2.0 (Searson et al., 2010), and defined the fitness function as $\|\mathbf{z}_t - \mathbf{z}_g\|^2 + 2^{|D_t - D_g|}$, where \mathbf{z}_t and D_t are the location vector and the dimensionality of the target function, and \mathbf{z}_g and D_g are the location vector and the dimensionality of a candidate generated function. The first term in the fitness function penalizes functions with a projected feature vector too far from the target point, while the second term penalizes functions utilizing more or less variables than the dimensionality of the functions we seek to generate. While we generated 2D and 10D functions, it is the 2D functions that we focus on here, since their contour plots create the artwork discussed in the remainder of this paper. Finally, we proceeded to generate hundreds of new test functions at target locations, a sample of which are shown in Figure 2, highlighting those beyond the convex hull defined by the existing COCO benchmarks. Several hundred new 2D and 10D test functions have been generated, and have added significant diversity to the available test suites. These functions are available to download as MATLAB functions, and the 2D functions can also be downloaded as contour plot images (Muñoz & Smith-Miles, 2020b).

Figure 4 presents a selection of the new 2D test functions, highlighting their intricacy and surprising complexity given the limited vocabulary they were permitted to utilize as

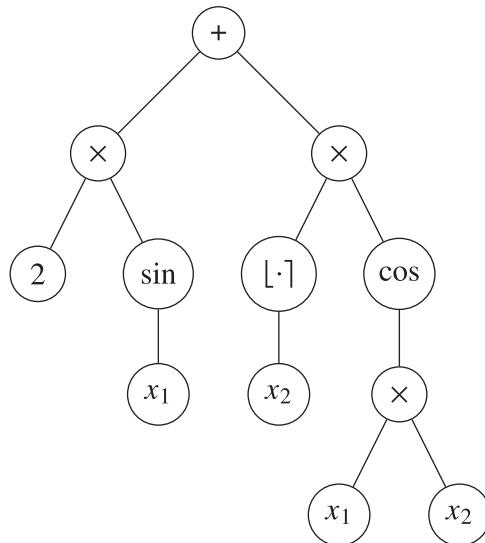


Figure 3. Example of a binary tree representation of the function $y = 2 \sin(x_1) + |x_2| \cos(x_1 x_2)$.

operators on two variables. Most notably, this selection of images reveals properties such as asymptotic behaviours, periodicity, and nonlinearities created by the trigonometric and exponential functional forms permitted in the genetic programming structures. In addition to adding diversity to the benchmark test suites, these new test functions are also more challenging for the BBO algorithms tested (Muñoz & Smith-Miles, 2020a), since they come from a part of the instance space where several algorithms have weaknesses. It is clear from inspecting these images why BBO algorithms would find many of these test functions challenging, since their task is to locate the global minima, represented by the darkest blue, which are often hidden from the search path by premature convergence to the many local minima, plateaus, or deceptive tricks that complicate the landscape. Compared to the relatively simple examples shown in Figure 1(a,b), it is clear that the ‘blue river’ structures of these new test functions are much more complex for BBO algorithms, and more interesting as visual artworks.

Scanning through hundreds of such images in the quest to identify the most aesthetically appealing one for printing as a poster, we realized the difficulty of selecting just one image; the opportunity to highlight the collection’s diversity by creating a montage became more appealing. To this end, we shortlisted 306 of the more mathematically interesting images to commence work on a new artistic outcome of the research, as described in the remaining sections of this paper.

3. Aesthetic consideration in montage arrangement

Random placement of the 306 images into an 18×17 array immediately highlighted the tendency for the human eye to be drawn to any regions where dark blue connects the individual images together. The random arrangement shown in Figure 5 illustrates this phenomenon, where the imbalance created by too much contiguous dark blue essentially distracts the viewer from noticing the wider perspective of the array, and draws attention

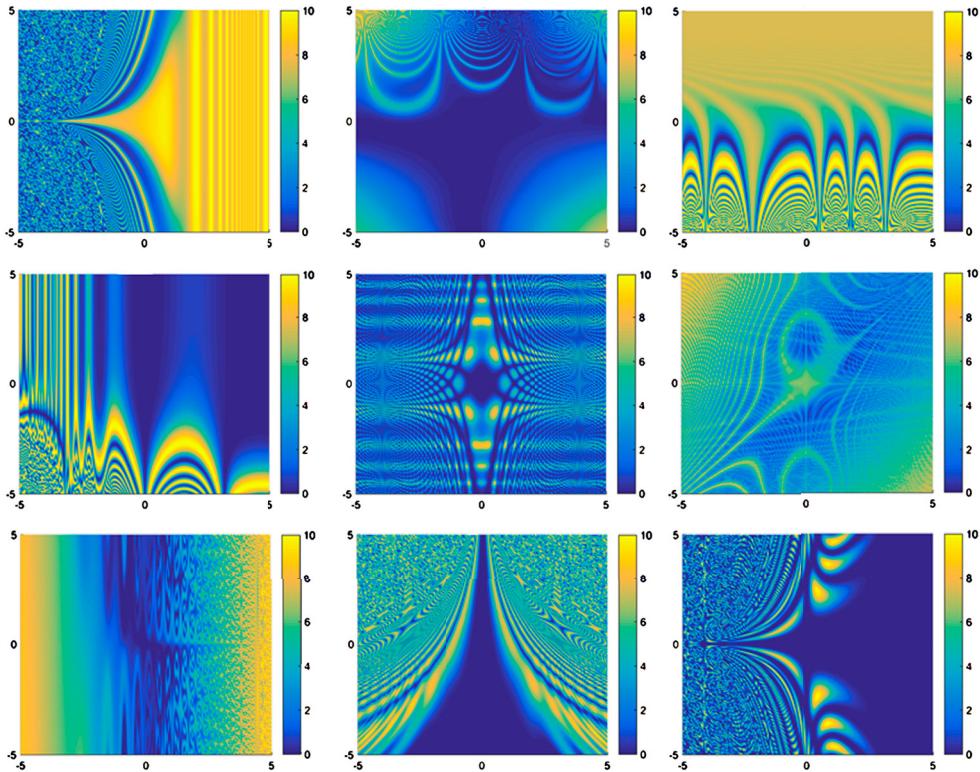


Figure 4. A sample of 9 contour plots from hundreds of new 2D test functions generated by genetic programming to lie at target points across the instance space. Each contour plot shows the value of the bivariate function in the (x_1, x_2) plane, with colour indicating the magnitude of the function value. These unique functions present considerable challenges for BBO algorithms due to their complexity and difficulty locating global minima, shown as dark blue.

to small regions that start to dominate the montage, such as the blue cross located around rows 5–6 and columns 9–10.

In order to arrive at a more balanced randomization, we generated 10 different random montages, and conducted an initial informal survey of approximately 50 friends and colleagues (the majority of whom were not mathematically trained beyond high school level) to ask which they found most aesthetically pleasing. While there was little consensus about which they liked best, there was consensus that none of the randomly generated mosaics were particularly aesthetic. Further insights were gained when survey participants were asked to explain what they didn't find appealing, so that we could try to improve the aesthetic appeal of the montage construction. Two sources of disappointment emerged. Some people reported disappointment that small fragments of 'blue river' connectivity accidentally created by some random configurations had terminated abruptly, and did not continue into larger rivers that meandered continuously around the montage. Other people reported that any regions of dominant blue were a visual distraction that created imbalance from admiring the diversity as a whole.

Based on this initial feedback, we decided to improve the aesthetic appeal of the montage by swapping some individual image locations to either increase or decrease the 'blue river'

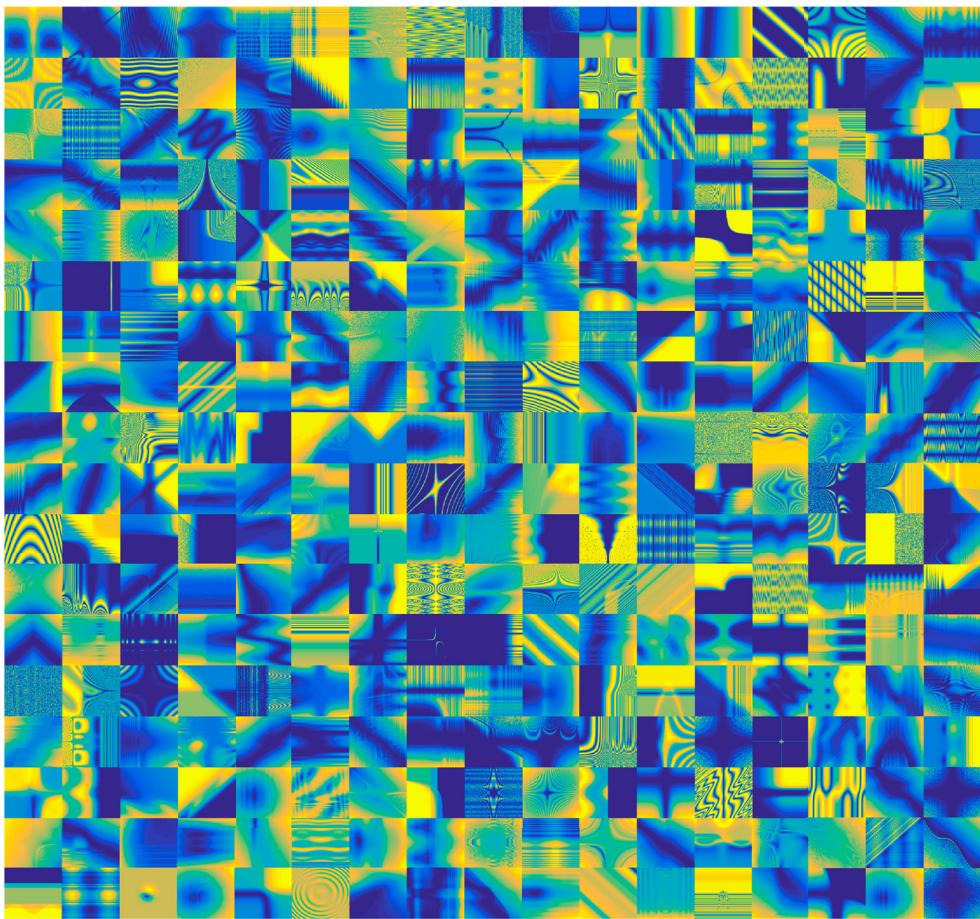


Figure 5. A random arrangement of 306 contour plot images into an 18×17 array.

connectivity. Guided by the human eye, we moved images with significant blue features to either join previously disconnected blue rivers, or to break up contiguous blocks of blue. After around 30 swaps for each objective our eyes felt that convergence had been reached, in the sense that while there were probably more swaps that could increase or decrease blue river connectivity, the gains in continuing were not likely to be perceptible, and our eyes deemed that the two exercises had resulted in two modifications to the montage that should better reflect the divergent aesthetic preferences of those surveyed.

The result is *Negentropy Triptych*, shown in Figure 6, which has the initial random montage in the centre, and the modified montages with less blue river connectivity (left) and more blue river connectivity (right). Surveying the same group of friends and colleagues again to now ask whether they found the left or right montage more aesthetically appealing revealed that they were almost 50:50 split between the left and right extremes. Those who preferred the right image found the enhanced blue river connectivity to be much more satisfying than the partial structures seen in the random image, and enjoyed their eye tracing the blue rivers through the montage as a background global structure. Those who preferred the left image reported that it was much more balanced, without the distraction of the blue

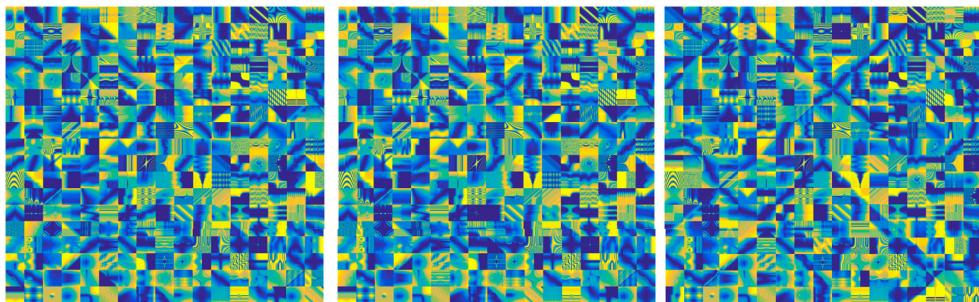


Figure 6. A triptych of montages depicting the emergence of order from disorder, known as *Negentropy Triptych*. The middle montage is a random arrangement of 306 images into an 18×17 array. The right montage has swapped some images to enhance the connectivity of 'blue rivers' as a background global structure. The left montage has swapped some images to destroy the accidental connectivity of the random arrangement, creating a more balanced composition.

Table 1. Demographic data for 48 human subjects informally surveyed for aesthetic preference for left (disordered) or right (ordered) images in *Negentropy Triptych*.

Age	Disordered image		Ordered image		Total
	Male	Female	Male	Female	
< 25 years	2	4	2	3	12
25–50 years	2	8	5	6	21
> 50 years	2	7	1	5	15
Total	6	19	9	14	48

rivers, which were described by some as 'hideous' and 'childish'. One subject explained that she found the right image to be quite insulting, with the imposition of an ordering to help the viewer make sense of the complexity of the montage, whereas she preferred to admire the complexity with the sense of order provided only by the grid structure itself. Since these human subjects were close friends and colleagues, we had the opportunity to also observe some correlations between their aesthetic preferences and personality traits. Table 1 provides some demographic data for the informal human survey, and their preferences for the left or right image. There appears to be no evidence of significant bias in either age or gender, given the limited sample size, to explain aesthetic preference and a natural hypothesis emerges, supported by the literature, that personality may well be playing a role in their aesthetic preferences.

Negentropy Triptych depicts the emergence of order (right) from disorder (left) – the negative of entropy, also known as negentropy – from the 'unity in variety' perspective, with unity provided by blue river connectivity. Of course, the order an individual seeks to provide balance with complexity can occur on different scales. From a blue river connectivity perspective, order is provided by imposing a background global structure to unify the images. From this perspective, the left image appears to be disordered. If we consider that the array itself provides an order from which to appreciate the diversity of the individual images, any attempt to distract our attention from the whole array, by drawing our eye to particular regions of the montage with a blue dominance, could be considered to create disorder. While the aesthetics literature explores the relationship between order and complexity, and there is acknowledgement that appreciation of complexity may depend on



cognitive state, our investigation also supports the notion that preference for order or disorder may not be an absolute concept, and the scale at which order is created and appreciated, micro or macro, may be a relevant factor.

4. Optimizing aesthetic appeal of montage arrangements

In this section, we present the methodology and results of several algorithmic efforts to swap images from a randomly arranged montage in order to either increase or decrease the order created by the background ‘blue river’ connectivity. If we can suitably quantify the presence or absence of such connectivity, it will be possible to use the power of optimization algorithms to efficiently explore the huge combinatorial search space of possible image swaps in order to achieve the required aesthetic objectives. We first introduce the computational representation of an individual montage, followed by a greedy heuristic for optimizing the arrangements. Next, we discuss the three tested objective functions based on mutual information, connected area, and pattern matching. Finally, the resulting montages are presented.

4.1. Computational representations of a montage

We represent a montage as a 18×17 matrix χ , whose element $\chi_{k,l}$ corresponds to a unique image from the set $\{X_1, \dots, X_{306}\}$. Each image is a 520×590 matrix, with two different types of images created for each function – a full colour version and a binary (black and white) version, whose elements $x_{i,j}$ are defined respectively as either:

- (1) colour: $x_{i,j}$ values are integers in the range $[0, 2^{16} - 1]$, representing a level within a 16-bit indexed colour spectrum based on MATLAB’s ‘parula’ colour map; or
- (2) binary: $x_{i,j}$ values are a binary value with ‘1’ representing a ‘dark enough’ blue hue, which corresponds to a colour level in the range $[0, 10^4]$, and ‘0’ otherwise.

The complete montage has 9360×10030 pixels, and it is denoted by \mathbb{X} . Figure 7 illustrates these representations, with the top left corresponding to the true colour image, the top right to the indexed spectrum, and the bottom left to the binary version. A visual inspection of the latter one reveals that the complex patterns each image contains could be simplified to ‘primitive’ shapes, i.e. binary matrices that represent the most salient feature of an image, as shown on the bottom right of Figure 7. We specified a vocabulary of 26 primitives, $\Pi_{(\cdot)}$, two of them being black or white squares, i.e. fully ‘0’ or fully ‘1’, and the remaining 24 being illustrated in Figure 8. We define the probability that an image looks like a specific primitive $\Pi_{(\cdot)}$ as:

$$\phi(\Pi_{(\cdot)}) = \frac{1}{520 \times 590} \sum_{i=1}^{520} \sum_{j=1}^{590} \mathbb{I}(x_{i,j} = \pi_{i,j}) \quad (3)$$

where \mathbb{I} denotes the indicator function and $\pi_{i,j}$ denotes a pixel of the primitive $\Pi_{(\cdot)}$. The selected primitive is the one that maximizes this probability, as illustrated on the bottom right of Figure 7. The montage can then be represented as a 18×17 matrix $\tilde{\chi}$, whose element $\tilde{\chi}_{k,l}$ corresponds to a primitive from the set $\{\Pi_{(1)}, \dots, \Pi_{(26)}\}$.

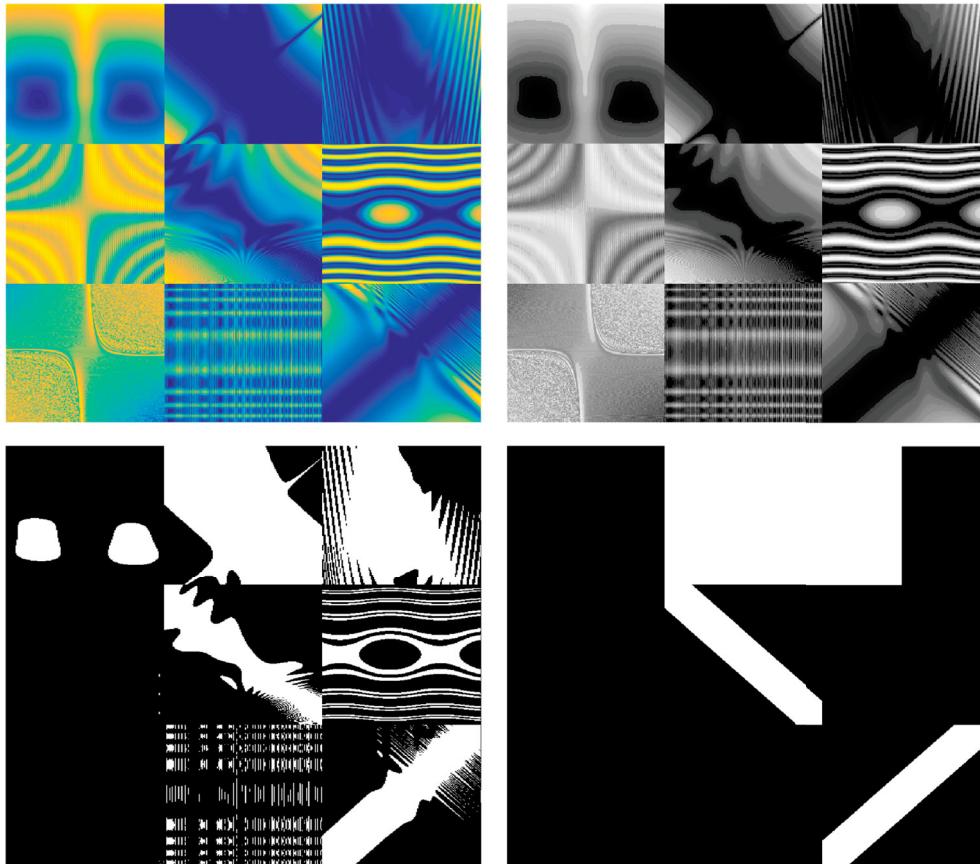


Figure 7. Representations employed to computationally describe the dark blue regions of a 3×3 portion of a montage. Top left corresponds to the true colour image, top right to the 16-bit indexed colour spectrum, bottom left to binary with '1' representing a 'dark enough' blue hue, and bottom right as 'primitive' shapes of 'dark enough' blue.

4.2. A greedy heuristic for optimal image placement

Algorithm 1 presents the pseudocode of AUTOART, a greedy search strategy that seeks to rearrange the montage by swapping or flipping images, such that one of the objective functions in Section 4.3 is maximized or minimized. This strategy is unwilling to temporarily accept a worse value of the objective function in order to achieve a better solution in the longer term, such as is achieved by hill-climbing strategies in, for example, simulated annealing. Therefore, we instead initialize from different starting points, or use different random seeds, to explore multiple local optima in order to find a reasonably good solution. While a greedy strategy may be suboptimal in terms of the number of function evaluations required to achieve a satisfactory result, it is the simplest to implement and test. Moreover, it provides information on the 'hardness' of the problem, i.e. if a simple strategy provides good results, a more complex one may just provide marginal performance increases.

The algorithm starts by creating a random permutation of the vector $[1 \dots 306]$, which is returned as the 18×17 matrix χ . Then, it calculates the objective function (to be defined

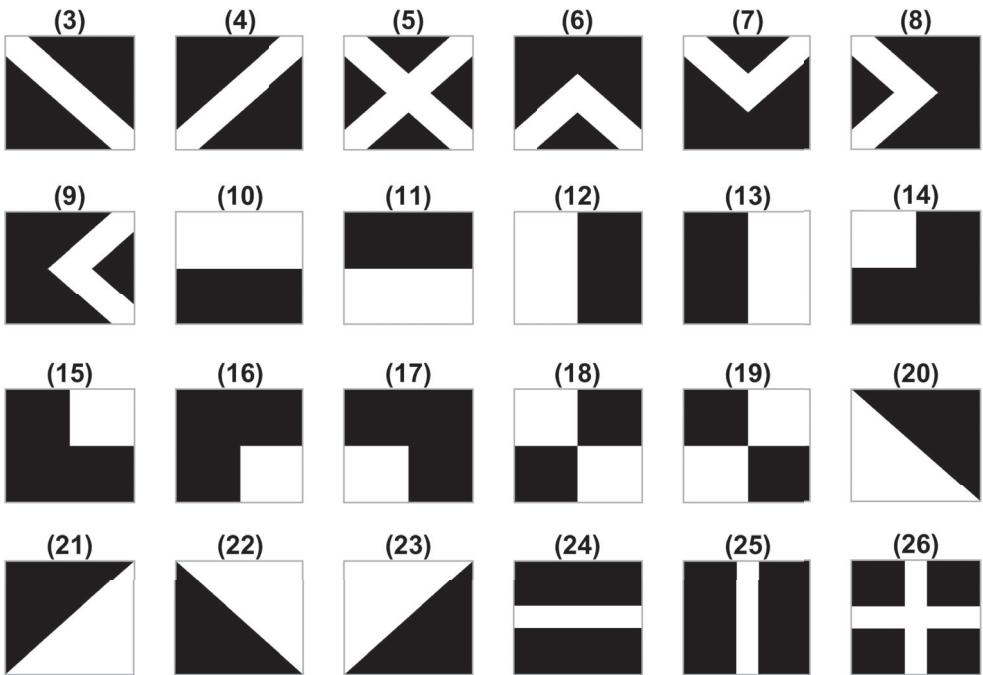


Figure 8. Vocabulary of ‘primitive’ shapes that represent the most salient ‘dark enough’ blue feature from an individual image. Besides the 24 illustrated above, primitives (1) and (2) correspond to black or white squares, i.e. fully ‘0’ or fully ‘1’, respectively.

in the next section), measuring the extent of its order or disorder of blue river connectivity. The algorithm runs for T iterations. In each iteration, it randomly selects a ‘mutation’ operator from the seven available, i.e.

- (1) a global swap of any two randomly selected images located at $\chi_{k,l}$ and $\chi_{k',l'}$;
- (2) a mirror operation in the horizontal axis of a randomly selected image at $\chi_{k,l}$;
- (3) a mirror operation in the vertical axis of a randomly selected image at $\chi_{k,l}$;
- (4) a local swap of a randomly selected image at $\chi_{k,l}$ with its left neighbour $\chi_{k,l-1}$;
- (5) a local swap of a randomly selected image at $\chi_{k,l}$ with its right neighbour $\chi_{k,l+1}$;
- (6) a local swap of a randomly selected image at $\chi_{k,l}$ with its top neighbour $\chi_{k-1,l}$; and
- (7) a local swap of a randomly selected image at $\chi_{k,l}$ with its bottom neighbour $\chi_{k+1,l}$.

While each of these types of mutation is equally likely, there is a bias towards local neighbourhood swaps (types (4)–(7)) which are four times more likely compared to the global swap type (1). It is also worth noting that the algorithm is given additional opportunities to find an optimal solution through mutation types (2) and (3) that the human did not consider, so it should have every opportunity to achieve the goal. After the mutation, the objective function is re-evaluated. If the objective has improved when considering whether the goal is to maximize or minimize the objective function, the montage is updated; if it has not improved, the proposed mutation is discarded. If a mutation is not possible, i.e. a

```

Input: The number of iterations,  $T$ , the random seed,  $\rho$ , a function handle to the objective function,
        ObjectiveFcn, a vector of function handles to the mutation operators, MutationOperator,
        and a variable determining whether to minimize or maximize the cost,  $\phi_{\min}$ .
Output: An optimized montage,  $\chi_{\text{opt}}$ .
1 Function GreedySearch( $T, \chi, \text{ObjectiveFcn}, \text{MutationOperator}, \text{OptimizationFcn}$ ) is
2    $\chi_{\text{opt}} \leftarrow \chi;$ 
3    $J_{\text{opt}} \leftarrow \text{ObjectiveFcn}(\chi);$ 
4   for  $t = 1$  to  $T$  do
5      $i = \text{RandomInteger}(1, 7);$  // Select the 'mutation' operator
6      $\chi_{\text{tmp}} \leftarrow \text{MutationOperator}[i](\chi_{\text{opt}});$ 
7      $J_{\text{tmp}} \leftarrow \text{ObjectiveFcn}(\chi_{\text{tmp}});$ 
8     if  $\text{OptimizationFcn}(J_{\text{opt}}, J_{\text{tmp}}) = J_{\text{opt}}$  then
9        $\chi_{\text{opt}} \leftarrow \chi_{\text{tmp}};$ 
10       $J_{\text{opt}} \leftarrow J_{\text{tmp}};$ 
11    end
12  end
13  return  $\chi_{\text{opt}};$ 
14 end
15 Function AUTOART( $T, \rho, \text{ObjectiveFcn}, \text{MutationOperator}, \phi_{\min}$ ) is
16    $\text{SetRandomSeed}(\rho);$ 
17   // Use a random permutation to initialize the positions of the 306
      images.
18    $\chi \leftarrow \text{RandomPermutation}([1 \dots 306]);$ 
19   if  $\phi_{\min}$  then  $\text{OptimizationFcn} \leftarrow @\text{Min}$  else  $\text{OptimizationFcn} \leftarrow @\text{Max};$ 
20   return GreedySearch( $T, \chi, \text{ObjectiveFcn}, \text{MutationOperator}, \text{OptimizationFcn}$ );
20 end

```

Algorithm 1: Pseudocode for optimal montage arrangement by a greedy algorithm utilizing the seven neighbourhood swap operators defined in Section 4.2, and the three types of objective functions defined in Section 4.3.

left neighbour swap when the selected image is on the left margin, then the mutation is also discarded.

4.3. Objective functions

As discussed above, our objective of constructing a montage can be thought of as an image mosaicing task where multiple, albeit overlapping, images are aligned into a large composition representing a part of a scene (Ghosh & Kaabouch, 2016). Mosaicing involves four steps, i.e. registration, reprojection, stitching and blending. For our purpose, registration is the only step of interest because there is no exploitable overlap in our images, and our aim requires us to only connect (or destroy) blue regions on the joining edges rather than to seamlessly join images with similarity across the entire length of each edge. According to Ghosh and Kaabouch (2016), spatial registration techniques can be classified into ‘area’ and ‘feature’ based. In the next sections we will describe two ‘area’ based approaches, i.e. mutual information and connected pixels, and one ‘feature’ based approach, i.e. identification of primitive patterns. Experimentation with other state-of-the-art ‘feature’ based approaches, such as Harris, FAST or SURF detectors, yielded poor results due to the lack of overlap.

4.3.1. Mutual information

Our first approach is to use mutual information to extract the relationship between the pixel location and its change in intensity. Our assumption is that neighbouring pixels with small changes in intensity will have similar colour. Therefore, by maximizing the relationship

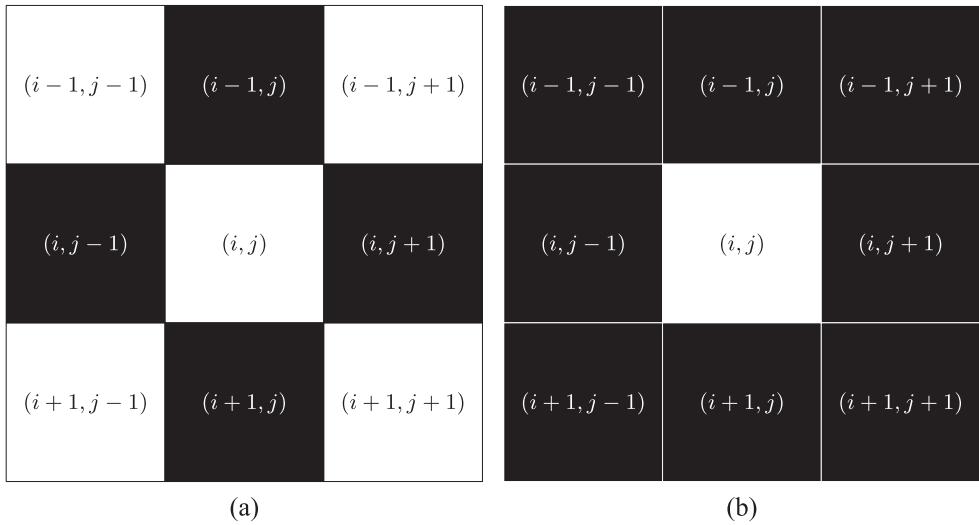


Figure 9. Neighborhood definitions used for the Mutual Information and Connected Pixels functions: (a) a 4-connected and (b) an 8-connected neighbourhood. Pixels marked as black are considered neighbours and used in the calculation.

between the position and the intensity change, as captured by the mutual information, we force the creation of large blocks of similar colours, with the expectation that blue areas emerge. On the other hand, by minimizing this relationship, we force the destruction of such blocks. Mutual information is defined in terms of entropy:

$$J_{MI}(\mathbb{X}) = \widehat{H}(R, C) + \widehat{H}(\Delta) - \widehat{H}(R, C, \Delta)$$

where R and C correspond to the row and column indices of the pixel, and Δ corresponds to the second difference in intensity level, with the difference defined over a 4-connected neighbourhood as illustrated in Figure 9(a) and intensity defined as the 16-bit indexed colour spectrum described in Section 4.1. More formally:

$$\Delta_{i,j} = \frac{1}{4} (\mathbb{X}_{i-1,j} + \mathbb{X}_{i+1,j} + \mathbb{X}_{i,j-1} + \mathbb{X}_{i,j+1}) - \mathbb{X}_{i,j}$$

To estimate Δ , we made use of MATLAB's discrete Laplacian function `DEL2`, whereas the entropy was calculated using the estimator based on k-d trees by Stowell and Plumbley (2009), which partitions the space Ω defined by the variables R , C or Δ into a set of non-overlapping cells ω , each one containing $P(\omega)$ pixels within its hyper-volume $V(\omega)$. The entropy is then defined as:

$$\widehat{H}(\cdot) = \sum_{\omega=1}^m \frac{P(\Omega)}{P(\omega)} \log \left(\frac{P(\Omega)}{P(\omega)} V(\omega) \right)$$

where $P(\Omega) = 9360 \times 10030$ or the total number of pixels. Because this objective function operates over the indexed spectrum and the complete montage \mathbb{X} , it is expected to be the most computationally expensive.

4.3.2. Connected pixels

Our second approach seeks to maximize the area of the largest set of ‘connected’ pixels in the complete montage \mathbb{X} , using the binary representation of each image to locate only the dark blue regions as described in Section 4.1. This representation focuses the algorithm on the dark blue regions while ignoring all other colours, and when we ask it to maximize (or minimize) the total area of connected dark blue regions we expect it to find swaps that will result in the creation (or destruction) of large contiguous regions of dark blue. For this purpose, we adopt the 8-connected definition, that is, two pixels are connected if both of them are equal to ‘1’ and their edges or corners touch, as illustrated in Figure 9(b). That is, for a given region α , a pixel $\mathbb{X}_{i,j}$ belongs to the region if any of its 8-connected neighbours also belongs to α . More formally:

$$\mathbb{X}_{i,j} \in \alpha \quad \left| \bigcup_{m,n=\{-1,0,1\}, m \neq n \neq 0} \mathbb{X}_{i+m,j+n} \in \alpha \right.$$

With the area of α being its cardinality, i.e. $A(\alpha) = |\alpha|$. That is, $J_{CP}(\mathbb{X}) = \max_i A(\alpha_i)$. To calculate this objective function, we utilize the MATLAB functions `BWAREAFILT`, which finds the largest connected area in the binary image, and `REGIONPROPS` which returns the area. Because this objective function is over the binary representation and the complete montage \mathbb{X} , it is expected to be the second most computationally expensive.

4.3.3. Pattern identification

Our third approach makes use of the ‘primitive’ representation described in Section 4.1. For this purpose, we define 24 pattern windows, each one being a 2×2 matrix of selected primitives forming a desirable configuration. Figure 10 illustrate these patterns. For example, the patterns $\{B, C, L, U\}$ were formed by primitives:

$$B = \begin{bmatrix} 1 & 4 \\ 5 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 4 \\ 7 & 1 \end{bmatrix} \quad L = \begin{bmatrix} 3 & 4 \\ 4 & 3 \end{bmatrix} \quad U = \begin{bmatrix} 25 & 1 \\ 26 & 24 \end{bmatrix} \quad (4)$$

where the number indicates the index in the shape vocabulary defined in Figure 8. We then calculate the probability that a section of the montage is similar to one of our patterns, as follows:

$$\Phi_{k,l} = \sum_{i=\{0,1\}} \sum_{j=\{0,1\}} \phi_{k+i,l+j} \mathbb{I}(\tilde{\chi}_{k+i,l+j} = p_{i+1,j+1})$$

where $\phi_{k,l}$ corresponds to the probability that the given image looks like the selected primitive, calculated through Equation (3) using the binary representation of each image, and where $\tilde{\chi}_{k,l}$ is an element of the 18×17 matrix $\tilde{\chi}$, and $p_{i,j}$ is an element of one of the 2×2 pattern matrices, with the selected pattern being the one that maximizes $\Phi_{k,l}$. Then the objective function is defined as:

$$J_{PI}(\tilde{\chi}) = \sum_{k=1}^{17} \sum_{l=1}^{16} \Phi_{k,l} \quad (5)$$

In other words, the objective function is the sum of the probabilities that overlapping windows in the montage are similar to user-defined patterns of ‘blue river’ connectivity. Therefore, by maximizing the probability, we are constructing larger connected areas,

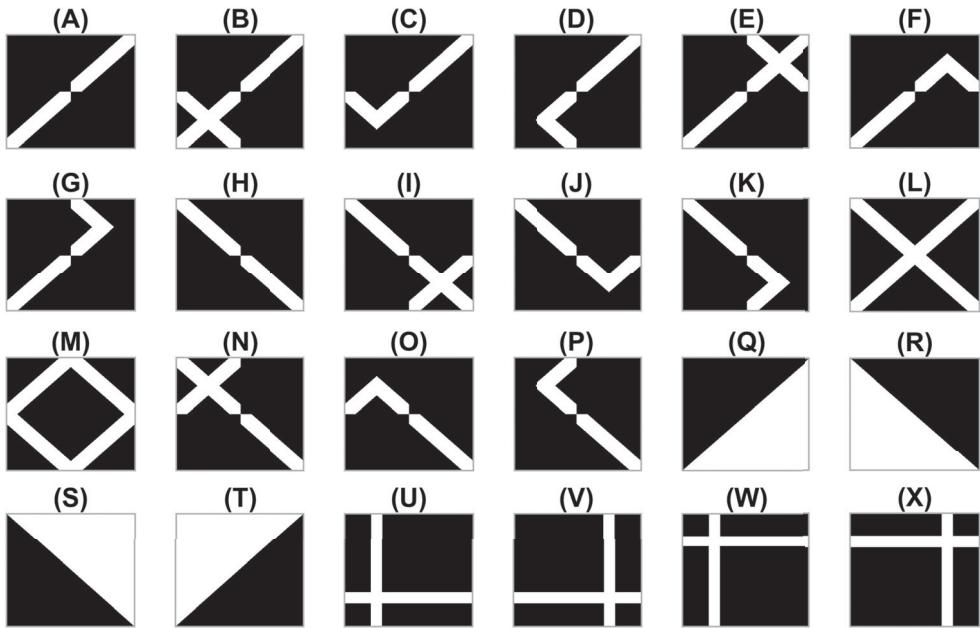


Figure 10. Vocabulary of ‘primitive’ pattern windows. Four of them, i.e. $\{B, C, L, U\}$, correspond to the matrices in Equation (4).

whereas by minimizing the probability we destroy those sections that could appear as connected. This approach provides information about the desired macrostructure sought by the human eye, which is not available through the other two methods. Moreover, because this objective function operates over the 18×17 matrix $\tilde{\chi}$, it is expected to be the least computationally expensive.

4.4. Results

We carried out 20 runs of AUTOART for each of the objective functions described in Section 4.3, ten of them to maximize and the other ten to minimize. In both cases, the algorithm had a budget of $T = 10^4$ iterations. All iterations had as starting point the centre montage shown in Figure 5, although with different random seeds. As a baseline, we also generated 10^6 random permutations of the 306 images. Table 2 shows the average time in seconds taken by one run of AUTOART depending on the objective function, which confirms our expectations on the computational expense of each function. For example, J_{CP} is roughly between 30 and 50 times less expensive than J_{MI} , and J_{PI} is roughly between 15 to 25 times less expensive than J_{CP} , and between 860 to 870 times less expensive than J_{MI} . The higher consistency between J_{MI} and J_{PI} can be explained by the fact that both act over a structure of the same size, i.e. either \mathbb{X} or $\tilde{\chi}$.

Figure 11 illustrates our results, with the top panel showing the results for mutual information (MI) as the chosen objective function, the middle panel for connected pixels (CP), and the bottom panel for pattern identification (PI). The histogram represents the random permutations, the lines represent each one of the panels comprising *Negentropy Triptych*,

Table 2. Average time in seconds taken by one run of autoart depending on the objective function and whether it is maximization or minimization.

	Minimize			Maximize		
	J_{MI}	J_{CP}	J_{PI}	J_{MI}	J_{CP}	J_{PI}
Mean	330378.7	6891.1	380.1	324186.8	9399.0	376.8
Standard Deviation	17233.9	286.4	8.4	22113.0	698.5	12.0

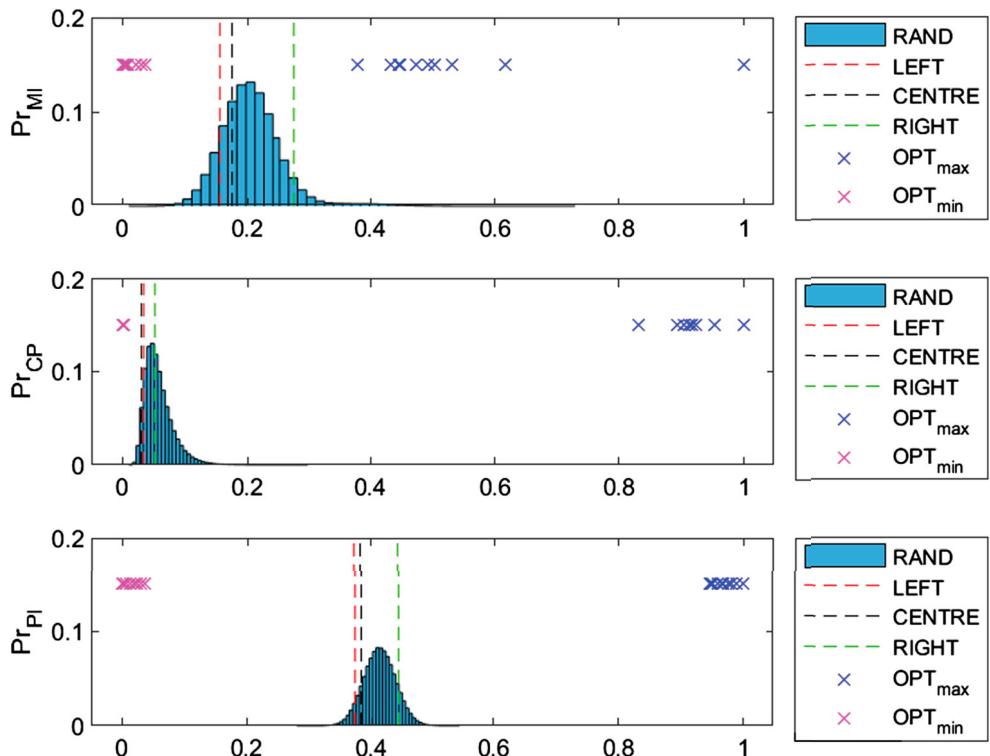


Figure 11. Results from 10 maximization runs from autoart (\times), 10 minimization runs (\times) compared with 10^6 randomly generated montages (histogram) and the left (red—), centre (black—) and right (green—) panels comprising Negentropy Triptych. The cost function values on the x-axis have been scaled between [0, 1] for visualization purposes. Each panel represents a different objective function to measure ‘blue river’ connectivity.

while the \times marks represent the 20 iterations of AUTOART. The x-axis represent the cost value, scaled between the [0, 1] range. The results from each objective function have been scaled between the absolute maximum and minimum obtained during our runs.

Our results confirm that the MI and PI objective functions certainly describe a spectrum of order to disorder, as defined by those objective functions. That is, the objective functions increase when we compare the order of the left panel to the right in *Negentropy Triptych*. Although the CP function does not fulfil completely the requirement, as the centre panel has a lower value than the left panel, the difference is small. Moreover, our results confirm that AUTOART pushes the boundaries of what is obtainable through simple random search, with the differences being significant for the PI function.

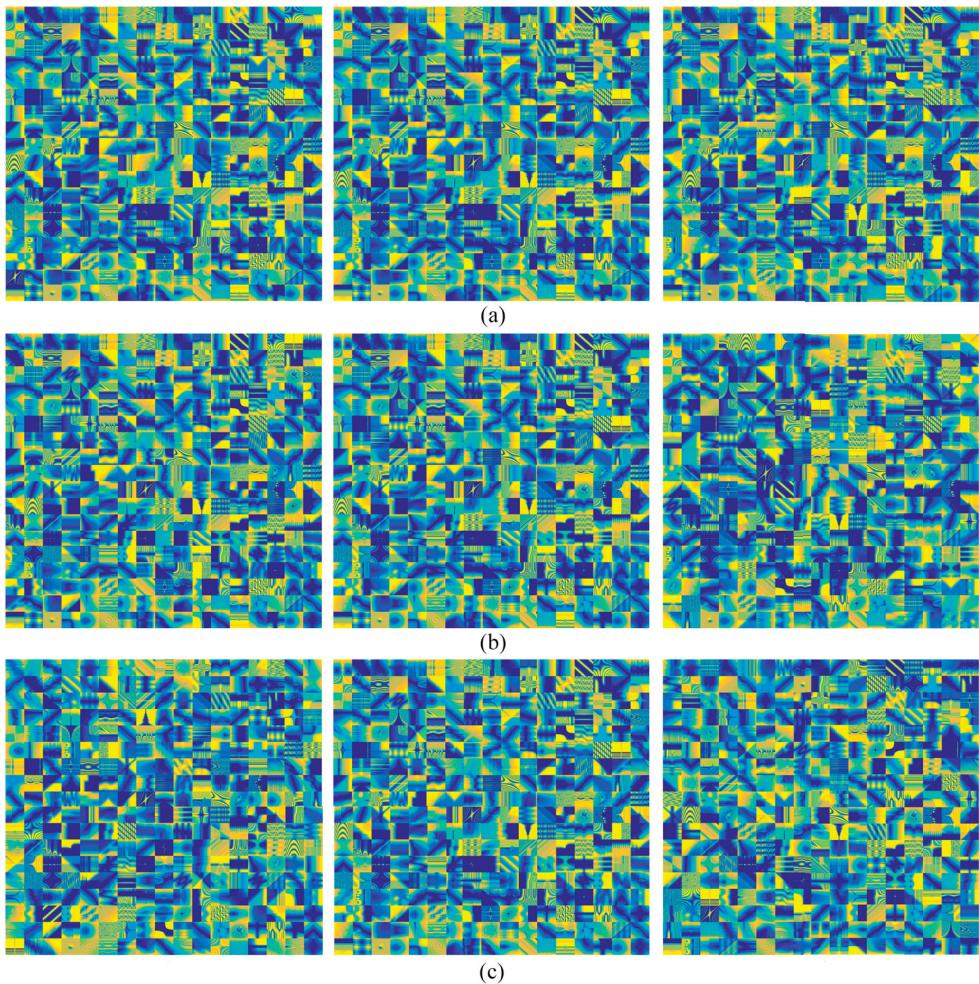


Figure 12. Three automated versions to improve the intent of *Negentropy Triptych*, preserving the centre frame, but replacing the left and right frames with the minimum and maximum obtained through `autoart` for (a) MI, (b) and (c) PI objective functions.

Figure 12 shows three versions of *Negentropy Triptych* montages obtained by preserving the centre frame, but replacing the left and right panels with the minimum and maximum obtained through AUTOART for each one of our objective functions. Each one has a varied level of success of enforcing order, with the PI function in Figure 12(c) showing the desired tendency to form long diagonal lines of connected images much like the human artistic choices, whereas the MI and CP functions tend to cluster images without forming a clear pattern. The ability of PI to form such diagonal structures is supported by the fact that many diagonal patterns were included, as illustrated in Figure 10. We expect that if some of these patterns were excluded, PI would not have been as successful.

The results in enforcing disorder are harder to quantify, as the algorithm's tendency is to push the blue blocks towards a corner. For the PI function, there is less visible microstructure remaining, as it has been effectively destroyed by minimizing the probability of

patterns emerging. Although these computational efforts can be considered successful, the ‘aesthetic’ of each version is harder to evaluate.

5. Discussion and conclusions

The results presented in the previous section clearly indicate that it is easier to ask an algorithm to destroy relationships between adjacent images than it is to ask it to create background structure, particularly when it is so difficult to mathematically define the kind of blue river connectivity that the human eye can identify so easily. While the algorithm is capable of more rapidly exploring the massive search space of potential swaps that could improve the objective function, it cannot compete with human pattern searching ability, which can quickly identify which images would be prime candidates to swap. The superior processing speed of the algorithm does not lead to improved results for two main reasons: (a) the greedy nature of the algorithm means that when it eventually converges it will only have found a locally optimal solution, and will need multiple restarts in order to increase the chance of finding a globally optimal solution; (b) the algorithm has been given objective functions that cannot completely capture the concept of blue river connectivity, despite our best efforts to focus its attention on blue regions of the montage, and various concepts of how connectivity can be measured.

There is no doubt that more sophisticated optimization algorithms could be deployed in order to escape from local minima, such as simulated annealing, or global search methods like genetic algorithms. Without a more compelling objective function, however, this is likely to be quite futile, requiring a considerable computational effort given the observed low probability of effective random swaps. Improving the computational complexity of the cost functions, by updating only the changed regions, could open the door to more sophisticated optimization algorithms with a small memory trade-off. There are numerous other ideas that could be explored, borrowing from image processing methods for image stitching and image registration (Razlighi et al., 2013), but these methods are more focussed on identifying similar images, which is a different goal compared to our efforts to create a very specific background structure. Another interesting direction would be to approach the challenge from a different perspective: instead of asking how we can enhance or destroy ‘blue river’ connectivity from a random arrangement, we could construct a random walk of ‘blue rivers’ as the right image, and then use the minimization of mutual information to destroy the global structure. This latter idea is akin to the concept of image scrambling used in image encryption (Arumugam & Annadurai, 2021; Zhang et al., 2007) and digital watermarking (Soliman et al., 2016), and could open avenues for new applications of our work in cryptography. We have made available the collection of 306 images at Muñoz and Smith-Miles (2020b) and hope that other researchers may take on the challenge of devising more human-like objective functions, and pursuing other such opportunities.

There are many interesting questions raised by this research, including how the choice of colour scale affects aesthetic taste, and whether the desire for global background structure depends on the chosen colour palette? Our images are available for download and may add value to aesthetics literature and psychology studies. In particular, neuro-aesthetics research may benefit from studying visual cortex stimulation and eye tracking of subjects when presented with some of our individual complex images, as well as the various montages, to explore the relationship between attention, aesthetic taste, complexity and



preference for order or disorder. We also propose in this paper that the scale of the order at which people seek unity in variety to make sense of complexity, whether it be at a micro or macro scale, may also be a factor worthy of consideration. A natural extension of this work, as a companion study, would be to conduct a rigorous psychometric evaluation of a larger group of human subjects to confirm the untested hypothesis that motivated our artistic goal described in this paper. Indeed, a formal psychometric study of human subjects combining their aesthetic preference with the five-factor model personality scores (Goldberg, 1992) and additional demographic data would be a rich source of data to confirm the hypothesis, and add value to the literature on computational aesthetics and the role of personality. Perhaps the most interesting question highlighted by this paper though is one with more significant societal consequences: given the challenges of programming an algorithm to make decisions like a human, can we devise an algorithm to learn to replicate an artist's decision-making process? Machine learning methods could certainly be deployed, as they have for other aesthetic prediction studies (Carballal, Fernandez-Lozano, Heras et al., 2019), to train a model to recognize good 'blue river' connectivity, either across the whole montage or in small regions, and then this model could form the basis of an objective function to evaluate the effectiveness of any proposed swaps. This is an extension of the template pattern matching approach explored in this paper. However, we leave the pursuit of an artificial 'artistic' intelligence to future work.

Acknowledgments

The authors are grateful to their colleagues who participated in the informal survey of their aesthetic preferences. Conceptualization, K.S-M; methodology, M.A.M, K.S-M; software, M.A.M; writing – original draft preparation, K.S-M.; writing – review and editing, M.A.M; visualization, M.A.M, K.S-M; project administration, K.S-M; funding acquisition, K.S-M. All authors have read and agreed to the published version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the Australian Research Council, under the Australian Laureate Fellowship scheme [grant number FL140100012].

ORCID

Kate Smith-Miles <http://orcid.org/0000-0003-2718-7680>

Mario Andrés Muñoz <http://orcid.org/0000-0002-7254-2808>

References

- Arnheim, R. (Ed.). (1966). *Toward a psychology of art: Collected essays*. University of California Press.
- Arumugam, S., & Annadurai, K. (2021). An efficient machine learning based image encryption scheme for medical image security. *Journal of Medical Imaging and Health Informatics*, 11(6), 1533–1540. <https://doi.org/10.1166/jmihi.2021.3470>
- Audet, C., & Hare, W. (2017). *Derivative-free and blackbox optimization*. Springer International Publishing.

- Berlyne, D. E. (1970, September). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, 8(5), 279–286. <https://doi.org/10.3758/BF03212593>
- Birkhoff, G. D. (1933). *Aesthetic measure*. Harvard University Press.
- Carballal, A., Fernandez-Lozano, C., Heras, J., & Romero, J. (2019, February). Transfer learning features for predicting aesthetics through a novel hybrid machine learning method. *Neural Computing and Applications*, 32(10), 5889–5900. <https://doi.org/10.1007/s00521-019-04065-4>
- Carballal, A., Fernandez-Lozano, C., Rodriguez-Fernandez, N., Castro, L., & Santos, A. (2019, January). Avoiding the inherent limitations in datasets used for measuring aesthetics when using a machine learning approach. *Complexity*, 2019, 1–12. <https://doi.org/10.1155/2019/4659809>
- Chamorro-Premuzic, T., Burke, C., Hsu, A., & Swami, V. (2010, November). Personality predictors of artistic preferences as a function of the emotional valence and perceived complexity of paintings. *Psychology of Aesthetics, Creativity, and the Arts*, 4(4), 196–204. <https://doi.org/10.1037/a0019211>
- Cupchik, G. C. (1986, December). A decade after Berlyne: New directions in experimental aesthetics. *Poetics*, 15(4–6), 345–369. [https://doi.org/10.1016/0304-422X\(86\)90003-3](https://doi.org/10.1016/0304-422X(86)90003-3)
- Geert, E. V., & Wagemans, J. (2020, May). Order, complexity, and aesthetic appreciation. *Psychology of Aesthetics, Creativity, and the Arts*, 14(2), 135–154. <https://doi.org/10.1037/aca0000224>
- Ghosh, D., & Kaabouch, N. (2016, January). A survey on image mosaicing techniques. *Journal of Visual Communication and Image Representation*, 34, 1–11. <https://doi.org/10.1016/j.jvcir.2015.10.014>
- Goldberg, L. R. (1992). The development of markers for the big-five factor structure. *Psychological Assessment*, 4(1), 26–42. <https://doi.org/10.1037/1040-3590.4.1.26>
- Hansen, N., Auger, A., Finck, S., & Ros, R. (2014, December). *Real-parameter black-box optimization benchmarking BBOB-2010: Experimental setup* (Tech. Rep. No. RR-7215). INRIA.
- Inglis, M., & Aberdein, A. (2014, July). Beauty is not simplicity: An analysis of mathematicians' proof appraisals. *Philosophia Mathematica*, 23(1), 87–109. <https://doi.org/10.1093/philmat/nku014>
- Kerschke, P., & Preuß, M. (2019). Exploratory landscape analysis. *GECCO '19: Proceedings of the genetic and evolutionary computation conference companion*.
- Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004, November). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95(4), 489–508. <https://doi.org/10.1348/0007126042369811>
- Marín, J. (2011, October). How landscape ruggedness influences the performance of real-coded algorithms: A comparative study. *Soft Computing*, 16(4), 683–698. <https://doi.org/10.1007/s00500-011-0781-5>
- Markovic, S. (2010). Aesthetic experience and the emotional content of paintings. *Psihologija*, 43(1), 47–64. <https://doi.org/10.2298/PSI1001047M>
- Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C., & Rudolph, G. (2011). Exploratory landscape analysis. In N. Krasnogor (Ed.), *Proceedings of the 13th annual conference on genetic and evolutionary computation – GECCO '11* (pp. 829–836). ACM Press.
- Mersmann, O., Preuss, M., Trautmann, H., Bischl, B., & Weihs, C. (2015, March). Analyzing the BBOB results by means of benchmarking concepts. *Evolutionary Computation*, 23(1), 161–185. https://doi.org/10.1162/EVCO_a_00134
- Muñoz, M. A., & Smith-Miles, K. (2020a, September). Generating new space-filling test instances for continuous black-box optimization. *Evolutionary Computation*, 28(3), 379–404. https://doi.org/10.1162/evco_a_00262
- Muñoz, M. A., & Smith-Miles, K. (2020b). Negentropy triptych image collection. https://matilda.unimelb.edu.au/matilda/problems/opt/black_box.
- Muñoz, M. A., & Smith-Miles, K. A. (2017, December). Performance analysis of continuous black-box optimization algorithms via footprints in instance space. *Evolutionary Computation*, 25(4), 529–554. https://doi.org/10.1162/evco_a_00194
- Muñoz, M. A., Villanova, L., Baatar, D., & Smith-Miles, K. (2018, December). Instance spaces for machine learning classification. *Machine Learning*, 107(1), 109–147. <https://doi.org/10.1007/s10994-017-5629-5>
- Nadal, M., Munar, E., Marty, G., & Cela-Conde, C. J. (2010, July). Visual complexity and beauty appreciation: Explaining the divergence of results. *Empirical Studies of the Arts*, 28(2), 173–191. <https://doi.org/10.2190/EM.28.2.d>



- Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013, January). Visual aesthetics and human preference. *Annual Review of Psychology*, 64(1), 77–107. <https://doi.org/10.1146/psych.2013.64.issue-1>
- Pandey, A., & Pati, U. C. (2019, September). Image mosaicing: A deeper insight. *Image and Vision Computing*, 89, 236–257. <https://doi.org/10.1016/j.imavis.2019.07.002>
- Rawlings, D., & Furnham, A. (2000, November). Personality and aesthetic preference in Spain and England: Two studies relating sensation seeking and openness to experience to liking for paintings and music. *European Journal of Personality*, 14(6), 553–576. [https://doi.org/10.1002/\(ISSN\)1099-0984](https://doi.org/10.1002/(ISSN)1099-0984)
- Razlighi, Q., Kehtarnavaz, N., & Yousefi, S. (2013, October). Evaluating similarity measures for brain image registration. *Journal of Visual Communication and Image Representation*, 24(7), 977–987. <https://doi.org/10.1016/j.jvcir.2013.06.010>
- Searson, D. P., Leahy, D. E., & Willis, M. J. (2010). GPTIPS: An open source genetic programming toolbox for multigene symbolic regression. In S. I. Ao, O. Castillo, C. Douglas, D. D. Feng, & J.-A. Lee (Eds.), *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 1, pp. 77–80). International Association of Engineers (IAENG).
- Seo, D. I., & Moon, B. R. (2007, June). An information-theoretic analysis on the interactions of variables in combinatorial optimization problems. *Evolutionary Computation*, 15(2), 169–198. <https://doi.org/10.1162/evco.2007.15.2.169>
- Smith-Miles, K., Baatar, D., Wreford, B., & Lewis, R. (2014, May). Towards objective measures of algorithm performance across instance space. *Computers & Operations Research*, 45, 12–24. <https://doi.org/10.1016/j.cor.2013.11.015>
- Soliman, M. M., Hassanien, A. E., & Onsi, H. M. (2016). An adaptive watermarking approach based on weighted quantum particle swarm optimization. *Neural Computing and Applications*, 27(2), 469–481. <https://doi.org/10.1007/s00521-015-1868-1>
- Stein, M. (1987, May). Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2), 143–151. <https://doi.org/10.1080/00401706.1987.10488205>
- Stowell, D., & Plumley, M. (2009, June). Fast multidimensional entropy estimation by k-d partitioning. *IEEE Signal Processing Letters*, 16(6), 537–540. <https://doi.org/10.1109/LSP.2009.2017346>
- Swami, V., & Furnham, A. (2020). The influence of personality on aesthetic preferences. In M. Nadal & O. Vartanian (Eds.), *The oxford handbook of empirical aesthetics*. Oxford University Press (OUP).
- Wang, W., Yang, S., Zhang, W., & Zhang, J. (2019, November). Neural aesthetic image reviewer. *IET Computer Vision*, 13(8), 749–758. <https://doi.org/10.1049/cvi2.v13.8>
- Zhang, H. x., Lv, H., & Weng, X. (2007). Image scrambling degree evaluation method based on information entropy. *Journal of Circuits and Systems*, 12(6), 95–98.

Appendix. Summary of technical specifications of evolved functions from Muñoz and Smith-Miles (2020a)

A.1. Mathematical definition of BBO

In the formal description that follows we will focus on the task of minimizing a function, without loss of generality for maximization, since the latter can be achieved by minimizing the negative of the function. Specifically, BBO involves a set of decision variables $\mathcal{X} \subset \mathbb{R}^D$ defining the input space, an unknown function $F : \mathcal{X} \mapsto \mathcal{Y}$, that generates a value $\mathcal{Y} \subset \mathbb{R}$ defining the output space, and $D \in \mathbb{N}^+$ is the dimensionality of the problem. We seek a solution as a decision vector in \mathcal{X} that gives the minimal value of \mathcal{Y} . A candidate solution $\mathbf{x} \in \mathcal{X}$ is a D -dimensional vector, and $y \in \mathcal{Y}$ is the candidate's objective or cost value. BBO algorithms explore a sequence of (\mathbf{x}, y) input-output pairs in order to find a solution, and differ in their search strategies for determining the (\mathbf{x}, y) pairs they evaluate. The most successful algorithms find the minimal cost solution with a minimal number of function evaluations.

Table A1. Features selected in Muñoz and Smith-Miles (2020a) for characterizing BBO test functions based on ELA metrics based on Surrogate models (Mersmann et al., 2011), Cost distributions (Marín, 2011; Mersmann et al., 2011), Entropic significance (Seo & Moon, 2007).

Feature	Type	Description
f_1	Surrogate model	Adjusted coefficient of determination of a purely quadratic model
f_2	Surrogate model	Ratio between the minimum and the maximum absolute values of the quadratic term coefficients in the purely quadratic model
f_3	Surrogate model	Mean cross-validation accuracy of a Classification and Regression Tree (CART)
f_4	Surrogate model	Ratio of Mean cross-validation accuracies of a CART and quadratic discriminant analysis model
f_5	Cost distribution	Entropy of the cost distribution
f_6	Cost distribution	Skewness of the cost distribution
f_7	Cost distribution	Number of peaks of the cost distribution
f_8	Entropic significance	Significance of first order

A.2. BBO benchmark test functions

In order to test BBO algorithms, benchmark test functions are created, whose function $F : \mathcal{X} \mapsto \mathcal{Y}$ is defined analytically and enables y to be rapidly calculated for any x , but allow the algorithm only knowledge of the (x, y) pairs rather than the analytical form of F . One of the most well-studied test suites is the Comparing Continuous Optimizers (COCO) noiseless benchmark test functions (Hansen et al., 2014). The COCO functions are composed of 24 basis functions defined within $\mathcal{X} = [-5, 5]^D$, and are divided into five categories with various qualitative characteristics (Mersmann et al., 2015): separable functions; low or moderately conditioned functions; unimodal functions with high conditioning; multimodal functions with adequate global structure; and multimodal functions with weak global structure. The benchmark test functions are generated by scaling and transforming these 24 basis functions via linear translations, rotations and symmetry breaking that cause perturbations and translational shifts in the input and output spaces to create some limited variations.

A.3. Characterising BBO landscapes using features from exploratory landscape analysis

ELA metrics require a sample of (x, y) pairs of sufficient size in order to estimate the characteristics of the landscape based only on the sample observations. Typically, a Latin-hypercube sample (Stein, 1987) is generated, and the number of samples is often related to the dimensionality D of the problem. Popular ELA metrics include those based on autocorrelation, estimates of the size of the basins of attraction for low-cost solutions, metrics based on goodness of fit of surrogate models of various functional forms such as linear or quadratic models, statistical properties of the cost function distribution, and information theoretic measures such as entropy and mutual information. We refer the interested reader to Kerschke and Preuß (2019) for a recent review of ELA methods.

For each of the 24 basis functions in the COCO test suite, we generated 30 instances with variations in rotations, translations, etc. for a fixed dimensionality, with D chosen successively from the set of 9 values $[2, 3, 5, 8, 10, 20, 40, 60, 100]$. This process created 6480 test functions from which to generate an instance space. Each test function F_i for $i = 1 \dots 6480$ is summarized by a feature vector of ELA metrics, $\mathbf{f}(i)$ calculated using a Latin-hypercube sample of size $D \times 10^3$. The chosen 8 ELA features, defined in Muñoz and Smith-Miles (2020a) and summarized in Table A1, span a range of information theoretical and statistical distribution measures.

A.4. Instance space analysis of BBO test functions

An instance space is created by finding an optimal projection of the instances from their initial representation as a feature vector, in this case in 8-dimensions, to a point in a 2D plane. This dimension



reduction is achieved via the linear transformation $\mathbf{z} = \mathbf{Af}$ provided by Equation (A1), with full details of the 8D feature vector \mathbf{f} , and the method used to find the linear transformation matrix \mathbf{A} , provided in Muñoz et al. (2018). For the purposes of this paper, it suffices to note that each instance is a test function that is now represented as a point in the 2D plane (z_1, z_2) defined by the axes in Equation (A1), which are linear combinations of the 8 ELA features summarizing an instance, as defined in Table A1. The location of the 6480 test functions within the instance space can be seen as the grey dots in the background of Figure 2, with each test function F_i given by a (z_1, z_2) coordinate in the instance space, once its 8D feature vector $\mathbf{f}(i)$ is calculated and projected to (z_1, z_2) using Equation (A1).

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} -0.28454 & 0.61038 \\ 0.20991 & 0.16267 \\ 0.02979 & 0.53365 \\ 0.53101 & 0.01772 \\ -0.31959 & -0.02927 \\ -0.30095 & -0.43889 \\ -0.42920 & -0.22560 \\ -0.46389 & 0.26739 \end{bmatrix}^T \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \end{bmatrix} \quad (A1)$$

A.5. Evolving new test functions using genetic programming

The GP routine described in Muñoz and Smith-Miles (2020a) was run 10 times from different initial random populations of binary trees. We used a population size of 400 trees, with a maximum tree depth of 10, and ran for 100 generations. Lexicographic tournament selection was employed with ten individuals per tournament, while the best 10% of the population was kept as an elite set. The probability of a mutation, crossover and direct transfer were set to 30%, 60% and 10% respectively. We set as stopping criteria a fitness value of 10^{-3} . Constants were generated from a uniform distribution in the $[-100, 100]$ range.