Breaking the Mould
An Evolutionary Quest for Innovation Through Style Change

João Correia, Penousal Machado, Juan Romero, Pedro Martins, and Amílcar Cardoso

Abstract An autonomous creative system able to learn, create and innovate is presented. Following previous work on the same topic, the approach explores the interplay between a classifier and an evolutionary system. The classifier is trained with famous paintings and images created by the system, learning to distinguish between these two categories. It is then used to assign fitness, leading to the discovery of imagery that deviates from the one previously created by the system. Additionally, by taking phenotype similarity into account, we further promote the discovery of diverse images during the course of the evolutionary runs. The images created throughout the evolutionary runs are added to the training set and the process is repeated. This iterative process, which includes retraining the classifier, sets the system into a permanent quest for novelty and innovation. The experimental results obtained across several iterations are presented and analysed, showing the ability of the system to consistently produce novel imagery and to identify atypical images.

1 Introduction

As posited by McCormack (2007), the development of Aesthetic Judgement Systems (AJSs) is one of the biggest challenges in the field of Computational Creativity. Over the course of the years, two main approaches emerged: the development

João Correia
CISUC, University of Coimbra, Portugal email: jncor@dei.uc.pt
Penousal Machado
CISUC, University of Coimbra, Portugal email: machado@dei.uc.pt
Juan Romero
Faculty of Computer Science, University of A Coruña, Spain email: jj@udc.es
Pedro Martins
CISUC, University of Coimbra, Portugal email: pjmm@dei.uc.pt
Amílcar Cardoso
CISUC, University of Coimbra, Portugal email: amilcar@dei.uc.pt

of hardwired fitness measures that try to encapsulate some sort of aesthetic principle, and the use of Machine Learning (ML) techniques to learn aesthetic models (Romero, Machado, Carballal, & Correia, 2012).

As we stated in previous works (Machado & Cardoso, 1997; Romero, Machado, Santos, & Cardoso, 2003; Machado, Romero, & Manaris, 2007; Machado, Romero, Santos, Cardoso, & Pazos, 2007; Romero, Machado, Carballal, & Correia, 2012), our long term goal is the development of Artificial Artists (AAs) that display the full range of abilities of human artists. In this context, the ability to learn aesthetic models is indispensable, since it gives the system the ability to experience, assess and react, not only to its own artistic production, but also to the artworks of other, artificial or human, artists (Machado & Cardoso, 1997). Furthermore, it also creates the preconditions that allow the system to be inspired by other artists, to detect trends, and to deliberately innovate and deviate.

The ability to consistently generate innovative and adequate artifacts is a key trait of creative, human or computational, agents. In this Chapter, we present an AA that is characterised by the ability to build its own aesthetic model from a set of examples, and by its permanent quest for novelty and innovation through style variation and change.

The approach resorts to an expression-based evolutionary art engine and adaptive classifiers, in this case Artificial Neural Networks (ANNs). The ANNs are trained to discriminate among the artistic production of the system and that of famous artists. The evolutionary engine is used to generate images that the ANNs do not recognise as being products of the system, and that, as such, are novel in relation to its previous artistic practice. During the evolutionary runs, we also promote the discovery of a diverse set of imagery by taking phenotype similarity into account when assigning fitness.

When a set of evolutionary runs is concluded, the novel imagery they produced is added to the training set, enlarging the area of the search space covered by the system, and the ANNs are retrained. This leads to a refinement of the classifiers, which, in turn, forces the evolutionary algorithm to explore new paths and styles to break with its past. Thus, the consecutive discovery of new styles is attained through the revision and refinement of the aesthetic criteria for novelty of the AA, while variation within style is attained by promoting phenotype diversity.

The research presented in this Chapter builds upon our previous efforts on the same topic (e.g., (Machado, Romero, & Manaris, 2007; Machado, Romero, Santos, et al., 2007; Correia, Machado, Romero, & Carballal, 2013b; Romero, Machado, Carballal, & Correia, 2012)) expanding previous approaches by:

- Performing in each framework iteration a set of parallel evolutionary runs instead of a single one;
- Considering phenotype similarity to promote the discovery of a wide set of diverse images in the course of each evolutionary run;
- Using classifiers with access to a larger number of image features;
- Using a significantly larger set of training examples;
- Using an archive to summarise the innovative imagery produced by the system.

Although we consider that framework presented herein could also be applied to other domains, the scope of this Chapter is limited to the image domain.

The experimental results obtained across several iterations are presented and analysed, showing the ability of the system to consistently produce novel imagery and to identify atypical images without human intervention. We consider that the results obtained in the course of the first iteration are evocative of images produced by user-guided evolution. Furthermore, we claim that the images evolved in the last of the presented iterations are of a significantly different nature, breaking the mould with the previous artistic production of the AA. As such, we hypothesise that a limited form of h- and t-creativity (Boden, 2004) may have been attained.

The Chapter is structured as follows: we begin by making an overview of the state of the art in this field identifying and summarising the works that were more pertinent to the research presented in this Chapter; in Section 3 we make an overview of the EFECTIVE framework, presenting the details of its instantiation in Section 4; this is followed by the presentation and analysis of the experimental results; finally we draw conclusions and indicate future research.

2 State of the Art

The seminal work of Karl Sims (1991) led to the emergence of a new art form, evolutionary art, which is characterised by the use of evolutionary computation to evolve populations of artworks. In Sims' work, users assign fitness to the images, indicating their favorite ones and, by these means, steering evolution towards regions of the space that match their criteria. This process is, in many ways, similar to selective breeding, a practice that humans have been following for centuries in the context of animal and plant breeding, to develop, enhance, exaggerate and create particular phenotype traits. This approach to fitness assignment became known as interactive evolutionary computation (IEC). While IEC has many merits and applications, systems based on IEC are dependent on human users. Therefore, although several Computer Aided Creativity systems have been developed based on IEC (Machado, Romero, Santos, et al., 2007), this approach is not viable for the development of Artificial Artists.

In what concerns automation of fitness assignment, the central question is how to develop a scheme that strongly correlates to human aesthetics or, at least, some aspects of it. One of the most popular approaches to tackle this problem is the use of hardwired fitness functions. There are several notable examples of systems in this category (Machado & Cardoso, 2002; G. Greenfield, 2002, 2003; Machado, Dias, & Cardoso, 2002; G. Greenfield, 2005; Ross, Ralph, & Hai, 2006; Neufeld, Ross, & Ralph, 2007; Machado, Correia, & Assunção, 2015) and also recent works comparing the merits of such aesthetic measures (Ekárt, Joó, Sharma, & Chalakov, 2012; den Heijer & Eiben, 2010; Atkins, Klapaukh, Browne, & Zhang, 2010; Romero, Machado, Carballal, & Santos, 2012).

The use of ML techniques for fitness assignment purposes has also been explored. In their seminal work, Baluja et al. (1994) used an ANN trained with a set of images generated by user-guided evolution to assign fitness. Machado et al. (2007; 2007) study the development of AAs able to perform style changes over the course of several runs. In a related work, Li et al. (2012) investigate aesthetic features to model human preferences. The aesthetic model is built by learning both phenotype and genotype features, which are extracted from internal evolutionary images and external real world paintings. Kowaliw et al. (2009) compared biomorphs generated randomly, through interactive evolution, and through automatic evolution using a classifier system inspired by content-based image retrieval metrics. The experimental results indicate that the results of the automatic system were comparable to those obtained by interactive evolution. The use of co-evolutionary approaches (Saunders, 2001; G. R. Greenfield, 2002) and hybrid approaches that combine interactive evolution with hardwired fitness functions (Machado, Romero, Cardoso, & Santos, 2005) have also been explored.

Another important contribution of Sims' work concerns the representation. Sims uses canonical Genetic Programming (GP) (Koza, 1992) to evolve images. The genotypes are symbolic expressions, which assume the form of a tree, composed of functions (internal nodes) and terminals (leafs), which may be variables or constants. The phenotypes, i.e. the images, are produced by calculating the outcome of the symbolic expression over a range of variable values. In other words, the outcome of $expression(x,y)$ yields the color values of the (x,y) pixel of the image. To produce the entire image, one iterates over the desired range of x and y values, with a given step.

One of the questions that naturally arises when considering a representation is its expressive power. In this case, what types of images can be represented by means of a symbolic expression of this kind. Although the answer depends, obviously, on the function and terminal set being used, Machado and Cardoso (2002) demonstrated that it is possible to represent any given image using a simple function and terminal set. Provided that the function set contains the `if-then-else` function and that the terminal set contains variables x, y and $constants$, it is trivial to design a symbolic expression to any image, and hence any image is representable. In simple terms, the argument is the following: using `if-then-else`, one can successively partition the image into smaller areas, eventually reaching pixel level size; then one only needs to use a constant to define the desired pixel color. Notice that the existence of `if-then-else` is not a strict requirement, as long as there is a way to partition and combine different regions of the image, the idea still holds. Furthermore, many other types of proof are viable. For instance, if the system has the ability to encode, explicitly or implicitly, an iterated function system, then one can rely on Barnsley's (1993) proof to demonstrate that all images are representable.

From the above, it is safe to say that most expression-based evolutionary art systems are able to represent any given image. Thus, in theory, it is possible to recreate by evolutionary means any artwork that was ever made or that will be made (McCormack, 2007). Practice, however, is an entirely different matter. The images produced by expression-based evolutionary art tend to be abstract and have a mathematical appearance. As pointed out by Machado et al. (e.g. (2007)), each evolutionary art system tends to have its own signature, which is deeply related to the function set and to the genetic operators being used.

Romero et al. (2003) suggested combining a general purpose evolutionary art system with an image classifier trained to recognise faces, or other types of objects, to evolve images of human faces. In recent years, this idea has been put to practice by several researchers to evolve several kinds of figurative images such as faces, flowers, leafs, breasts, and font glyphs (Machado, Correia, & Romero, 2012a, 2012b; Correia, Machado, Romero, & Carballal, 2013a; Nguyen, Yosinski, & Clune, 2015; Martins, Correia, Costa, & Machado, 2015), as well as ambiguous images (Machado, Vinhas, Correia, & Ekárt, 2015). This kind of approach extended the realm of imagery produced by expression-based evolutionary art systems, by assigning fitness based on the resemblance to objects that are usually not present in the kind of images these systems tend to produce.

Another approach that has the potential to expand the realm of generated imagery is novelty search. It is important to notice that the use of techniques to promote the novelty of the solutions, predates the coining of the term novelty search algorithm, by Lehman and Stanley (2008). The works of Saunders et al. (2001) and Machado et al. (2007) are examples of early approaches, where novelty plays an important role in evolution, while in the work of Kowaliw et al. (2009), evolution is guided by novelty alone. Among the examples that strictly follow a novelty search mechanism as proposed by Lehman and Stanley, we highlight the works of Secretan et al. (2011) and Liapis et al. (2013).

Our biggest criticism to canonical novelty search is that we consider that novelty, alone, is not a sufficient criterion for creativity. Furthermore as it was analysed in the previous Chapter (McCormack, 2017), EC approaches have demonstrated success to better locate regions of high creative rewards. As such, the work presented in this chapter focuses on three central issues: (i) The automation of fitness assignment; (ii) The development of a system that innovates, overcoming the implicit bias of its representation and expanding the frontiers of its artistic production; (iii) The generation of artworks that relate to human aesthetics.

3 The Framework

The architecture proposed by Romero et al. (2003) argues that an AA should be composed of two main modules: a creator and a critic. The work presented in this chapter follows, roughly, this architecture, with the role of the creator being played by an evolutionary computation engine and the role of the critic being played by an ANN. We employ the Evolutionary FramEwork for Classifier assessmenT and ImproVEment (EFECTIVE). In abstract, EFECTIVE is a framework that assesses and improves classifier performance through the synthesis of new training instances. In our scenario, it fits the role of AA, who, based on the its judgment and its past

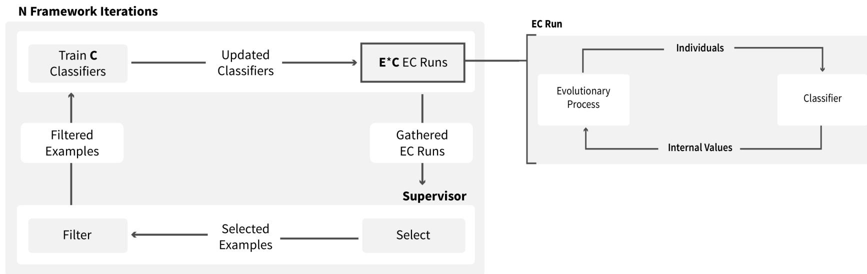

Fig. 1 Overview of the EFECTIVE Framework.

experience, iteratively learns from its inspiration and produced work, evolving its craft along its existence.

EFECTIVE is composed of three main modules: Evolutionary Computation (EC) engine; Classifier System (CS) and Supervisor; each with distinct roles. In brief, the EC engine is the one responsible for the evolutionary part, where the examples are synthesised and evolved. As for the CS, it constitutes the learning approach. The Supervisor module manages the examples that result from the interaction between the EC engine and classifier system, selecting and filtering synthesised examples that will be used to improve the training dataset. These modules come together to create an iterative process for improvement of classifiers. Fig. 1 presents an overview of the framework.

Before diving into the details of the instantiation of the framework, a succinct description follows:

1. A set of *external* images is selected; In the case of this chapter, this set is composed of famous artworks, representing a source of inspiration for the AA;
2. A set of *internal* images is selected; In this case, the EC engine is used to randomly create a set of images, thus creating a sample of the type of imagery the evolutionary engine tends to produce;
3. The ANN is trained to distinguish among *internal* and *external* images;
4. A new set of evolutionary runs is started; The output of the ANN is used to assign fitness; Images classified as *external* have higher fitness than those classified as *internal*; Additionally, phenotype diversity is also taken into consideration;
5. During the course of each evolutionary run, an archiving module keeps track of the artistic production of the AA, storing images that are classified as external and diverse from other images classified as external evolved during the course of the run;
6. When the set of evolutionary runs is concluded, the Supervisor module gathers and merges the archives resulting from each evolutionary run;
7. The consolidated archive is added to the set of *internal* images;
8. The process is repeated from step 3.

One of the key aspects of this approach is the definition of two classes of images. The first class contains *external imagery*. Images that were not created by the GP system and that are usually considered “interesting” or of “high aesthetic value”. Conceptually, the external set should be seen as an “inspiration” for the AA. It provides a stable attractor that is meant to ensure that the evolved imagery tends to incorporate aesthetic qualities recognized by humans. The second class contains *internal imagery*, it is composed of images generated by the evolutionary engine, and describes the previous artistic production of the AA. For the purposes of the present work, this class represents undesirable imagery, since we are interested in innovation through style change.

In the present case, the task of the evolutionary module is to evolve images that the ANN classifies as external. This may be accomplished by evolving images that are:

1. Similar to those belonging to the *external* set;
2. Different from the set of internal images (e.g., images that are entirely novel, hence dissimilar from both sets).

Note that in this context, the concept of similarity and dissimilarity is deeply connected to the features serve as input to the ANN. Therefore, it may deviate from human perception.

The approach relies on promoting a competition between the evolutionary engine and the CS. In each iteration, the evolutionary engine must evolve images that are misclassified by the CS, otherwise no progress is achieved. By assigning fitness using a classifier and valuing examples that belong to a predefined class, the approach evolves several misclassified examples (Machado et al., 2012b). These examples can be potentially useful for improving the performance of the CS.

The systematic expansion of the internal set, and the subsequent retraining of the ANN, causes an “arms-race” between generator and classifier. As such, from iteration to iteration, the evolutionary engine is forced to explore new paths, which results in stylistic change and in the expansion of the diversity of the artistic production of the system.

As pointed out by Machado et al. (2007), in the long run, there are two possible final scenarios, which correspond to natural termination criteria for the approach: (i) the evolutionary engine becomes unable to find images that are classified as external; (ii) the ANN becomes unable to discriminate between internal and external imagery. The first outcome reveals a weakness of the evolutionary engine, which can be caused by a wide variety of factors (deceptive fitness landscape, incorrect parametrisation, lack of computational resources, etc.). In the second outcome, there are two possible sub-scenarios: (ii.a) the images created by the EC system are similar to some of the external images, which implies that the EC and the CS are performing flawlessly; (ii.b) the images created by the EC system are stylistically different from the external imagery, which indicates a flaw of the CS.

In the next Section we present several details pertaining the instantiation of the framework to the scenario discussed in this Chapter. Applications of the same framework in other, non-artistic, domains can be found in (Machado et al., 2012b, 2012a; Correia et al., 2013a; Machado, Vinhas, et al., 2015).

4 Instantiation of the EFECTIVE Framework

The EFECTIVE framework is instantiated for this scenario with one classifier system, an evolutionary engine and a supervisor. It starts by training a classifier with an initial dataset. Then E parallel evolutionary runs are started. When all evolutionary runs are finished, the supervisor gathers the individuals and decides which ones are going to be added to the dataset. This cycle is iteratively repeated until a termination criterion is met. The global parameters of the framework are presented in Table 1.

Table 1 Global parameters of the framework.

Parameter	Setting
Classifiers per iteration (C)	1
EC runs per classifier (E)	30
Adequacy threshold	0.5
Dissimilarity threshold	0.01

4.1 Classifier System

The CS is composed of a Feature Extraction module and an ANN. The classifier participation in the approach is crucial for several reasons: it evaluates the images that are generated by the evolutionary engine; its performance dictates the number of examples that are added and/or deleted before retraining the classifier.

In this work, the CS is trained under certain conditions before it is used to assign fitness during the evolutionary runs. On each training phase the ANN is trained with the full dataset. If training is entirely successful, meaning that the ANN is able to fully discriminate between the internal and external sets, we proceed to the evolutionary runs. However, if false externals exist, i.e. if there are human produced artworks being classified as evolved images, these images are removed from the external dataset and a new training attempt is made. Training is only concluded when no external images are classified as being internal.

The removal of these images has two motivations. First, from the perspective of the classifier, one can consider that the style these images embody has already been explored. As such, they should no longer be classified as external. Second, from a more pragmatic perspective, these images tend to be atypical in relation to the rest of the images of the external dataset, removing them arguably simplifies the task of the classifier, which may, in turn, result in classifiers that provide fitness landscapes that are more favourable for the evolutionary engine.

The existence of false externals, i.e. evolved images classified as human made, does not have a direct solution. Deleting them would solve nothing. Instead, they remain in the internal dataset. Future iterations are likely to explore the same shortcoming of the classifier, increasing the number of examples of the same style present in the internal dataset, and forcing, due to the increased cardinality of the subset, the classifier to learn that such images are internal.

4.1.1 Feature Extraction

In our approach, the ANNs do not have direct access to the images, instead each image is described by a set of image features, which serve as input to the ANN. As such, we developed a Feature Extractor used to extract relevant features from each image.

The pipeline of the feature extractor is the following: the input image is resized to 128×128 pixels; converted to the Hue Saturation Value (HSV) colourspace, and a copy of the each image channel is stored for further computation; several pre-processing operations are computed on demand, depending on the feature to be extracted, i.e. apply Canny filter to the image and extract information from the edges. In total, the feature extraction process yields a total of 120 features, that are later used as input for the classifier.

A throughout description of the feature extractor would be long, and is outside the the scope of this chapter. Therefore, we present a brief description of the features extracted, indicating bibliographic references that may provide to the interested reader a complete description. Most of the features implemented originate from previous work concerning the aesthetic analysis of images (Datta, Joshi, Li, & Wang, 2008). The features collected were inspired by the work of Datta et al. (2006), Li et al.(2009), Faria et al. (2013), Romero et al. (2003), den Heijer (2012) and, based on our previous work Machado et al. (2007; 2007) and Correia et al.(2013b).

To make the description of the feature set tractable, we introduce a taxonomy 2. Some of the features could be classified in several categories, in these cases we followed the literature consensus for the feature's category. When selecting and developing this group of features, our goal was to cover several aspects of the images' style and aesthetics.

Although most of the features are implementations based on the state of the art, we have also introduced some new features in this work. These are briefly described in the following paragraphs.

As the name suggests, the edge density feature captures information regarding the number of edges present in the image. This is achieved by applying a Canny filter to the image and counting the percentage of pixels that correspond to edges, i.e. white pixels.

We also introduce the Palette analysis features, which are intended to provide additional information regarding the image's colour palette. The core idea is to analyse the contrasting colours present in the image (Machado, Correia, & Assunção, 2015). First, we apply a colour quantisation algorithm to reduce the number of colours using k-means clustering. The colour occurrences are counted and sorted in descending order. We compute the distances among the colours of the resulting image using the HSV space, as follows: considering two colour vectors (H,S,V) to represent the colour, the distances in the S and V colour components are calculated

Table 2 The proposed feature taxonomy is composed of five categories: colour, complexity, composition, salience and texture.

Category	Features	Reference
Colour	Palette analysis	
	Average of pixel values	(Romero et al., 2003; Datta et al., 2006; Machado, Romero, Santos, et al., 2007)
	Standard deviation pixel values	(Romero et al., 2003; Datta et al., 2006; Machado, Romero, Santos, et al., 2007)
	Weber contrast	(Faria et al., 2013)
	Michelson contrast	(Faria et al., 2013)
	Warm and cool colors	(Faria et al., 2013)
	Contrasting colors	(C. Li & Chen, 2009; Faria et al., 2013)
	Background simplicity	(Rubner, Tomasi, & Guibas, 2000; Datta et al., 2006)
Complexity	Fractal dimension	(Romero et al., 2003; Machado, Romero, & Manaris, 2007)
	Zipf size	(Romero et al., 2003; Machado, Romero, & Manaris, 2007)
	Zipf rank	(Romero et al., 2003; Machado, Romero, & Manaris, 2007)
	JPEG and fractal compression	(Romero et al., 2003; Machado, Romero, & Manaris, 2007; Correia et al., 2013b)
Composition	Horizontal and vertical symmetry	(den Heijer, 2012)
	Liveliness	(den Heijer, 2012)
	Edge density analysis	
	Lighting	(C. Li & Chen, 2009)
	Hue Count	(Datta et al., 2006; C. Li & Chen, 2009)
	Blur analysis	(C. Li & Chen, 2009)
	Rule of thirds	(Datta et al., 2006)
Salience	Edge distribution	(Datta et al., 2006)
	Spatial frequency	(Faria et al., 2013)
	Subject size	(Faria et al., 2013)
Texture	Tamura contrast	(Tamura, Mori, & Yamawaki, 1978)
	Tamura coarseness	(Tamura et al., 1978)

using the Euclidean norm; for the H channel, which is circular, we use the formula $dist(a,b) = \min\left(|a-b|, |a-MAX-b|, |b-MAX-a|\right)$ to compute the distance, where a and b are two colours and MAX is the maximum value of H. After calculating the distances, we discard the colours that are closer to each other than a predetermined threshold. This results in n colours, which we consider the images palette. With the palette we calculate a frequency histogram and we compute the following metrics: number of palette colours; percentage of occurrences; the mode, minimum value and maximum value for each component of the colour; histograms linear regression and error; average distance to the next colour; average and standard deviation of the differences between the histograms bins; components of the maximum and minimum distance from one colour to the others. We make the same analysis by considering the purity of the colours, which translates to only considering the S and V components of the images channels to compute the metrics, ignoring the H channel.

4.1.2 Artificial Neural Network

The ANN is a feed-forward network, with one hidden layer and two output neurons. It is trained with standard backpropagation. The classifier was built using WEKA's[1] FastNeuralNetwork. WEKA is a workbench for machine learning with a significant number of algorithms and tools available (Hall et al., 2009). The choice of an ANN based classifier is justified by the success of this approach in previous works of related nature (Machado, Romero, & Manaris, 2007; Correia, 2009).

[1] `http://www.cs.waikato.ac.nz/ml/weka/`WEKA 3: Data Mining Software in Java

The ANN receives as input the feature vector. The output indicates its confidence in classifying the input instance as belonging to the either the internal or the external class. To avoid a "binary" output, i.e. both neurons returning either 0 or 1, which would result in an unsuitable fitness landscape, we employ a tolerance threshold during the training stage. This translates on a modification of the training algorithm, where, during the backpropagation of the error, if the difference between the output of the network and the desired output is below the tolerated threshold, then the error is propagated back as zero (no error). The parameters of the ANN are summarised in Table 3.

Table 3 Parameters related to the ANNs and their training.

Parameter	Setting
Initialisation of weights	random, $[-0.1, 0.1]$ interval
Learning function	backpropagation
Tolerance threshold	0.3
Learning rate	0.3
Momentum	0.2
Epochs	1000

4.2 Initial Datasets

The initial sets of external and internal images play an important role in the performance of our system. We use an external set containing 26238 images including works of artists such as: Cézanne, de Chirico, Dalí, Gauguin, Kandinsky, Klee, Klimt, Matisse, Miró, Modigliani, Monet, Picasso, Renoir, van Gogh. The images where gathered from different online sources. The rationale was to collect a wide and varied set of artworks. Although we avoided repetitions, it is relatively common for an artist to paint several versions of the same theme. In these cases, and in order to avoid the subjectivity of deciding what was sufficiently different, we decided to include the different variations.

The set of internal images is created using the evolutionary engine, described in the next subsection, to generate 30 initial random populations of size 1600. These images are added to the internal dataset until the same amount of examples exist in the two datasets. Although the images were created randomly, some of the phenotypes may appear more than once. Figure 2 presents samples of the images belonging to the internal dataset, illustrating the type of imagery that the EC engine produced in these circumstances.

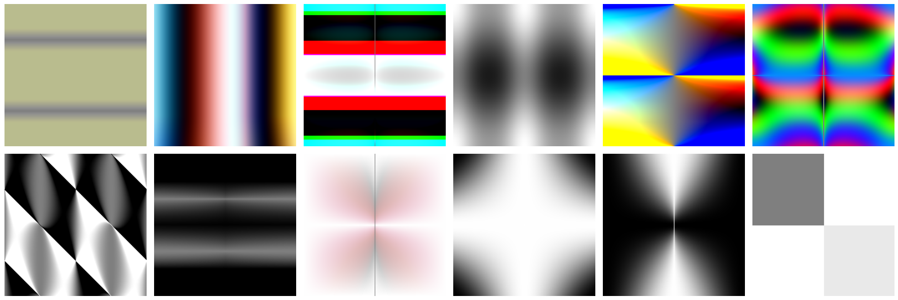

Fig. 2 Samples of the internal dataset.

4.3 Evolutionary Engine

For this work we used the ge**N**eral purp**O**se exp**R**ession **B**ased **E**volutionary a**R**t **T**ool (norBErT) as the EC engine (Vinhas, 2015). Inspired by the work of Sims (1991) and Machado and Cardoso (2002), it is a general purpose, expression-based, GP image generation engine that allows the evolution of populations of images. The genotype uses a tree representation to encode individuals and create images from those trees, using a rendering process which consists in generating an output value for each image pixel. Thus, the genotypes are trees composed of a lexicon of functions and terminals. The functions include mathematical and logical operations; the terminal dataset is composed of two variables, x and y, and random constant values and vectors. The phenotypes are images, rendered by evaluating the expression-trees for different values of x and y, which serve both as terminal values and image coordinates. In other words, to determine the value of the pixel in the (0,0) coordinates, one assigns zero to x and y and evaluates the expression-tree. In this instantiation the fitness of the individuals is given by the output of the CS, more precisely, by the ANN's output as described in subsection 4.1.2.

As mentioned, we employ a phenotype diversity mechanism by using a novelty search algorithm, designed to evolve a diverse set of adequate images. The main goal of this algorithm is to generate a broader set of images than the set that would be created by a traditional fitness based EA. In essence, it is a method capable of evolving images according to two criteria that are chosen automatically by analysing the quality of the images produced in each generation. One criterion is to look for the best images according to a fitness function and the other consists in taking novelty and fitness as two different objectives to be maximised. The reason why novelty is not considered alone is because prior tests have shown how big is the search space and, consequently, how difficult is to discover suitable images (Vinhas, 2015). Similar behaviour has occurred when using a single criterion or considering both fitness and novelty (Vinhas, 2015).

The algorithm's flowchart is similar to the traditional EA one, differing only in two main aspects: (i) the creation of an archive to store the most novel solutions

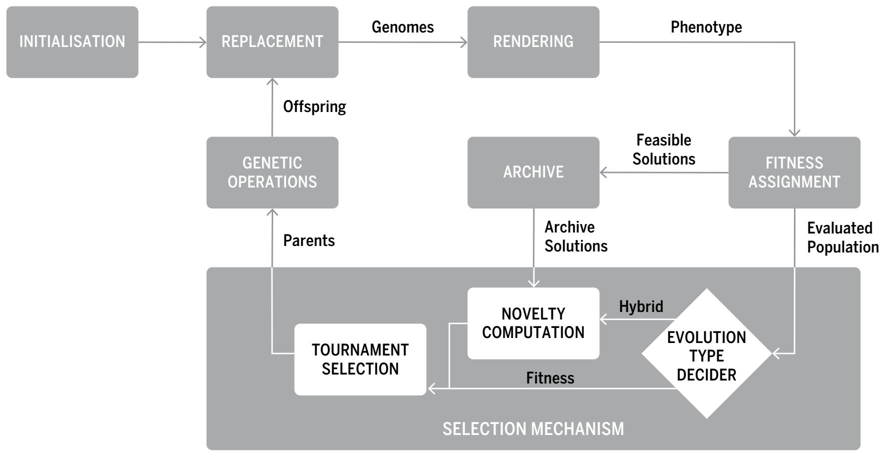

Fig. 3 Flow of the proposed hybrid algorithm.

and, (ii) a customised selection mechanism, which is able to consider single or multiple objectives using a tournament based strategy. The algorithm's flow is shown in Figure 3, and can be summarised as follows:

1. Randomly initialise the population;
2. Render the images (phenotypes) from the individuals' genotypes;
3. Apply the fitness function to the individuals;
4. Select the individuals that meet the criteria to be in the archive (archive assessment);
5. Select the individuals to be used in the breeding process. The individuals are picked using one of the following criteria: (i) according to their fitness, as a standard EA; (ii) taking into account both the fitness and the novelty metric, which is computed using the archive members;
6. Employ genetic operators to create the new generation of solutions, that will replace the old one;
7. Repeat the process starting from step 2, until a stop criterion is met.

4.3.1 Archive Assessment

In this work, the archive has an unlimited size and it plays an important role, because it is used to evaluate our solution and prevents the algorithm from exploring areas of the search space already visited. The idea is that the archive should represent the spectrum of images found to date, and for this reason, the bigger the archive is, the more the algorithm is able to generate suitable and diverse images. Whereas in the previously mentioned works the archive size is limited, we opted for not restricting it.

At this stage, a candidate individual has its fitness assigned and it has to meet two requirements in order to be added to the archive: (i) its fitness must be greater or equal than an adequacy threshold f_{min}; (ii) it needs to be different from those that already belong to the archive. This process is performed by computing the average dissimilarity between the candidate and a set of k-nearest neighbours. When the average dissimilarity is above a predefined dissimilarity threshold, $dissim_{min}$, the individual is added to the archive. The values for f_{min} and $dissim_{min}$ are presented in Table 4.1.

The dissimilarity metric for an image i is computed as:

$$\text{dissim}(i) = \frac{1}{\text{max}_{arch}} \sum_{j=1}^{\text{max}_{arch}} d(i,j), \tag{1}$$

where max_{arch} is a predefined parameter which represents the number of most similar images to consider when comparing with image i, and $d(i,j)$ is a distance metric that measures how different two images (i and j) are. From this dissimilarity measure there are two exceptions that should be highlighted. If there are no entries in the archive, the first individual that has a fitness above f_{min} is added. Moreover, if the number of archive entries is below max_{arch}, Equation (1) is used with the number of archive entries instead of max_{arch}.

For archive assessment, we resorted to an image similarity metric. Similarity metrics provide us with a notion of distance between pair of images. The development of image distance metrics is an relevant and rich area of research with several applications. A revision of the state of the art is beyond the scope of this Chapter. To the interested reader, we suggest the consulting the works of Wang et al.(2005) and Goshtasby et al. (2012). Images distance metrics typically involve pixel based operations that can be less or more elaborated. Among the available state of the art options, we chose to employ the Normalized Cross Correlation (NCC), that can be calculated, for two images X and Y with a m by n size, in the following way:

$$\text{NCC}(X,Y) = \frac{\sum_{i=1}^{m\times n} X_i Y_i}{\sqrt{\sum_{i=1}^{m\times n} X_i^2 \sum_{i=1}^{m\times n} Y_i^2}}, \tag{2}$$

where X_i and Y_i correspond to the pixels of images X and Y, respectively.

NCC similarity outputs a value in the interval $[0,1]$, where 1 indicates the best match. This measure, besides providing a fast calculation, is deemed more robust than most metrics for noisy scenes (Nakhmani & Tannenbaum, 2013). It suits our needs, in the sense that our approach involves a considerable quantity of images, and it can minimise the impact of noisy images on our dissimilarity assessment. As such, we use as distance metric $d(i,j) = 1 - NCC(i,j)$.

4.3.2 Selection Mechanism

The selection mechanism is important to shape how evolution will proceed, depending on the results obtained in a given generation. Our novelty approach has a customised selection mechanism that can switch between a fitness-based strategy and a hybrid mechanism that considers both fitness and novelty. It starts as a fitness guided evolution; however, that can change according to a decision rule, which is described as:

$$\begin{cases} change\_to\_fitness, & adequate_{inds} < T_{min} \\ change\_to\_hybrid, & adequate_{inds} > T_{max}, \end{cases}$$

where $adequate_{inds}$ is the number of individuals of the current generation that have a fitness above the threshold f_{min}; T_{min} is the threshold used to verify if evolution should be changed to fitness, and T_{max} is used to verify if it should be changed to hybrid.

In fitness guided evolution, the tournament selection is based on the fitness values of the candidate solutions, as in a standard EA. If hybrid evolution is chosen, it is necessary to compute the novelty of each selected individual, and perform a Pareto-based tournament selection, using the novelty and fitness of each selected individual as two different objectives to maximise.

The novelty computation process is inspired by Lehman and Stanley's work (2008), with one small change: the k most similar images are considered from the set of the selected individuals and the archive, instead of considering the whole population and the archive. An example of this novelty computation is illustrated in Figure 4: considering $k = 4$ and a tournament size of 5, the dashed lines denote the chosen individuals to compute novelty, and it is possible to see that from the 4 nearest individuals picked, 3 were chosen from the tournament while the remaining one was chosen from the archive.

At this stage, each selected individual has a fitness and novelty value, and there is the need to determine the winner of the tournament. This process is inspired by multi-objective EAs, namely the Pareto-based approaches, which select the best individuals based on their dominance or non-dominance when compared to other individuals. In this work, the hybrid tournament selection determines the non-dominant solutions by comparing, among the selected individuals, on the basis of both fitness and novelty. After computing the set of non-dominant individuals, we have the so-called Pareto front. The tournament winner will be selected by randomly retrieving one of the solutions of the Pareto front.

The settings of the GP engine and the archive assessment for each EC run are presented in Table 4.

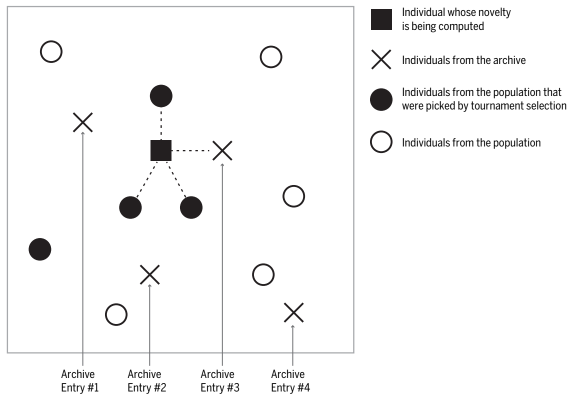

Fig. 4 Novelty computation for an individual.

Table 4 Parameters of the GP engine.

Parameter	Setting
Population size	100
Number of generations	50
Crossover probability	0.8 (per individual)
Mutation probability	0.05 (per node)
Mutation operators	sub-tree swap, sub-tree replacement, node insertion, node deletion and mutation
Initialization method	ramped half-and-half
Initial maximum depth	5
Mutation max tree depth	3
Archive assessment width	32 px
Archive assessment height	32 px
T_{min}	5
T_{max}	15
Function set	$+, -, \times, /$, min, max, abs, neg, warp, sign, sqrt, pow, mdist, sin, cos, if
Terminal set	x, y, random constants

5 The Experimental Results

In this section we present the experimental results obtained using our approach. As previously stated, one of the key characteristics of our approach is its iterative nature. In each iteration we perform 30 evolutionary runs, and once these runs end, the “external” images produced by the system, i.e. the images that expand the range of the artistic production of the system, are added to the internal set and the ANN retrained, promoting the discovery of novel images in subsequent iterations.

Table 5 Statistics regarding evolutionary process across iterations. The results pertain 30 independent evolutionary runs for each framework iteration.

Framework iteration	Evolved external images	Images added	Seeds with ev. external	Avg. generations for ev. external
1	38283	30110	30	3.63
2	1426	250	22	18.22
3	816	195	17	20.52
4	752	178	24	25.33
5	366	57	10	29.3
6	1105	433	21	20.24
7	620	131	22	31.64
8	191	31	7	23.86
9	422	115	21	24.90
10	692	62	20	28.5
11	374	126	22	32.27
12	267	101	11	31.09
13	842	352	17	21.76

We are, therefore, primarily interested in analysing the differences, in terms of produced imagery, that occur from iteration to iteration. It is impossible to show all the images produced in the course of the evolutionary runs. Even if we only presented the images classified as external, this would imply presenting 38283 images for the first iteration alone. As such, we will present a synthesis of the results, which aims to convey the key experimental findings. We divide our analysis into subsections as follows: first, we present and examine the results concerning the evolution of fitness throughout iterations; next, we will inspect the images produced; finally, we analyse the classifier's training and performance in each iteration.

Although we present results concerning 13 iterations of the framework, it is important to stress that further iterations are still being performed. Therefore, the process is not concluded, and all evidence indicates that a significantly higher number of iterations would be necessary before a *breakdown* of the EC engine or classifier takes place.

5.1 Analysis of the Numeric Results Concerning Evolution

Table 5 depicts a series of statistics concerning the evolutionary process across iterations, namely: the total number of images evolved in the course of the 30 evolutionary runs of each iteration that were classified as external (*Evolved external*); the number of these that was added to the internal set used to train the classifier guiding the next iteration after supervision (*Images added*); the number of seeds in which the EC engine was able to find at least one image classified as external (*Seeds with ev. external*); the average number of generations necessary for finding an image classified as external (*Avg. generations for ev. external*). This average is calculated taking only into account the seeds where at least one image classified as external was found.

As it can be observed, a striking number of images classified as external was found in the course of the first iteration, 38283, which corresponds to an average of 1276.1 per evolutionary run. All of the evolutionary runs were able to find "external" images and, on average, they took 3.63 generations to find the first image classified as external.

Although this number is somewhat surprising, it is far from being unexplainable. In essence, this result means that it is easy for the system to break from his "past" and produce novel imagery.

The initial set of internal images was created by randomly generating genotypes and their corresponding phenotypes. As such, the images of the initial internal dataset did not undergo evolution. By supplying an aesthetic model, and a mechanism that steers evolution towards regions of the search space that where not covered by the initial dataset, we are fundamentally changing the nature of the images that the system tends to produce. When confronted by images that are novel, and that probably do not fit in either of the categories (internal or external), the classifier is forced to make a choice, eventually classifying some of these novel images as external. Once such image is found, evolution quickly explores and exploits such type of imagery, leading to the discovery of a high number of images classified as external.

On a second stage, the phenotype diversity mechanisms kicks in, contributing to the discovery of a diversified set of images classified as external. The importance of the phenotype diversity mechanism can be verified by the fact that out of the 38283 classified as external, 30110 were added to the internal dataset. Thus, only 8173 of the evolved images classified as external, roughly 21%, was considered similar to the ones already in the archive of their corresponding evolutionary runs and, therefore, discarded. This result shows that the phenotype diversity mechanism is able to prevent stagnation of the evolutionary runs and convergence to a fixed type of image.

After the "explosion" of novelty that occurs in the first iteration, the task of the evolutionary engine and, as will be seen, of the classifier, becomes increasingly harder and an abrupt decrease of productivity is verified. In the course of the 30 generations of the second iteration, the EC engine found 1426 images that were classified as external. Although this is still an impressive number, it pales in comparison with the numbers observed in the first iteration. This increase in difficulty can also be observed by the increase on the average number of generations necessary to find an external image 18.22 and by the fact that only 22 out of the 30 evolutionary runs were able to find images classified as external. The chart presented in Figure 5, concerning the evolution of the fitness of the best individual of each generation across iterations, further highlights the differences in the difficulty of the task of the EC engine in the first and second iteration.

Out of the 1426 external images found in the course of the second iteration, 250 were added to the internal dataset, since the remaining 1176 were considered sufficiently similar by our archiving algorithm to these 250. This illustrates a well-known fact concerning novelty search algorithms (as defined by Lehman and Stan-

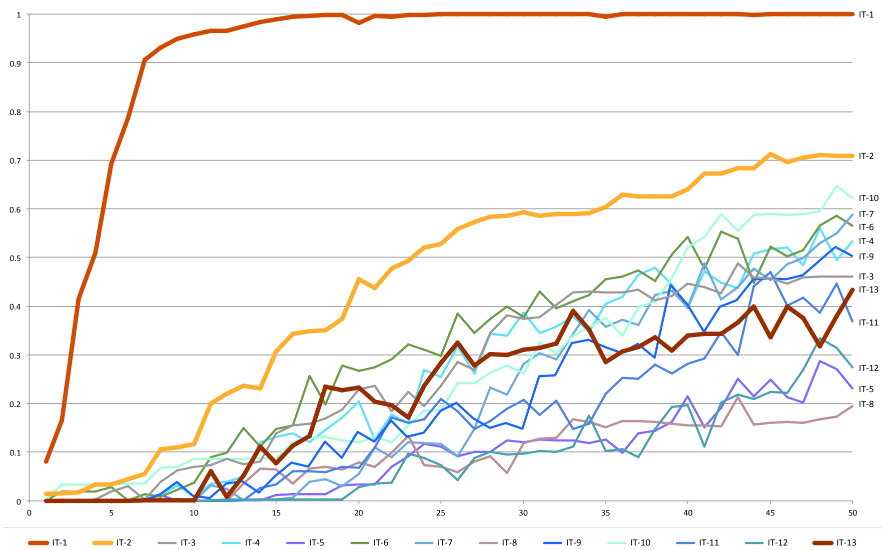

Fig. 5 Evolution of the fitness of the best individual of each generations. Results are averages of 30 independent evolutionary runs for each iteration.

ley (2008)): as optimising fitness becomes harder, it becomes significantly more difficult to find solutions that are both novel and fit. In other words, although the phenotype diversity mechanisms are activated and contribute to the diversity of the population, finding images that are simultaneously novel, in relation to the ones evolved in the course of the evolutionary run, and adequate, i.e. classified as external, becomes increasingly difficult.

As the number of iterations increases, and as the internal dataset becomes larger, one would expect an increasing difficulty in finding images classified as external (and also an increasing difficulty in learning to differentiate between the two sets). Although this tends to be true, it is not always the case. As Figure 5 illustrates, although there is a clear differentiation among the lines representing the evolution of fitness of the first two iterations and the remaining ones, and although these differences are statistically significant, the same does not happen for the remaining iterations. The explanation for this fact is twofold: (i) the number of images added in each iteration is not sufficient to make the task visibly harder; (ii) the training of the classifier includes a stochastic component, and as such, even if trained with the same datasets, different classifiers may induce different fitness landscapes with different difficulties.

Fig. 6 Fittest individual from each population of a typical evolutionary run of the first iteration. The image in the upper-left corner corresponds to population 0; remaining images in standard reading order.

5.2 *Analysis of the Visual Results*

Next we make an analysis of the visual results, i.e. the images, produced in the course of the 13 iterations. The complexity of the setup, and the vast number of images classified as external that were evolved make this analysis particularly hard. Furthermore, and although we will try to be as objective as possible, the analysis entails a degree of subjectivity that cannot, and perhaps, should not, be avoided. We divide this analysis in three subsections, focusing, respectively, on the analysis of the visual results of the first iteration, intermediate iterations, and thirteenth iteration.

5.2.1 First Iteration

We begin by trying to convey what happens within each of the 30 evolutionary runs of the first iteration. For this purpose, Figure 6 depicts the fittest individual from each of the 50 generations of a typical evolutionary run of the first iteration. As it can be observed, the fittest images of the first two generations are quite amorphous. By the third generation, the EC engine finds its the first image classified as external. From this point onwards, the phenotype diversity mechanism kicks in, promoting the discovery of images that are, simultaneously adequate, i.e. classified as external, and different from the ones previously evolved in the course of this specific run. This mechanism does not produce immediate effects in terms of the fittest image of the generations, but it prevents the algorithm from converging, and creates the conditions for the discovery, within the evolutionary run, of different images that are also classified as external. As such, the apparently abrupt changes that can be observed in Figure 6 result, mainly, from a progressive evolutionary process that promotes the diversity of the population.

In Figure 7 we present a sample of the images classified as external evolved in the course of the same run as the one depicted in Figure 6. Since we were unable to find a reasonable algorithm for automatically sampling the set of evolved images in a convincing manner, this and other samples presented in this Chapter were selected by hand, trying, in all cases, to maximise the diversity of the sample and making it as representative as possible. As it can be observed, the diversity of the populations and of the images being classified as external is larger than what Figure 6 suggests, showing the adequacy of the phenotype diversity mechanisms.

Figure 8 depicts the fittest individual of each of the 30 evolutionary runs of the first iteration. All of these images have been classified as external. There are, at least, three predominant traits: most of the images tend to be dark and with low contrast; several exhibit a star-like shape; many of them include some sort of noise. In some cases, the contrast is so low that the images appear to be of uniform color for the human eye; however, a color adjustment and equalization operation will reveal the hidden structure. Regarding this point, it is relevant to point out that several of the features that serve as input to the ANN are invariant regarding contrast among colors, so these results also highlight the differences in the perception of images between humans and ANNs. It also appears to be safe to state that several of the runs

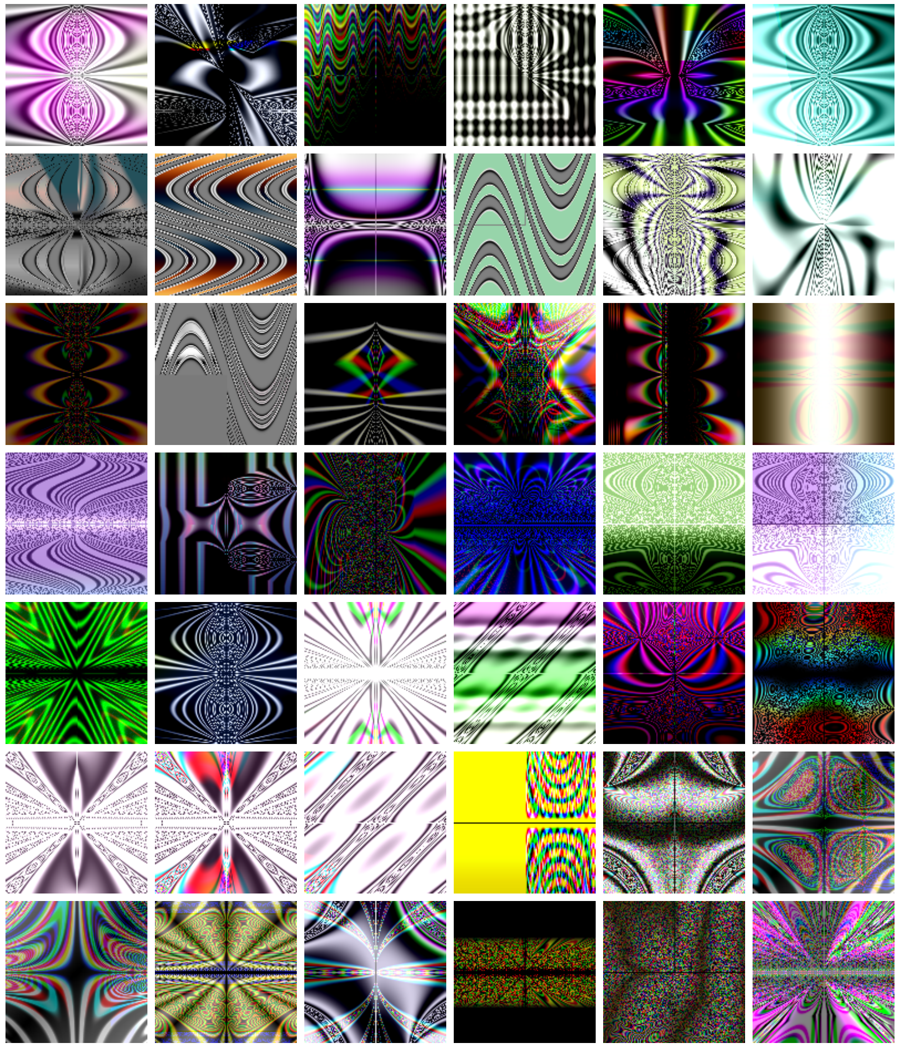

Fig. 7 Samples of the images classified as external, generated throughout the course single typical evolutionary run of the first iteration.

converged to the same type of imagery, which is an expected result. The runs are performed in parallel and the classifier, which ultimately defines the fitness landscape is common to all. Therefore, the fitness landscape has the same local and global optimum, an optima with a larger basis of attraction are bound to be explored more often. Additionally, each evolutionary run has its own archive and no access to the archives of others, therefore the phenotype diversity mechanisms cannot avoid

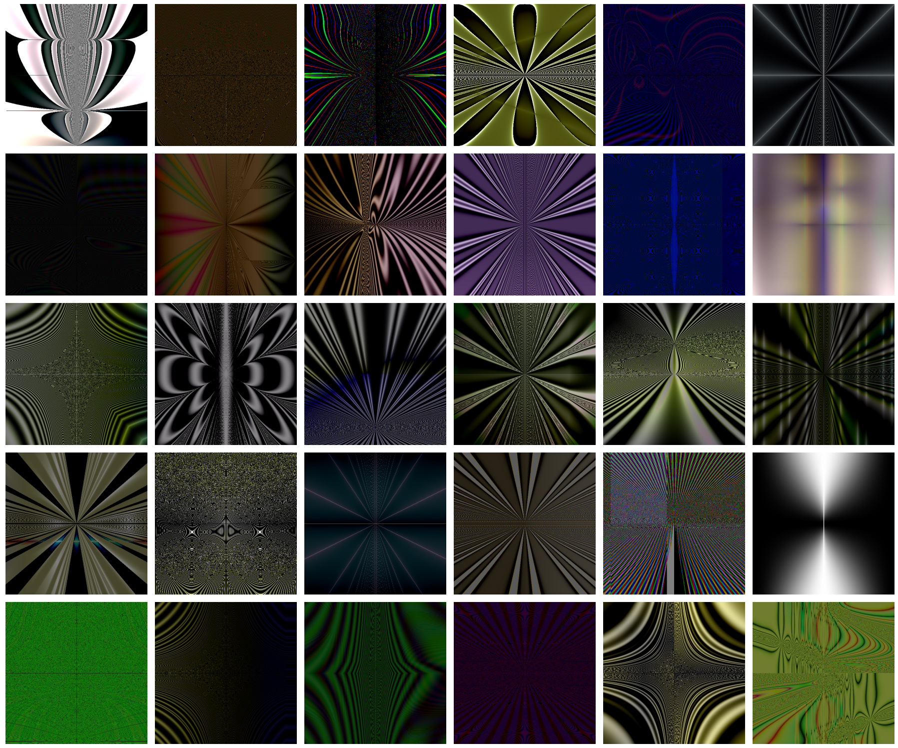

Fig. 8 Fittest individuals of the last generation of each of the 30 seeds of the first iteration.

imagery being explored in other evolutionary runs – they only operate within the production of a specific run.

Figure 9 presents a sample of the 38283 images evolved in the course of the first iteration and classified as external. Obviously, the visual inspection of 38283 and the selection of a representative set sufficiently small to present in this Chapter is close to impossible. Nevertheless, we believe that the selected samples illustrate the diversity of images classified as external that were evolved throughout the course of this iteration.

Based on the results presented, we believe it is safe to claim that the images classified as external are substantially different from the ones belonging to the initial dataset. On the other hand, it is also safe to state that they are substantially different from the external dataset composed of human-made artworks. In a nutshell, the EC engine is producing images that are distinct from both initial datasets, and that the classifier, which is forced to classify them into one of these two sets, identified as external. We also believe that it is safe to claim that these images are novel in relation to the ones previously produced by the EC system (i.e. the initial set of

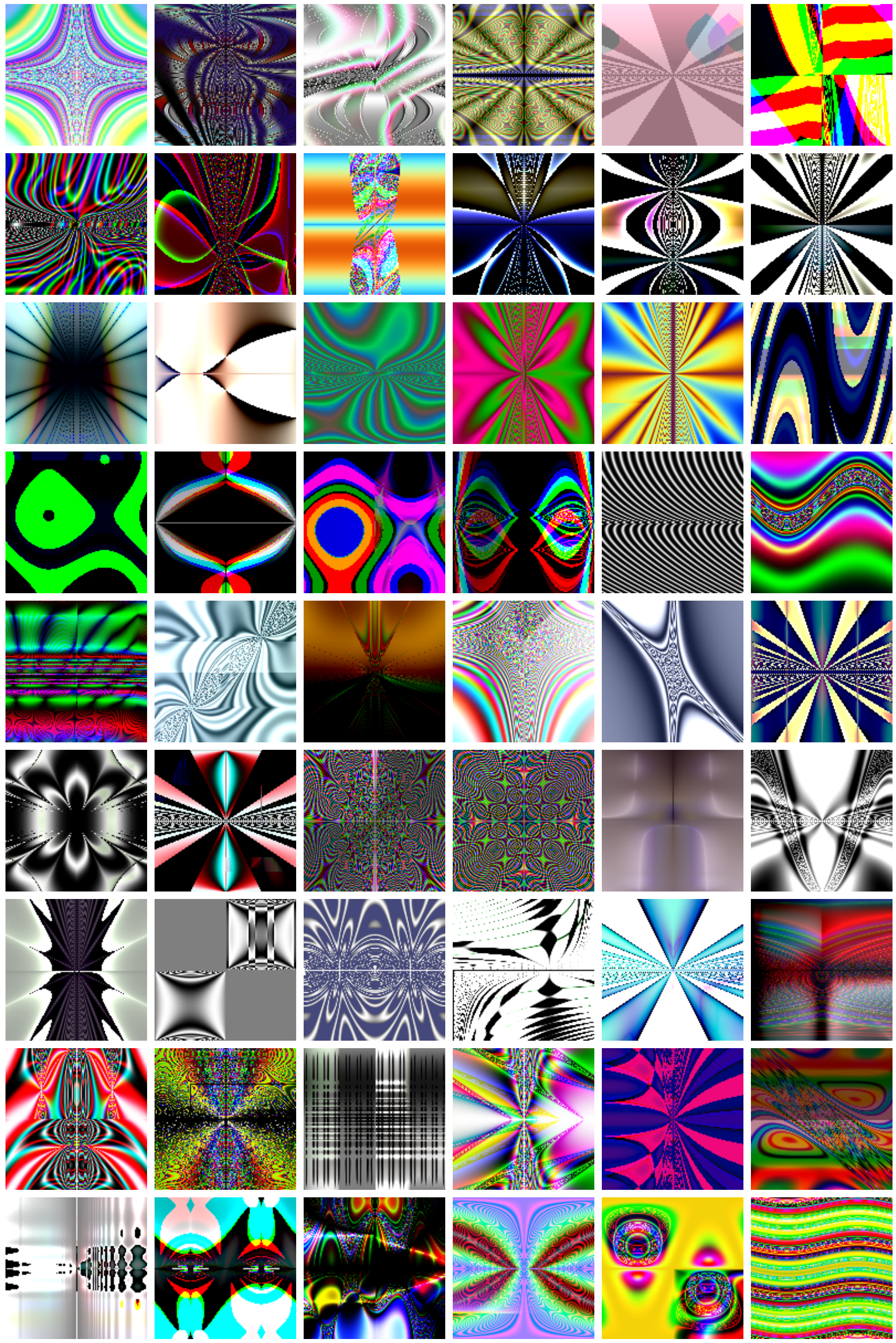

Fig. 9 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the first iteration.

internal images), not only from a computational perspective, but also to the human eye.

In our subjective opinion, several of these images are aesthetically interesting and appealing. Considering our background and experience using user-guided evolutionary art systems, which spans more than a decade, it is relevant to make the following observation: these images are, in many ways, similar to the ones we evolved through user-guided evolution in the course of these years. An anecdotal evidence of this fact is that, when confronted with Figure 9, one of the authors asked "Why are we including user-guided images?". Proving that this resemblance is real is beyond the scope of the present paper, nevertheless, even without strong evidence to make this claim, we consider this one of the most unexpected, and possibly relevant, results of this Chapter.

5.2.2 Intermediate Iterations

In this subsection we make an overview of the visual results obtained in the second to the twelfth iterations. These results are illustrated by the samples of the images classified as external presented in Figures 10 to 20. It is important to remember that, in most cases, for each of the images presented in the figures a significant number of images of similar nature was evolved throughout the corresponding run.

Rather than making a detailed analysis, we will focus on highlighting some of the most striking results obtained in each iteration, identifying, whenever possible, trends that emerge in several runs and that, as such, represent optima with large basis of attraction for the classifier being used in that particular interaction.

In the course of the second iteration the EC engine evolved 1436 images classified as external. These images result from 22 of the 30 runs. The images presented in the figures are ordered by evolutionary run. As such, two similar images presented side by side typically indicate that they were evolved in the same run, similar images that are not adjacent to each other typically indicate the rediscover of the same type of imagery in two different runs.

A brief scrutiny of the images presented in Figure 10 reveals that most of the evolutionary runs converged to different imagery, but also the recurrence of some themes. Among these, we highlight the stripped star-like shapes, which also emerged in the first iteration, and that continue to be present, although "rendered" in a different style. One of the interesting results concerns the evolution of several "minimalistic" images (e.g. the two rightmost images of the first row and the leftmost images of the fifth row), which occurs in several runs. Although they appear minimalistic, this type of image is particularly hard to evolve, and their simplistic nature contrasts with the size of their genotypes. In fact, an inspection of the learning process after the second iteration appears to indicate that the emergence of these images is deeply related to the presence in the initial dataset of external imagery that are also minimalistic and monochromatic (see Subsection 5.3. In fact the use of a reduced color palette occurs in several of the evolutionary runs. This is consistent with the color schemes used in many of the images belonging to the external dataset and

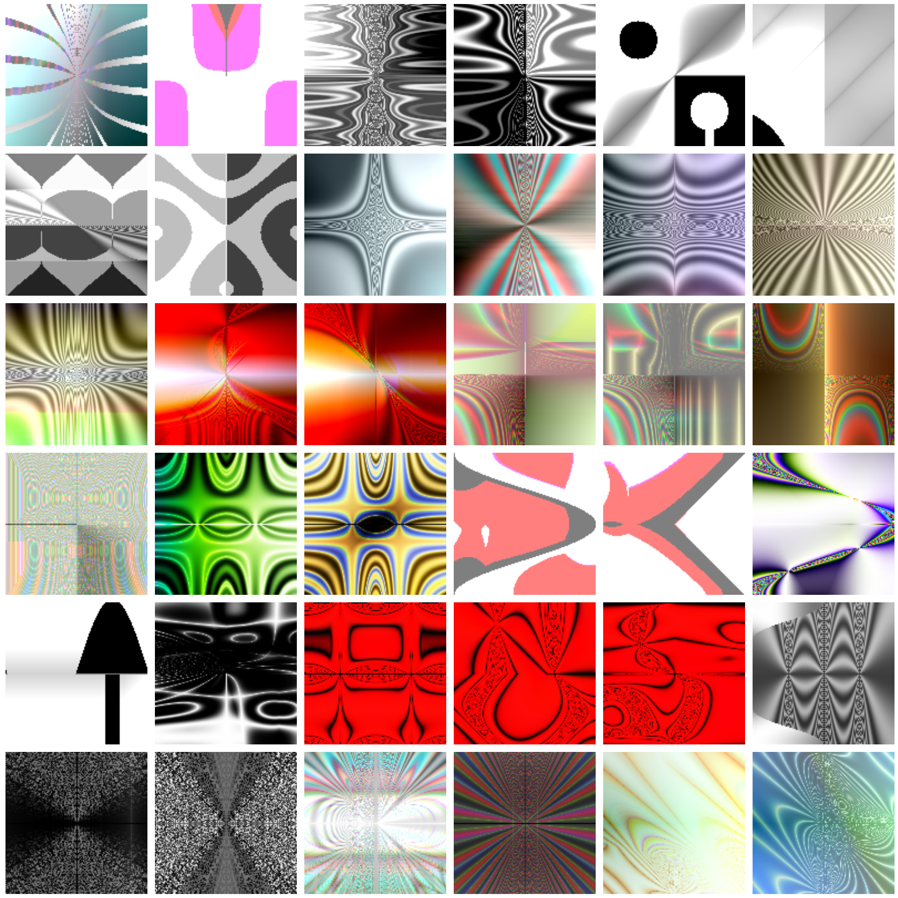

Fig. 10 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the second iteration.

contrasts with with the typical imagery produced by the EC engine. More importantly, considering the nature of this Chapter, the appearance of the evolved images classified as external appears, in most cases, to be different from the initial dataset of external images and from the images evolved in the course of the first iteration.

Analysing the images produced in the course of the third iteration, of which a sample is presented in Figure 11, one can observe the same overall patterns: most runs tend to converge to different types of images; most evolved novel imagery in relation with the previous production of the system; there are some recurring themes, namely the star-like images, which are "rendered" in different styles. The emergence of images with strong and contrasting colors (magenta, green, yellow, white, black) occurs in several evolutionary runs. This type of imagery is highly atypical of the

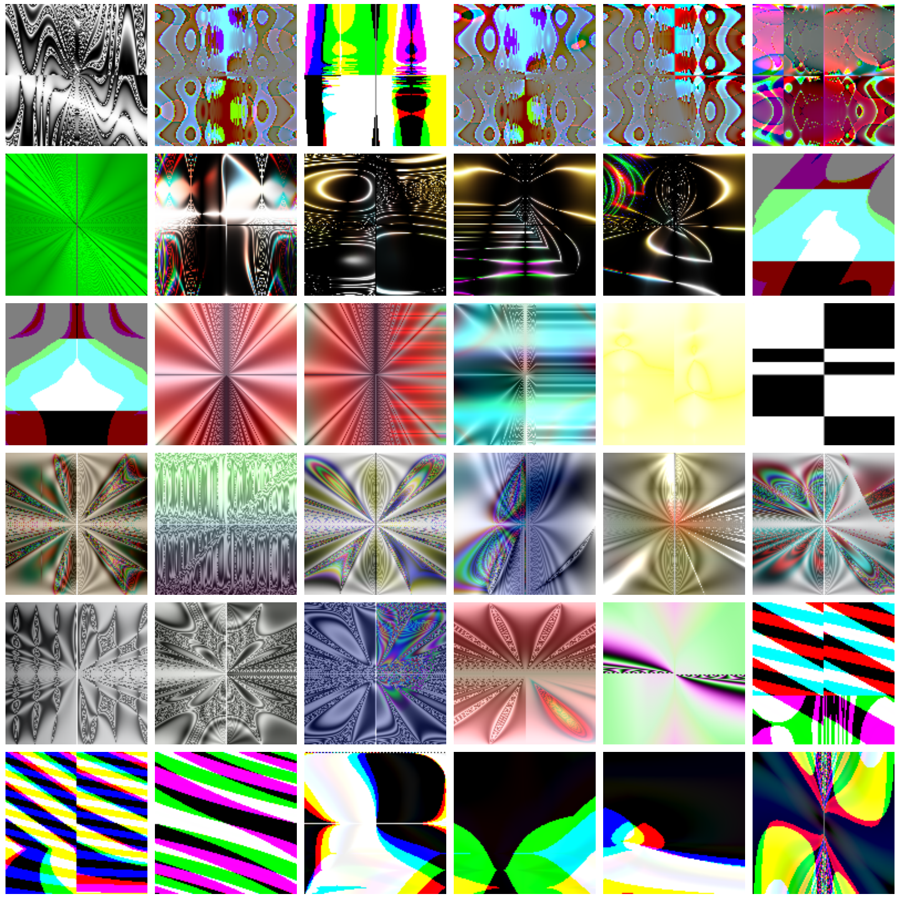

Fig. 11 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the third iteration.

EC engine and matches the chromatic characteristics of several of the artworks of the external set.

As we will see when analysing the results of other iterations, the emergence of graphic elements such as lines, points and planes, also characterises some of the evolved images. Although these are usually considered graphic primitives for humans, the EC engine has no explicit way of creating such elements. As such, their emergence is deeply linked with the fitness landscape induced by the classifiers.

Much of what was stated regarding the images evolved in the third iteration also applies to the ones evolved in the fourth (see Figure 12). Many of the images are characterised by the emergence of organic lines and planes. Others appear to be composed of multiple layers with transparencies (e.g. leftmost image of the third row).

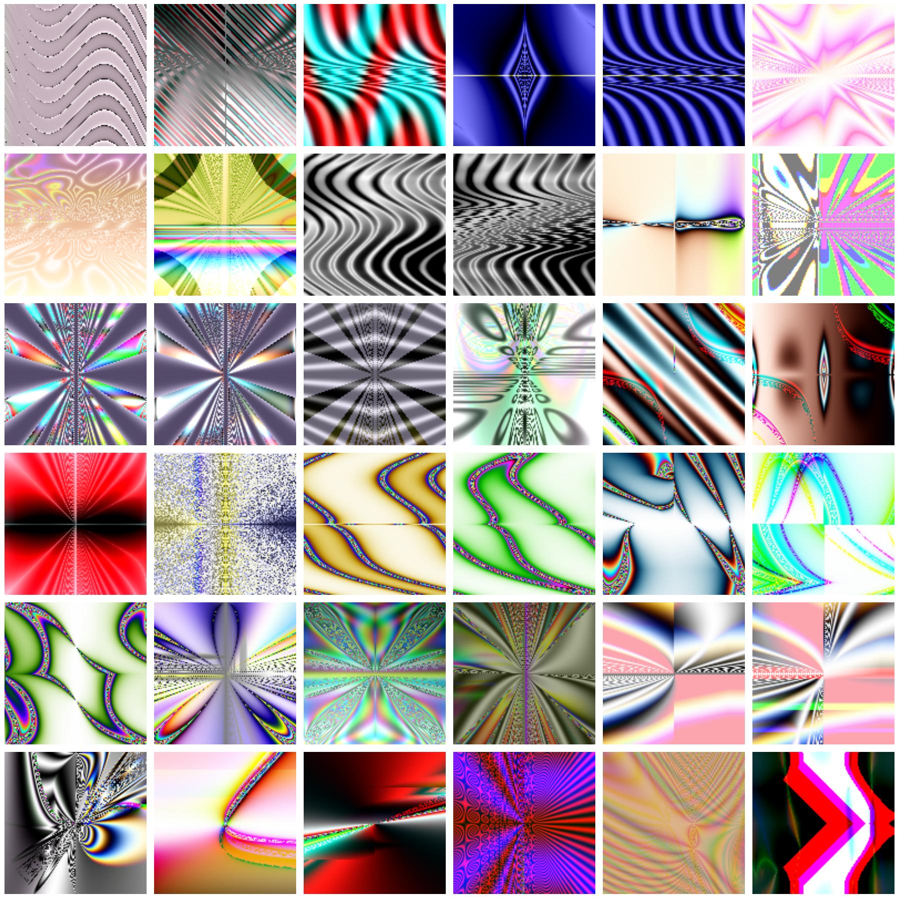

Fig. 12 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the fourth iteration.

In the fifth iteration, the EC engine experienced difficulties in finding images classified as external. Only 10 of the 30 evolutionary runs found such images and, on average, these took 29.3 generations to evolve. In total, 366 images classified as external were evolved, a number that is reduced to 57 by our archiving algorithm. For these reasons, the diversity of the images presented in Figure 13 is not as large as in previous iterations. The feature common to all of these images is the presence of "noise" patterns. It is also interesting to notice that a vast percentage of the images is monochromatic and with intricate detail. In several of the cases (e.g. the black and white images of the first and last row) the lines are discontinued, in the sense that they emerge from the arrangement of several white or gray dots that are not actually connected.

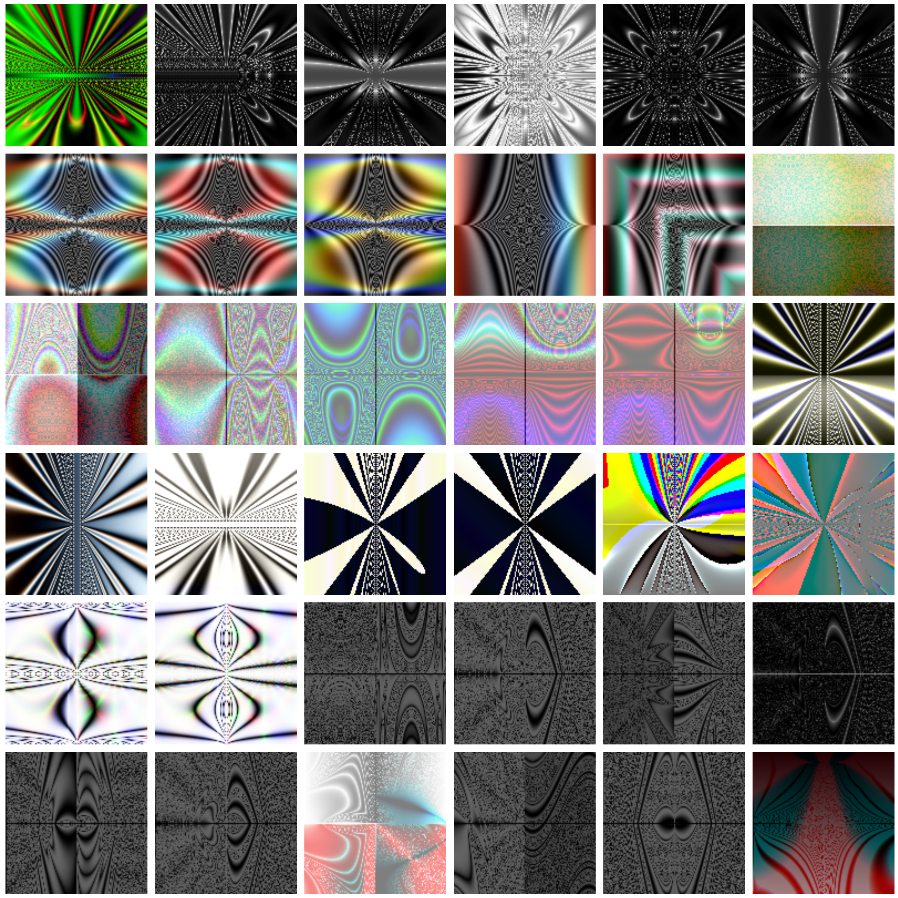

Fig. 13 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the fifth iteration.

Although the productivity of the AA during the fifth iteration was not high, the addition of these images to the internal dataset, coupled with the removal of some of the external images (see Subsection 5.3), appears to cause profound changes in the classifier. There is a burst of productivity in the course of the sixth iteration, 1105 images of which 433 are added to the archive, a number that is only surpassed by the first iteration. As Figure 14 illustrates, this burst of productivity coincides with a change of style in comparison with the previous iterations. This sudden increase of productivity can be explained by the performance of the classifier, and will be discussed in Subsection 5.3.

Productivity decreases during the seventh iteration, see Figure 15, and reaches an all time low in the eighth iteration (refer to Figure 16). Generally speaking, one can state that the images classified as external evolved in the course of the seventh

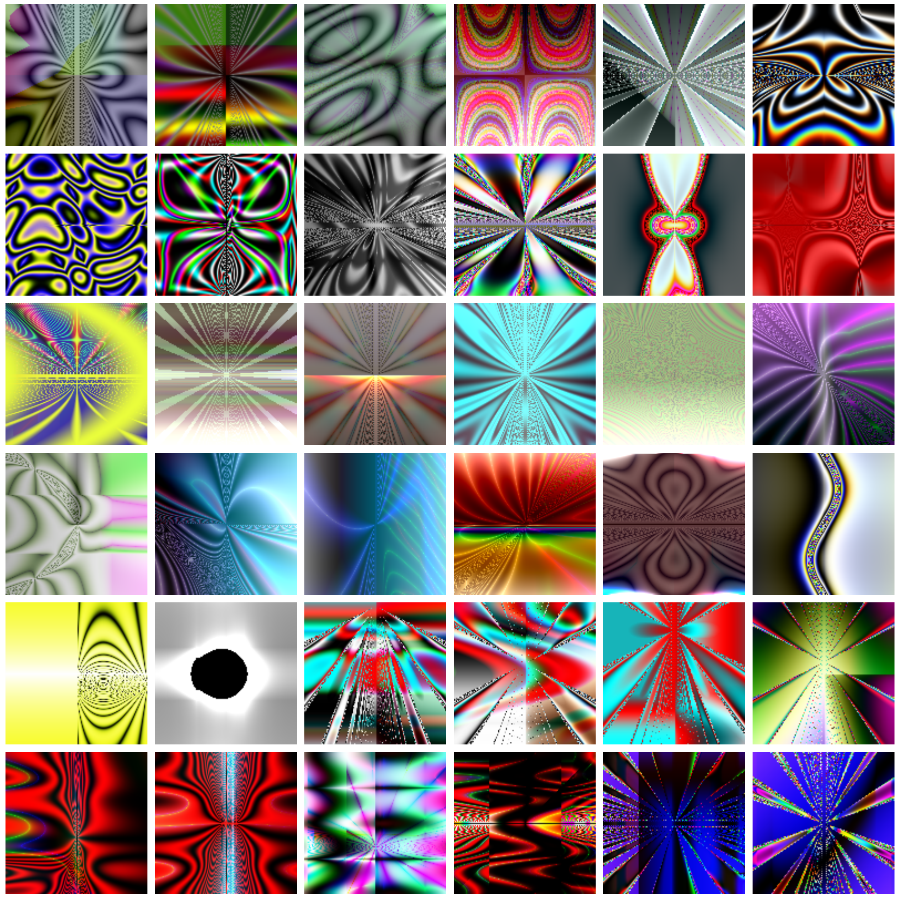

Fig. 14 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the sixth iteration.

iteration correspond to variations in style of themes already explored in previous iterations, almost as if the AA further refined and included additional detail to previously explored images. The ones evolved in the eighth iteration appear, in our eyes, to be of the same style as images evolved in some of the previous iterations. The classifier is not able, even during training, to fully discriminate among the internal and external datasets. As previously explained, this opens the door for the repetition of styles and imagery that was not sufficiently explored in previous iterations. Our analysis indicates that this is what happened in the course of the eighth iteration, the AA artist explored styles that, although already present, were not sufficiently explored. As the cardinality of such images increases the Classifier system is “forced” to recognise such images as internal and, therefore, the EC engine will no longer be able to explore them in future iterations.

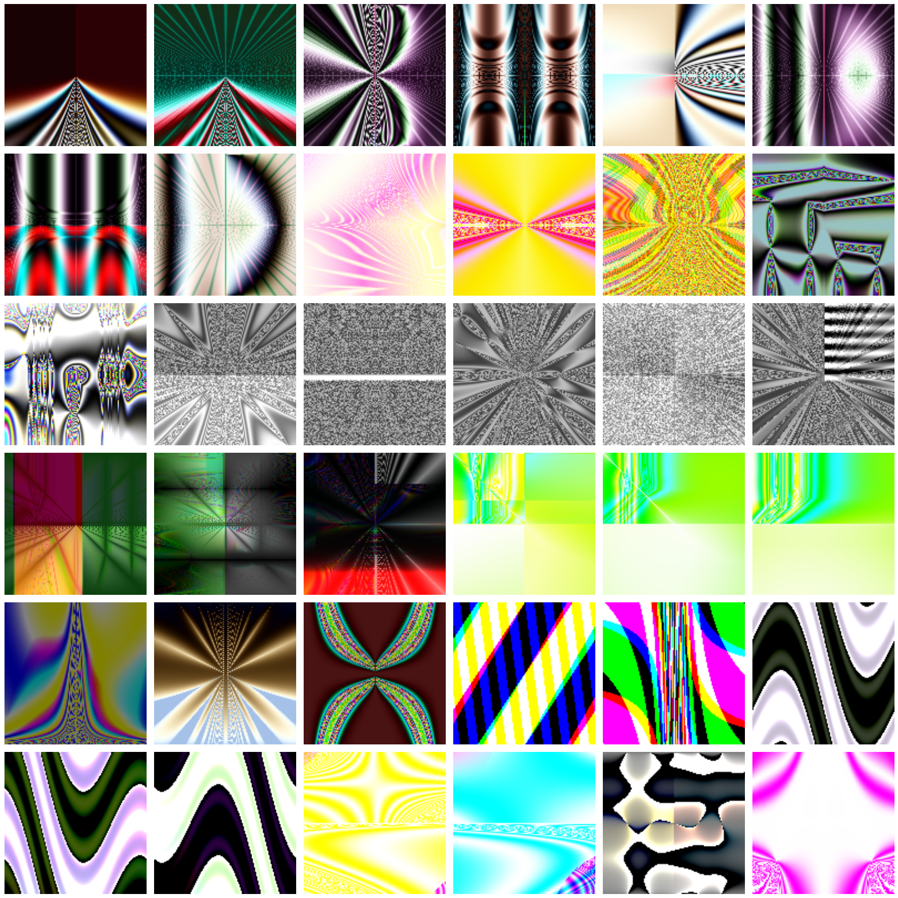

Fig. 15 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the seventh iteration.

Like previously, the changes to the datasets gave rise to a classifier that bases its assessments on different premises, inducing a different fitness landscape, which happens to be more prone to evolution. In the ninth iteration, the EC engine evolved a total of 422 images, of which 115 where archived, finding images classified as external in 21 of the 30 runs. As can be observed by inspecting Figure 17, there is a mixture between new and old themes and styles. Interestingly, several images that are evocative of landscapes (three leftmost images of the third row) were evolved.

The tenth iteration was one of the most productive ones in terms of the total number of images classified as external, 692, but of these only 62, less than 10% made it to the archive. As it can be observed in Figure 18, several runs converged to the same type of imagery, reducing the overall diversity and productivity of the set. Visually, we identify three main styles which emerge in several runs: the black and

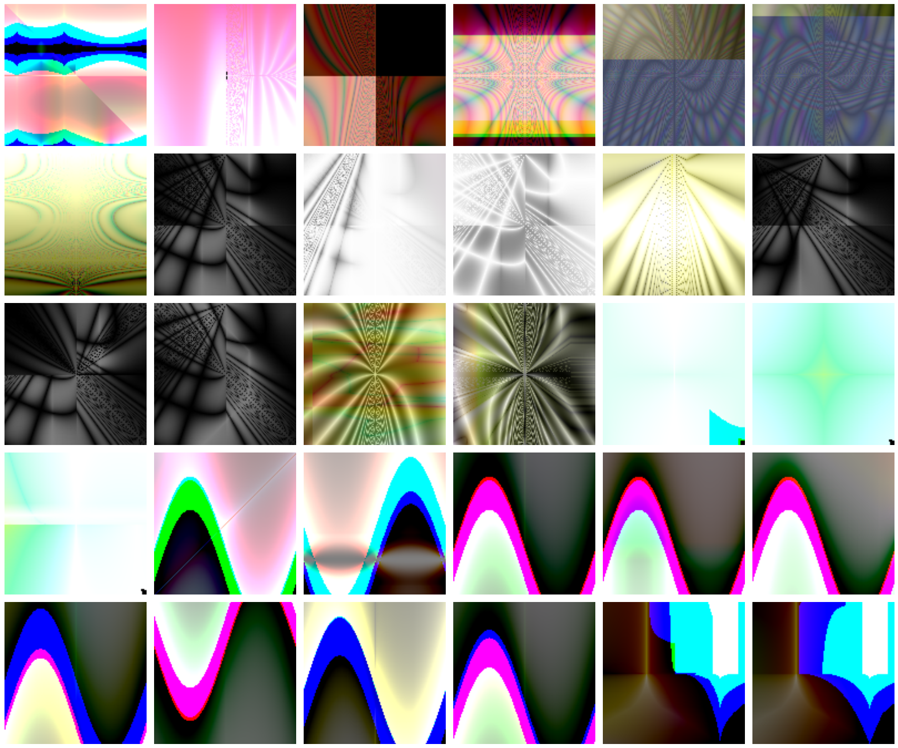

Fig. 16 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the eighth iteration.

white minimalistic images; images that appear to have a white transparent layer (e.g. the five leftmost images of the first row, but also the two rightmost of the last row); The images exploring a combination of magenta and green. Like in several of the previous iterations, the star-like shape continues to be one of the favorite “themes” of the AA.

The lack of diversity of the tenth iteration contrasts to the visual diversity of the eleventh. Although only 374 images classified as externals were found, 126 of these images were archived. On average it took 32.27 generations to find the first image classified as external. Although there is some stylistic agreement among several evolutionary runs (see Figure 19) the overall diversity is significantly higher than in the previous iteration. The purely black and white images disappear from this iteration onwards, likely due to the combination of two factors: the inclusion of several of these images in the internal dataset and, most importantly, the removal of a large number of strictly black and white images from the external dataset.

The twelfth iteration is among the least productive ones, only 267 images classified as external were found and only 11 of the 30 runs found such images. In spite

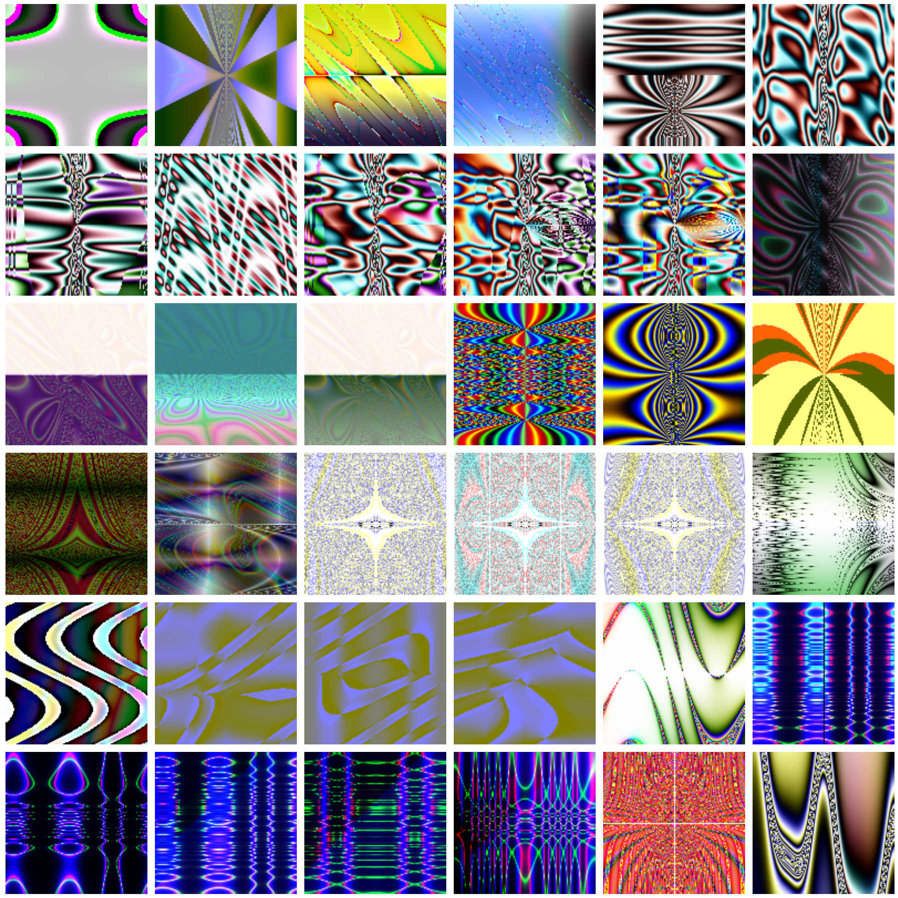

Fig. 17 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the ninth iteration.

of this lack of productivity, visible in Figure 20, some of the evolutionary runs were able to find novel imagery that contrasts both in terms of style and theme from the previous artistic production of the system.

5.2.3 Thirteenth Iteration

The thirteenth, and last iteration presented in this Chapter, corresponds to burst of novelty and productivity of the system. Although a similar burst occurred in the sixth iteration, the nature of the burst appears significantly different. In this case, the increased productivity is coupled with significant stylistic variations and may be seen as a moment where the AA actually “broke the mould”.

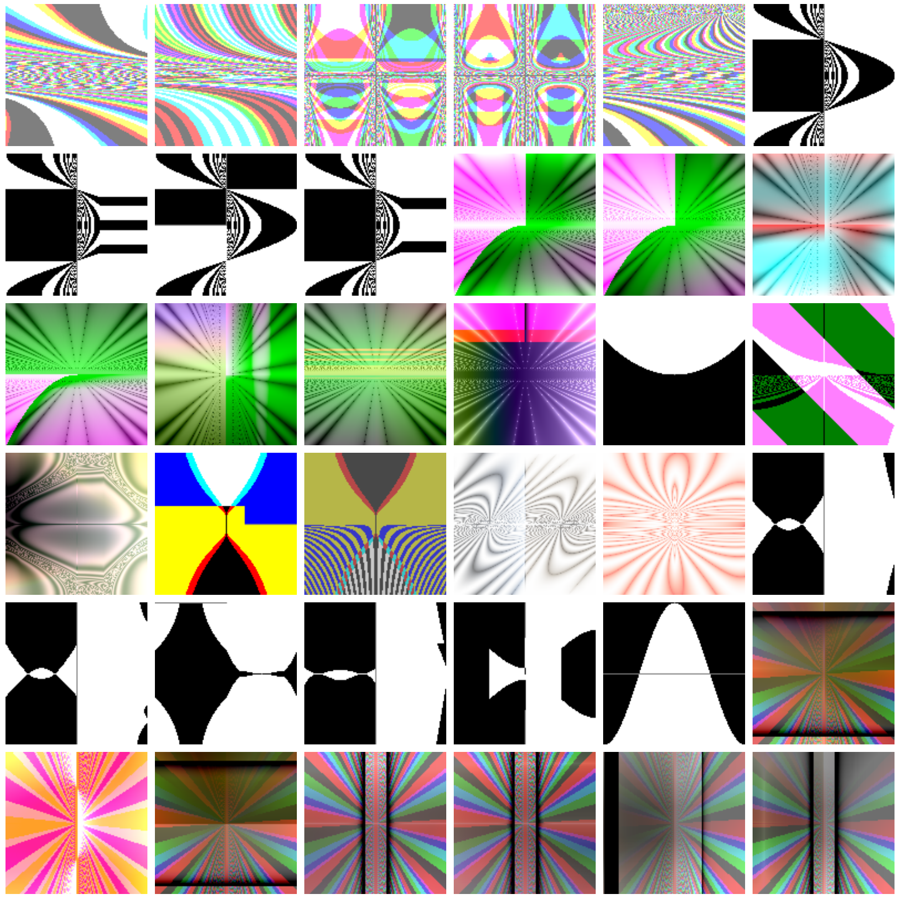

Fig. 18 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the tenth iteration.

As we will see in the next section, while the increase of productivity in the sixth iteration seems to be linked with shortcomings of the classifier, here it appears to be linked with significant changes to the way the classifier system differentiates between the class of internal and external imagery. Thus, while in the other intermediate iterations the AA seems to be making minor stylistic variations of images that it has already produced, and opportunistic exploitations of shortcomings of the classifier, what happens in the thirteenth iteration seems to be rather different, resulting from profound changes of the aesthetic model, caused by the cumulative revision of the internal and external set. Making an analogy, this can be seen as an "Eureka" moment, where the system discovers substantially different styles, expanding and enriching the range of its artistic production.

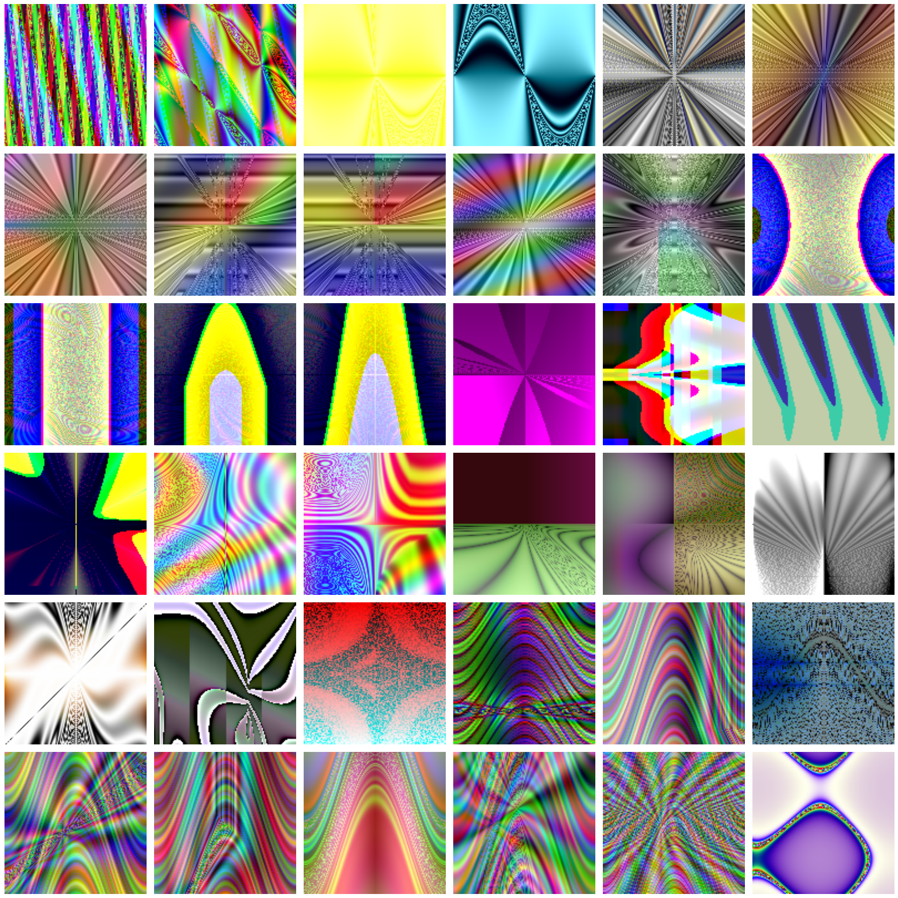

Fig. 19 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the eleventh iteration.

As it can be observed by the sample of the images presented in Figure 21, although there are some recurring themes, the detail of "execution" of the images of the thirteenth generation classified as external is a lot higher than in previous iterations. The images seem to be more elaborate, detailed, and refined, when compared with previous iterations. At the same time, some novel ornamentation techniques, such as the one depicted in the three rightmost images of the first column, were discovered, and some novel themes seem to emerge. The exploration of "light" (see, e.g. leftmost image of the third row and the rightmost image of the fifth row) also emerges as a visible and distinctive traits.

We considered the images resulting from the first iteration comparable to the ones evolved through user-guided evolution. Although, in fairness, the same could be stated for a significant portion of the images evolved in the course of the thir-

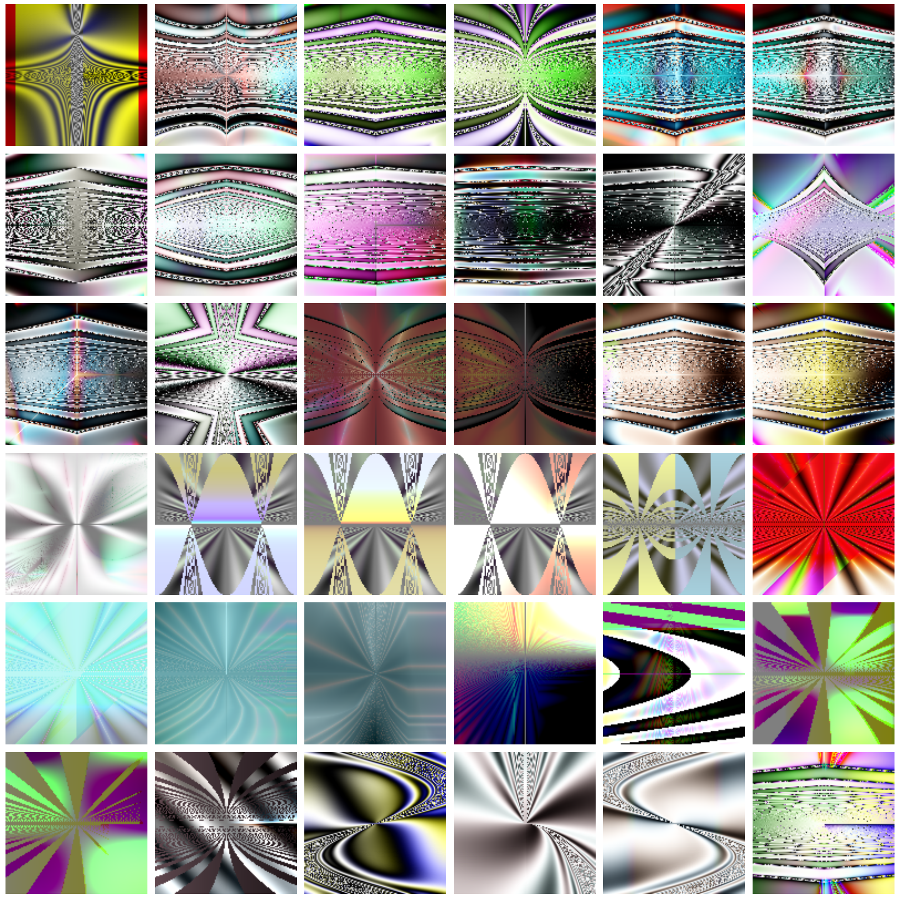

Fig. 20 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the twelfth iteration.

teenth, it is equally fair to state that some of the runs created imagery that is stylistic dissimilar from what we have evolved through user-guided evolution, or other means. Thus, many of these images strike us not only as novel in relation to the previous artistic production of the AA, but also as novel and surprising in relation to our own experience and production.

5.3 Training of the Classifiers

In this subsection we make an overview of the results pertaining the training of the classifier. As previously mentioned (see Section 4.1), when an iteration is concluded

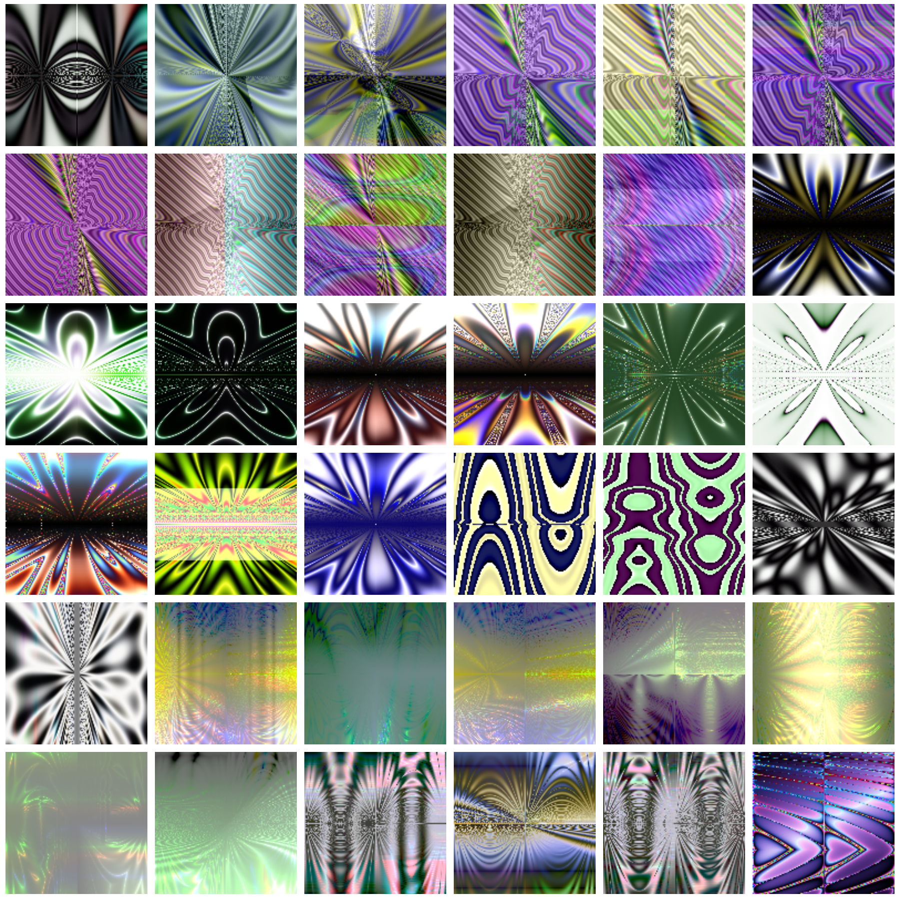

Fig. 21 Samples of the images classified as external, generated throughout the course of the 30 evolutionary runs of the thirteenth iteration.

the images classified as external, evolved in the course of the iteration, are gathered by the supervisor and added to the internal dataset. This is followed by several training attempts, which may imply removing images from the external dataset. Training is concluded when no external images are classified as being internal. The existence of internal images classified as external implies that the EC engine may revisit previous styles, but, when this occurs, the consequent increase of the number of images of those styles present in the internal dataset will eventually force the classifier to recognise such images as internal.

Table 6 presents a summary of pertinent statistics regarding the training phase. It details, per iteration: the number of training attempts necessary to reach a classifier without false internals (*attempts*); the total number of false externals identified in the course of these attempts (*False Externals*); the number of false externals and internals after the training attempts are concluded (*CS False External* and *CS False Internal*, respectively), which reflects the ability of the classifier that is going to be used to guide the following iteration to discriminate between the sets; the size of the external and internal datasets after training is concluded (*Total External* and *Total Internal*).

As it can be observed, the training of the classifier for the first iteration required two attempts. One external image, a black and white photograph of a detail of a painting, was removed from the external dataset. Unfortunately, due to copyright issues this and other external images cannot be replicated in the Chapter. After the second attempt, no external images were classified as internal, but one internal image, a black and white a star-like shape was deemed as external. The existence of this image, and the difficulty in classifying it may, at least partially, explain the recurrence of such theme in several iterations.

As previously mentioned, the first iteration generated a wide and varied number of images. As a consequence, 30110 images have been added to the internal dataset and the training of the classifier that guided the second iteration took a significantly larger number of attempts; in total 68 external images have been removed. These are, mostly, black and white engravings, and quite interestingly, some images of mathematical objects, an artwork of M. C. Escher, which also has a mathematical appearance, and a cartoon image. In what concerns the engravings, we believe that these images were removed by two main reasons: (i) at the resolution that the FE processes these images, they could easily be confused with images produced by the EC engine; (ii) more importantly, these images tend to be atypical in relation to the other images belonging to the external dataset, which makes them harder to classify. In what concerns the images of mathematical objects, they seem to be computer generated and, therefore, the confusion with the images produced by the AA is natural. After the removal of these images the classifier is able to fully distinguish between the two sets.

The same overall trend occurs in iterations 2 to 3, although the number of attempts varies (4 and 7, respectively) the types of external images being excluded is the same, including engravings, M. C. Escher artworks, black and white drawings, photographs of sculptures, and minimalistic paintings (some of them by Kazimir Malevich). The reasons for their misclassification are the same: they are either atypically in relation with the rest of the external set or, confusable with computer generated imagery, i.e., similar in style with the images the EC engine is prone to create.

The fourth generation provoked few changes of the external dataset, removing only 11 images. While 10 of these are black and white drawings, the remaining one is notable since it is the first Mondrian removed from the external set. The images produced in the fifth generation, provoked the exclusion of 18 images from the external dataset, among which two by Matisse and two by M.C. Escher. The most relevant issue concerning the training subsequent to the fifth iteration is that the resulting classifier, that will guide evolution in the sixth iteration, misclassifies 28 internal images. This gives the EC engine a large degree of freedom to explore previously visited imagery, which explains the burst of productivity observed in the sixth iteration. The exploitation of these “shortcomings” leads to the generation of images that, once added to the internal dataset prevent their future exploitation.

In the seventh iteration, a total of 248 external images were removed from the external dataset, these include: Black and white engravings, several M. C. Escher artworks (14 to be precise), several Mondrian paintings, and numerous line drawings. The eighth iteration, the least productive of all, provoked the removal of 31 external images, that tend to be of the same type as the ones previously identified. After these removals, the classifier is, again, able to fully distinguish among the two sets.

The ninth and tenth iterations caused the removal of few external images, 8 and 6, respectively. Confirming our previous observation that, although these iterations were productive, they were not particularly fruitful in terms of the novelty of the evolved imagery. As mentioned previously, although 692 images classified as external were evolved in the tenth iteration, only 62 of these made it to the archive.

These numbers contrast with the ones of the eleventh and twelfth iteration where, respectively, 47 and 63 were deleted. These include several Picasso, Dalí, Paul Klee, Mondrian, and Mark Rothko paintings, as well as several line drawings. The lack of texture, at low resolution, appears to be the binding trait of these paintings.

As mentioned previously, in our opinion, the thirteenth iteration is different from the others in the sense it corresponds to a pronounced shift in style. As such, it is particularly interesting to inspect what kind of changes to the external dataset the evolved imagery induces. In total 64 external images where removed. In previous iterations most of the removed images were black and white drawings or engravings, this is not the case in the thirteenth generation; only ten of the deleted images are black and white. The remaining images are Renascence style paintings (unfortunately, they do not include the Mona Lisa), two Van Gogh paintings, three by Kandisky and one by Miró. Quite interestingly, four paintings of Monet’s Waterloo Bridge, a theme that is present in several of his artworks, were also removed. In this case it was possible, and quite easy we might add, to identify the evolved images that promote the confusion between internal images and these artworks. Some of them are depicted in the bottom two rows of Figure 21, and we believe that the reader will also understand why, in the eyes of the classifier, they can be easily confused.

6 Conclusions and Future Work

In this Chapter we presented an Artificial Artist that is characterised by its permanent quest for novelty. The system is composed of two main modules: a generator and an evaluator. The role of the generator is played by an expression based evolutionary engine and that of the evaluator by an artificial neural network. The network is trained to discriminate among the images produced by the evolutionary engine and a set of famous artworks. The fitness of the images being evolved depends on the output of the classifier, promoting the discovery of images that the network classifies as external. In each iteration of the framework we perform 30 evolutionary

Table 6 Statistics regarding the training of the classifiers of each iteration in terms of: number of attempts, false internals and externals during training, false externals after training, and size of the internal and external and internal dataset after training.

		During Training Cycles		After Training Attempts			
Iteration	Attempts	False External	False Internal	CS False Externals	CS False Internals	Total Externals	Total Internal
initial	2	1	1	1	0	26238	26239
1	7	23	68	0	0	26170	56349
2	4	52	20	3	0	26150	56599
3	7	41	67	3	0	26083	56794
4	6	7	11	3	0	26072	56972
5	2	40	18	28	0	26054	57029
6	2	7	18	3	0	26036	57462
7	3	11	248	3	0	25788	57593
8	3	23	31	0	0	25757	57624
9	3	11	8	5	0	25749	57739
10	3	126	6	5	0	25743	57801
11	8	145	47	4	0	25696	57927
12	6	99	63	7	0	25633	58028
13	4	88	64	3	0	25589	58380

runs. When these are concluded, the relevant misclassified images are added to the set representing the production of the AA and the neural network is retrained.

For the above reasons, the approach promotes and explores a competition between generator and evaluator. From a theoretical standpoint – assuming that the evolutionary engine and the artificial neural network are adequate and always able to cope – the iterative expansion of the internal set leads, necessarily, to change since the evolutionary algorithm is forced to explore new paths. Moreover, assuming that a sufficiently large number of iterations is performed and that both systems cope, the convergence to the aesthetic model (or models) implicitly defined by the set of external images, which provides an aesthetic reference to the artistic production of the AA, is bound to eventually occur.

To increase the diversity within evolutionary runs, and prevent their early convergence and stagnation, we include mechanisms to promote the phenotype diversity of the populations. This implies taking two criteria into account when performing tournament selection: the adequacy of the image (which results from the output of the neural network) and its diversity in relation to the images produced in the course of the same run.

The analysis of the experimental results confirms the adequacy and potential of the approach, revealing that the system is able to consistently produce novel imagery of arguable, aesthetic merit. As such, we consider that we successfully developed a creative system that is able to learn, create and innovate in an entirely autonomous way.

The experimental results indicate that the images produced in the course of the first iteration of the framework are similar to those produced by expression-based interactive evolutionary art systems, where the role of the evaluator is played by a human. Analysing the results, we consider that the behaviour and production of the system during the first twelve iterations can be considered as e-creative. We consider that what happens during the thirteenth generation is significantly different and goes beyond e-creativity. In this case, the system made a qualitative and substantial change both in terms of production and aesthetic model, thus breaking the mould. We put forward the hypothesis that this behavior can be seen as a limited case of h-creativity – in the sense that the system produced images that appear to be different from those previously attained by evolutionary means – and of t-creativity, in the sense that the changes appear to be related from deep changes in the aesthetic model and, therefore, of a profound transformation of the search space.

Future work will focus on two main aspects: further improvement of the framework, which includes additional testing, namely testing some of the hypothesis raised. In what concerns the refinement of the framework, we consider that the framework can only be fully assessed once it is pushed to its limits. We present the results of thirteen iterations, but further ones are still being performed. Although it is impossible to predict, the number of iterations necessary to provoke the collapse of one of the modules is likely to be large. By taking the framework to the point of collapse, we hope to gain additional insights regarding the limitations of the evolutionary engine and, specially, of the classifier. Such insights will guide future developments. The inclusion of a curator module, which would select from the artistic production of the Artificial Artist a small set of artworks, would also be a valuable addition to the framework. From a conceptual standpoint, this confers to the system a degree of introspection and self-analysis that it currently lacks. From a practical perspective, it would avoid the need of hand-picking representative examples, which may induce a bias in the presentation of the experimental results. In what concerns testing, we are particularly interested in verifying if the images that were evolved in the course of the first interaction are indeed similar, to humans and computers, to those resulting from user-guided evolution. For that purpose, we are gathering a set of user evolved images and we will test if humans and computers are able to discriminate between them.

Acknowledgments

The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 611733. This research is also partially funded by: Fundação para a Ciência e Tecnologia (FCT), Portugal, under the grant SFRH/BD/90968/2012.

References

Atkins, D. L., Klapaukh, R., Browne, W. N., & Zhang, M. (2010). Evolution of aesthetically pleasing images without human-in-the-loop. In *Ieee congress on evolutionary computation* (p. 1-8). IEEE.

Baluja, S., Pomerlau, D., & Todd, J. (1994). Towards automated artificial evolution for computer-generated images. *Connection Science*, *6*(2), 325-354.

Barnsley, M. F. (1993). *Fractals everywhere* (2nd ed.). Cambridge, MA, USA: Academic Press.

Boden, M. (2004). *The creative mind: Myths and mechanisms*. Routledge.

Correia, J. (2009). *Evolutionary Computation for Assessing and Improving Classifier Performance* (Unpublished master's thesis). Department of Informatics Engineering of the University of Coimbra.

Correia, J., Machado, P., Romero, J., & Carballal, A. (2013a). Evolving figurative images using expression-based evolutionary art. In *Proceedings of the fourth international conference on computational creativity (iccc)* (p. 24-31).

Correia, J., Machado, P., Romero, J., & Carballal, A. (2013b). Feature selection and novelty in computational aesthetics. In P. Machado, J. McDermott, & A. Carballal (Eds.), *Evolutionary and biologically inspired music, sound, art and design - second international conference, evomusart 2013, vienna, austria, april 3-5, 2013. proceedings* (Vol. 7834, p. 133-144). Springer.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. In *Computer vision – eccv 2006, 9th european conference on computer vision, part iii* (p. 288-301). Graz, Austria: Springer.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008, May). Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, *40*, 5:1–5:60. Retrieved from `http://doi.acm.org/10.1145/1348246.1348248` doi: http://doi.acm.org/10.1145/1348246.1348248

den Heijer, E. (2012). Evolving art using measures for symmetry, compositional balance and liveliness. In *Ijcci* (pp. 52–61).

den Heijer, E., & Eiben, A. E. (2010). Comparing aesthetic measures for evolutionary art. In C. D. Chio et al. (Eds.), *Applications of evolutionary computation, evoapplications 2010: Evocomnet, evoenvironment, evofin, evomusart, and evotranslog, istanbul, turkey, april 7-9, 2010, proceedings, part ii* (Vol. 6025, p. 311-320). Springer. doi: 10.1007/978-3-642-12242-2-32

Ekárt, A., Joó, A., Sharma, D., & Chalakov, S. (2012). Modelling the underlying principles of human aesthetic preference in evolutionary art. *Journal of Mathematics and the Arts*, *6*(2-3), 107-124. Retrieved from `http://www.tandfonline.com/doi/abs/10.1080/17513472.2012.679489` doi: 10.1080/17513472.2012.679489

Faria, J., Bagley, S., Ruger, S., & Breckon, T. (2013). Challenges of finding aesthetically pleasing images. In *Image analysis for multimedia interactive services (wiamis), 2013 14th international workshop on* (pp. 1–4).

Goshtasby, A. A. (2012). Image registration: Principles, tools and methods. In (pp. 7–66). London: Springer London. Retrieved from http://dx.doi.org/10.1007/978-1-4471-2458-0_2 doi: 10.1007/978-1-4471-2458-0_2

Greenfield, G. (2002). Color dependent computational aesthetics for evolving expressions. In R. Sarhangi (Ed.), *Bridges: Mathematical connections in art, music, and science; conference proceedings 2002* (pp. 9–16). Winfield, Kansas: Central Plains Book Manufacturing.

Greenfield, G. (2003). Evolving aesthetic images using multiobjective optimization. In B. McKay et al. (Eds.), *Congress on evolutionary computation, cec 2003* (Vol. 3, pp. 1903–1909). Canberra, Australia: IEEE Press.

Greenfield, G. (2005). Evolutionary methods for ant colony paintings. In *Applications of evolutionary computing, evoworkshops2005: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoMUSART, EvoSTOC* (Vol. 3449, pp. 478–487). Lausanne, Switzerland: Springer Verlag.

Greenfield, G. R. (2002). On the co-evolution of evolving expressions. *International Journal of Computational Intelligence and Applications*, *2*(1), 17–31. Retrieved from http://dx.doi.org/10.1142/S1469026802000403

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009, November). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, *11*(1), 10–18. Retrieved from http://doi.acm.org/10.1145/1656274.1656278 doi: 10.1145/1656274.1656278

Kowaliw, T., Dorin, A., & McCormack, J. (2009). An Empirical Exploration of a Definition of Creative Novelty for Generative Art. In K. B. Korb, M. Randall, & T. Hendtlass (Eds.), *Acal* (Vol. 5865, p. 1-10). Springer.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by natural selection*. Cambridge, MA: MIT Press.

Lehman, J., & Stanley, K. O. (2008). Exploiting open-endedness to solve problems through the search for novelty. In *Proc. of the eleventh intl. conf. on artificial life (alife xi).* Cambridge, MA: MIT Press.

Li, C., & Chen, T. (2009). Aesthetic Visual Quality Assessment of Paintings. *IEEE Journal of Selected Topics in Signal Processing*, *3*(2), 236–252. Retrieved from http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4799314 doi: 10.1109/JSTSP.2009.2015077

Li, Y., Hu, C., Chen, M., & Hu, J. (2012). Investigating aesthetic features to model human preference in evolutionary art. In P. Machado, J. Romero, & A. Carballal (Eds.), *Evolutionary and biologically inspired music, sound, art and design - first international conference, evomusart 2012, málaga, spain, april 11-13, 2012. proceedings* (Vol. 7247, pp. 153–164). Springer. Retrieved from http://dx.doi.org/10.1007/978-3-642-29142-5_14 doi: 10.1007/978-3-642-29142-5_14

Liapis, A., Yannakakis, G. N., & Togelius, J. (2013). Sentient sketchbook: Computer-aided game level authoring. In *Fdg* (pp. 213–220).

Machado, P., & Cardoso, A. (1997). Model proposal for a constructed artist. In

N. Callaos, C. Khoong, & E. Cohen (Eds.), *First world multiconference on systemics, cybernetics and informatics, sci97/isas97* (p. 521-528). Caracas, Venezuela.

Machado, P., & Cardoso, A. (2002). All the truth about NEvAr. *Applied Intelligence, Special Issue on Creative Systems*, *16*(2), 101-119.

Machado, P., Correia, J., & Assunção, F. (2015). Graph-Based Evolutionary Art. In A. Gandomi, A. H. Alavi, & C. Ryan (Eds.), *Handbook of genetic programming applications.* Berlin: Springer.

Machado, P., Correia, J., & Romero, J. (2012a). Expression-based evolution of faces. In *Evolutionary and biologically inspired music, sound, art and design - first international conference, evomusart 2012, málaga, spain, april 11-13, 2012. proceedings* (Vol. 7247, p. 187-198). Springer. doi: 10.1007/978-3-642-29142-5_17

Machado, P., Correia, J., & Romero, J. (2012b). Improving face detection. In A. Moraglio, S. Silva, K. Krawiec, P. Machado, & C. Cotta (Eds.), *Genetic programming - 15th european conference, eurogp 2012, málaga, spain, april 11-13, 2012. proceedings* (Vol. 7244, p. 73-84). Springer. doi: 10.1007/978-3-642-29139-5_7

Machado, P., Dias, A., & Cardoso, A. (2002). Learning to colour greyscale images. *The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour – AISB Journal*, *1*(2), 209-219.

Machado, P., Romero, J., Cardoso, A., & Santos, A. (2005). Partially interactive evolutionary artists. *New Generation Computing – Special Issue on Interactive Evolutionary Computation*, *23*(42), 143-155.

Machado, P., Romero, J., & Manaris, B. (2007). Experiments in computational aesthetics: An iterative approach to stylistic change in evolutionary art. In J. Romero & P. Machado (Eds.), *The art of artificial evolution: A handbook on evolutionary art and music* (p. 381-415). Springer Berlin Heidelberg. doi: 10.1007/978-3-540-72877-1_18

Machado, P., Romero, J., Santos, A., Cardoso, A., & Pazos, A. (2007). On the development of evolutionary artificial artists. *Computers & Graphics*, *31*(6), 818-826. doi: 10.1016/j.cag.2007.08.010

Machado, P., Vinhas, A., Correia, J., & Ekárt, A. (2015). Evolving ambiguous images. In Q. Yang & M. Wooldridge (Eds.), *Proceedings of the twenty-fourth international joint conference on artificial intelligence, IJCAI 2015, buenos aires, argentina, july 25-31, 2015* (pp. 2473–2479). AAAI Press. Retrieved from `http://ijcai.org/papers15/Abstracts/IJCAI15-350.html`

Martins, T., Correia, J., Costa, E., & Machado, P. (2015). Evotype: Evolutionary type design. In C. Johnson, A. Carballal, & J. Correia (Eds.), *Evolutionary and biologically inspired music, sound, art and design - 4th international conference, evomusart 2015, copenhagen, denmark, april 8-10, 2015, proceedings* (Vol. 9027, pp. 136–147). Springer. Retrieved from `http://dx.doi.org/10.1007/978-3-319-16498-4_13` doi: 10.1007/978-3-319-16498-4_13

McCormack, J. (2007). Facing the future: Evolutionary possibilities for human-machine creativity. In J. Romero & P. Machado (Eds.), *The art of artificial evolution: A handbook on evolutionary art and music* (p. 417-451). Springer Berlin Heidelberg. doi: 10.1007/978-3-540-72877-1_19

McCormack, J. (2017). Creative systems: a biological perspective. In T. Veale & F. A. Cardoso (Eds.), *Computational creativity: The philosophy and engineering of autonomously creative systems.* Springer International Publishing.

Nakhmani, A., & Tannenbaum, A. (2013). A new distance measure based on generalized image normalized cross-correlation for robust video tracking and image recognition. *Pattern recognition letters*, *34*(3), 315–321.

Neufeld, C., Ross, B., & Ralph, W. (2007). The evolution of artistic filters. In J. Romero & P. Machado (Eds.), *The art of artificial evolution: A handbook on evolutionary art and music* (p. 335-356). Springer Berlin Heidelberg. doi: 10.1007/978-3-540-72877-1_16

Nguyen, A. M., Yosinski, J., & Clune, J. (2015). Innovation engines: Automated creativity and improved stochastic optimization via deep learning. In S. Silva & A. I. Esparcia-Alcázar (Eds.), *Proceedings of the genetic and evolutionary computation conference, GECCO 2015, madrid, spain, july 11-15, 2015* (pp. 959–966). ACM. Retrieved from `http://doi.acm.org/10.1145/2739480.2754703` doi: 10.1145/2739480.2754703

Romero, J., & Machado, P. (Eds.). (2007). *The art of artificial evolution: A handbook on evolutionary art and music*. Springer Berlin Heidelberg. doi: 10.1007/978-3-540-72877-1

Romero, J., Machado, P., Carballal, A., & Correia, J. (2012). Computing aesthetics with image judgement systems. In J. McCormack & M. d'Inverno (Eds.), *Computers and creativity* (p. 295-322). Springer Berlin Heidelberg. Retrieved from `http://dx.doi.org/10.1007/978-3-642-31727-9_11` doi: 10.1007/978-3-642-31727-9_11

Romero, J., Machado, P., Carballal, A., & Santos, A. (2012). Using complexity estimates in aesthetic image classification. *Journal of Mathematics and the Arts*, *6*(2-3), 125-136. Retrieved from `http://www.tandfonline.com/doi/abs/10.1080/17513472.2012.679514` doi: 10.1080/17513472.2012.679514

Romero, J., Machado, P., Santos, A., & Cardoso, A. (2003). On the development of critics in evolutionary computation artists. In R. Günther et al. (Eds.), *Applications of evolutionary computing, evoworkshops 2003: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoMUSART, EvoSTOC* (Vol. 2611). Essex, UK: Springer.

Ross, B. J., Ralph, W., & Hai, Z. (2006, 16–21 July). Evolutionary image synthesis using a model of aesthetics. In G. G. Yen et al. (Eds.), *Proceedings of the 2006 ieee congress on evolutionary computation* (pp. 1087–1094). Vancouver, BC, Canada: IEEE Press. Retrieved from `http://ieeexplore.ieee.org/servlet/opac?punumber=11108`

Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, *40*(2), 99–121.

Saunders, R. (2001). *Curious design agents and artificial creativity — a synthetic approach to the study of creative behaviour* (Unpublished doctoral dissertation). University of Sydney, Department of Architectural and Design Science Faculty of Architecture, Sydney, Australia.

Secretan, J., Beato, N., D'Ambrosio, D. B., Rodriguez, A., Campbell, A., Folsom-Kovarik, J. T., & Stanley, K. O. (2011). Picbreeder: A case study in collaborative evolutionary exploration of design space. *Evolutionary Computation*, *19*(3), 373–403.

Sims, K. (1991). Artificial evolution for computer graphics. *ACM Computer Graphics*, *25*, 319-328.

Tamura, H., Mori, S., & Yamawaki, T. (1978, June). Textural features corresponding to visual perception. *Systems, Man and Cybernetics, IEEE Transactions on*, *8*(6), 460-473. doi: 10.1109/TSMC.1978.4309999

Vinhas, A. (2015). *Novelty and figurative expression-based evolutionary art* (Unpublished master's thesis). Department of Informatic Engineering, Faculty of Sciences and Technology, University of Coimbra.

Wang, L., Zhang, Y., & Feng, J. (2005). On the euclidean distance of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *27*(8), 1334–1339.