Genetic Programming and Evolvable Machines (2022) 23:557–579
https://doi.org/10.1007/s10710-022-09445-9

Experiments in evolutionary image enhancement with ELAINE

João Correia[1] · Daniel Lopes[1] · Leonardo Vieira[1] · Nereida Rodriguez-Fernandez[2] · Adrian Carballal[2] · Juan Romero[2] · Penousal Machado[1]

Received: 30 July 2021 / Revised: 26 August 2022 / Accepted: 26 September 2022 / Published online: 1 November 2022

Abstract

Image enhancement is an image processing procedure in which the image's original information is refined, for example by highlighting specific features to ease post-processing analyses by a human or machine. This procedure remains challenging since each set of images is often taken under diverse conditions which makes it hard to find an image enhancement solution that fits all conditions. State-of-the-art image enhancement pipelines apply filters that solve specific issues; therefore, it is still hard to generalise these pipelines to all types of problems encountered. We have recently introduced a Genetic Programming approach named ELAINE (EvoLutionAry Image eNhancEment) for evolving image enhancement pipelines based on pre-defined image filters. In this paper, we showcase its potential to create solutions under a real-estate marketing scenario by comparing it with a manual approach and an existing tool for automatic image enhancement. The ELAINE obtained results far exceed those obtained by manual combinations of filters and by the one-click method, in all the metrics explored. We further explore the potential of creating non-photorealistic effects by applying the evolved pipelines to different types of images. The results highlight ELAINE's potential to transform input images into either suitable real-estate images or non-photorealistic renderings, thus transforming contents and possibly enhancing its aesthetic appeal.

Keywords Image enhancement · Image processing · Computer vision · Evolutionary computation · Genetic programming

✉ João Correia
jncor@dei.uc.pt

Extended author information available on the last page of the article

1 Introduction

Digital images have become increasingly prevalent as a medium. They are present in most online activities, since they are essential elements in visual communication tasks such as attracting people's interest to further content. Each image has several attributes that condition its perception. Often, these attributes are not well balanced or optimised for the context of the image, thus affecting its visual quality. Image Enhancement (IE) is an image processing approach for improving, manually or automatically, either the overall quality of an image or the perception of a single feature. Even so, using manual IE to process large amounts of images under different conditions and constraints might often be too complex and even unfeasible. Hence, this work focuses on the application of an automatic IE approach, more specifically, for automatically enhancing images in the context of real estate marketing. Automatic IE brings significant challenges, especially when it comes to manipulating multiple aspects of the image simultaneously since individual image features are not independent of each other. Therefore, there are several types of IE techniques with different purposes and characteristics. Some of these techniques are more detailed static filters that are applied to the spatial domain, while others adapt to the image context to avoid heterogeneous results across multiple images. In this paper, we explore IE methods that are focused on improving image aesthetics.

In this paper, we present a recent Evolutionary approach named ELAINE (EvoLutionAry Image eNhancEment) .for evolving image enhancement pipelines. The system implements a Genetic Programming (GP) engine that generates image enhancement pipelines based on image processing filters with decision components. These filters alter a pipeline's output depending on the input image state and features. Fitness assignment schemes are implemented by resorting to the response of the Neural Image Assessment (NIMA) classifier [9] as an aesthetic evaluator. The NIMA classifier was tested using a dataset of various real-estate pictures of different visual quality. Furthermore, the outcome images were evaluated using image-quality assessment tools. We evolved image enhancement pipelines that can successfully improve input images according to the fitness assignment metric and other image quality metrics used for validation. However, some solutions optimised by ELAINE led to the creation of pipelines that promote non-photorealistic renderings of arguably aesthetic merit. In the context of Computational Creativity, this can be seen as a moment of serendipitous discovery [22] as, partially due to chance, an exciting and unexpected output occurred. We further explore how the pipelines that create non-photorealistic renderings affect other types of images by further showing the value of these solutions.

The work presented in this paper is an extended version of a previous paper [8] which started with the following contributions: (i) the design of an approach that creates a sequence of image filters for image enhancement; (ii) analysis of results obtained with automatic fitness assignment schemes that quantify image aesthetics; (iii) comparison between evolved filters pipeline and a baseline subset of the state-of-the-art image enhancement filters; (iv) analysis of the non-photorealistic effect detected during the experiments; and (v) exploration of the application of the non-photorealistic rendering pipelines on other types of images. In this paper, we further expand the work on the following points: (i) comparison with an external automatic tool for image enhancement; (ii) further discussion on the image transformation by these evolved pipelines, to create suitable and arguably valuable aesthetic artefacts; and (iii) deployment of a website for the evolved filters used to transform different types of images.

The remainder of this paper is organised as follows: Sect. 2 discusses related work. Section 2.1 presents commonly used image filters in the area. Section 3 describes the ELAINE approach. Section 4 lays out the experimental setup. Section 5 presents and discusses the experimental results of the effectiveness of the enhanced images in the real estate context and then image transformation factor of the evolved pipelines. Finally, Sect. 6 concludes and pinpoints future work.

2 Related work

In this paper, we focus on aspects of the image that most appeal to the human eye since we are interested in enhancing the perceived aesthetics for humans. Perceptual IE is a sub-category of IE that includes models which consider the *"Principles of Human Visual System (HVS)"* to better enhance images [12]. The contrast sensitivity function and visual masking are among the most important principles that have played important roles in existing IE methods. The *contrast sensitivity function* maps how the human eye reacts to different levels of contrast in different situations, *multi-scale and multi-orientation decomposition*, which explains how the human eye adjusts objects at various scales, orientations and distances, visual masking refers to the phenomenon that occurs when an image appears to have lower contrast or brightness when surrounded by another stronger stimulus called the mask [12]. There are multiple types of IE techniques with different purposes and characteristics. The research in this field revolves around machine learning models, computer vision pipelines for applying filters, or both. This section reviews work on image filters that are related to our approach.

As image processing and filter approaches are concerned, W. Wencheng et al. proposed an IE pipeline that aims to improve the overall brightness and contrast of low-illumination images [32]. C. Y. Wong et al. proposed another approach to bridge the problem where intensity-based approaches may produce artefacts in "over-enhance" regions and lack enrichment on colour-based features [33]. Furthermore, H. Talebi et al. [30] proposed a novel way of improving image detail and contrast by expanding on Laplacian operators of edge-aware filter kernels. Closing on classical techniques, we refer to S. Zhuo et al. in [34], who proposed a noise reduction pipeline.

Evolutionary Computation (EC) has been used in several image processing and computer vision tasks with success, which can be considered related to the work presented in this paper. Bi et al. [3, 4] used GP for image feature learning, where image-related operators, filters, and feature extraction methods are all employed as functions; to solve different image classification tasks. Colton et al. [6] developed a GP approach that automatically evolves an image filter to approximate a target one, given the original image and the filtered one. Although it does not cover the same objective of this paper, the results attained show the feasibility of the GP approach to evolve a set of filters in an unconstrained way.

Furhermore, there are a few references that directly integrate evolutionary computation approaches into IE. L. Rundo et al. proposed an evolutionary method based on genetic algorithms to improve medical imaging systems [27]. C. Munteanu also proposed an IE method that relies on evolutionary techniques to improve grey-scale images by evolving the shape of the contrast curve [21]. The work of Shan et al. [29] used an Immune Clone Algorithm (ICA) which makes the enhancement method suppress noise and increase the visibility of the underlying signal at the same time on grayscale images.

Most of the approaches aforementioned are non-modular pipelines that use fixed parameterisation and, to the best of our knowledge, are used to solve specific issues on the input images. Based on the present review, we implemented a set of image filters that provide reasonable flexibility to the evolved pipelines. These filters are further described in Sect. 3.

2.1 Image filters

Based on the review of existing related work (see Sect. 2), 7 widely-stated filters were implemented and used in this paper. These filters focus on five main aspects of IE: contrast adjustment, brightness adjustment, colour balance, noise removal and edge enhancement (also referred to as *sharpening*). In this section, we present and explain each of these filters briefly and individually.

In image processing, contrast refers to the range of intensity values available in an image. Contrast stretching is a point operation method for improving image contrast by linearly rescaling the intensity values in an image, thus increasing the contrast level. Bazeille et al.[2] discussed and reported on this filter in their approach IE for underwater image restoration. Histogram Equalisation (HE) is another method for manipulating contrast by spreading the most common intensity values to the less common ones, in order to increase the global contrast of an image. This method is a widely used, and there are multiple iterations and discussions about its results [28]. Contrast Limited Adaptive Histogram Equalization (CLAHE) is a contrast enhancement method that can adapt to different use cases [28]. It is an iteration of the Adaptive Histogram Equalization (AHE) technique, which is an improved version of the regular histogram equalisation. CLAHE improves upon AHE by clipping the maximum intensity values of each region and redistributing the clipped values uniformly throughout the histogram before applying equalization. Gamma Correction (GC) accommodates the fact that HVS perceives brightness in a non-linear way. This is achieved by scaling each pixel brightness from *[0 - 255]* to *[0 - 1]* and applying an expression to map the original values. Non-local Means Denoising (NLMD) [5] reduces existing noise in an image by replacing the value of each pixel in each channel with the average of similar pixels. Unsharp Masking (UM) is an IE technique for sharpening the edges of an image [28] by subtracting a blurred version of the original image from the original one to create a mask. The mask is then applied to the original image, to enhance its edges and details.The Simplest Colour Balance (SCB) method proposed by *N. Limare* et. al in [17], removes incorrect colour cast by scaling the histogram of each channel to the complete *0-255* range via affine transform.

3 Evolving image enhancement pipelines

Both the analysis of the related work and the results from preliminary work using the implemented filters have shown that, individually, the filters can perform well under the right conditions. However, these may struggle with versatility; input images under different conditions may require parameter adjustments. Moreover, it is clear that applying filters in sequence may produce noteworthy results and that making slight adjustments in the order of the filters may cause significant changes to the output. Furthermore, some images may require the application of one or two filters depending on the different conditions. That led us to conclude that, for automatic IE, one must find a generic pipeline of filters that is suitable for different input images. Contemplating these insights, we have developed a method for automatically generating filter pipelines. We opted for GP using a tree representation since the problem inherently can be viewed as a program, i.e. a succession of steps and decisions for deciding which filters must be applied and by what order. GP provides a representation suitable for exploring solutions in a structured and flexible way, with variation operators that are well-defined and adaptable to our problem. Thus, the developed system was later named ELAINE, standing for EvoLutionAry Image eNhancEment, as it consists of a GP engine for evolving image enhancement pipelines based on pre-defined image filters.

Making use of GP requires the definition of the primitives and terminals that will be available to the population during the evolution. In our scenario, we have sought to evolve a sequence of filter functions that generally receive, at least, an image and a numeric value as an input, (i.e. the filter's parameters). We have defined a primitive set containing all the seven classical functions implemented previously. That is a terminal set containing the input image and *ephemeral constants* ranging from -1 to 1. Each function will then map the defined range to adapt it to the desired magnitude. The range of the parameters of each function was empirically defined so acceptable results were returned. Additionally, we introduced an "*if-then-else*" function that, depending on the Boolean value of a condition, returns the output of the "*then* tree" or the "*else* tree". This function allows the same program to behave differently according to the input characteristics. This aspect means that we are using strongly typed GP when considering the use of the "*if-then-else*" function.

Since one of the primitives is a whole image, we first needed to extract comparable values from it. Thus, five "conditional functions" were added to the primitive set to extract relevant features from the image: noise, contrast, saturation, brightness and sharpness. These five features can capture characteristics of the perceived quality of the image. To extract the noise, we used the approach proposed in [14] to estimate the image's Gaussian noise efficiently. For contrast, we calculated the *Root mean square (RMS)* contrast [24], which means the standard deviation of pixel intensities. For saturation, we averaged the pixels' intensity in the *S* channel of the

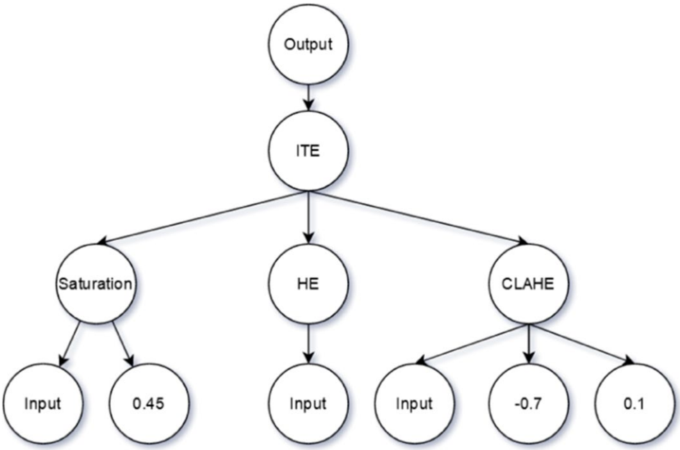


Fig. 1 Graphical example of a possible individual. The numbers represent the ephemeral constants, the *ITE* node represents the *if-then-else* primitive and the Saturation node represents the conditional function

Table 1 Summary of the GP functions, type of input and output

Acronym	Name	Input	Output
CRMS	Image contrast rms	Image, numeric	Boolean
SAT	Image saturation	Image, numeric	Boolean
PB	Perceived brightness	Image, numeric	Boolean
SHARP	Image sharpness	Image, numeric	Boolean
ITE	If then else	Boolean, image, image	Image
HE	Histogram equalize	Image	Image
CLAHE	Cilahe	Image, numeric, numeric	Image
UM	Unsharp masking	Image, numeric	Image
GC	Gamma correction	Image, numeric	Image
NLMD	Non local means denoising	Image, numeric	Image
CS	Contrast stretching	Image	Image
SCB	Simplest contrast balance	Image, numeric	Image

HSV colour system. For brightness, we used the *HSP* colour system [25], because it grants a brightness value closer to the real human perception compared with the luminance (L) channel of the *HSL* or the value (V) channel from the *HSV*. We then averaged the perceived brightness (P) channel to obtain a final value. Finally, for sharpness, we applied a Laplacian filter, calculated the variance of the output and used that as a sharpness score, as proposed in [23].

All these functions were modified to receive an image and an ephemeral constant as inputs. The ephemeral constant serves as a threshold for the condition. Figure 1 presents a graphical example of a possible individual. All the implementation was done using *DEAP* [10] for *Python* as the base evolutionary engine. Table 1 lists the different functions, inputs and outputs in this work.. The “if-then-else” condition can

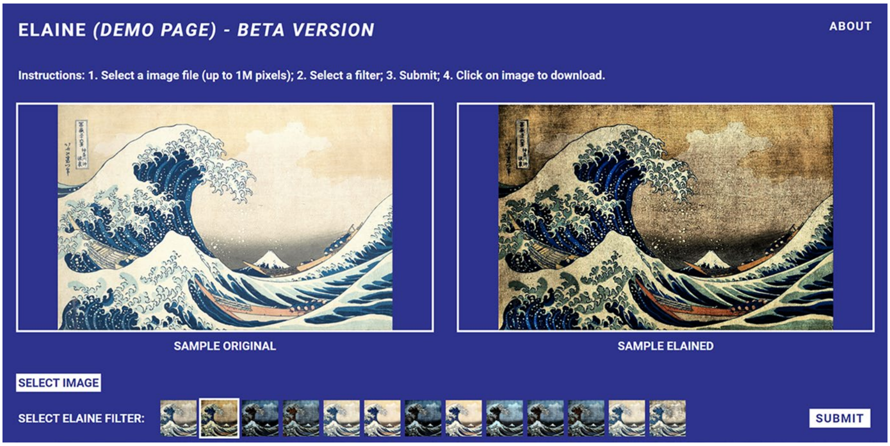


Fig. 2 ELAINE's filter tester website

only have functions that return a Boolean, as for the rest, the functions that apply filters are also strongly typed to accept the image and float arguments and the output is a filtered image.

Along with the evolutionary engine, we have created a website that allows a user to test the evolved filters with ELAINE[1] as shown in Fig. 2. The interaction goes as follows: uploading an image; selecting a filter from the pre-set evolved filters; submitting it to the system; and seeing the generated image, which can be downloaded. It is important to note that the filters are examples of outputs for the particular sample image. Since the pipelines have conditionals based on image properties, the same filter applied to different images can produce different effects.

4 Experimental setup

In this section, we present the experimental setup of our work. The GP approach creates image enhancement pipelines to generate enhanced outputs of an input image. Without a lack of generality, we deployed the approach in online real-estate marketing scenarios, where the images should be aesthetically appealing to the audience. To validate the outputs we used a set of Image Quality Assessment (IQA) tools that are presented in Sect. 4.1. 4.2 presents the test sample used in these experiments. Furthermore, in 4.3, we present the setup for our evolutionary approach.

4.1 Image quality assessment

Since our work focuses on IE, it is essential to understand how the quality of an image can be measured. L. He et al. proposed the definition of image quality in

[1] http://elaine.dei.uc.pt

three levels: fidelity, perception and aesthetics [13]. Fidelity is how well an image preserves the original information; Perception is how well an image is perceived according to every part of the HVS. An image's aesthetic is the most subjective level because it varies from person to person. As suggested by Machado et al. [19], it is possible to estimate stimuli that relate complexity to aesthetics in a controlled environment with proper testing and evaluation. Nevertheless, Van Geert et al. [11] states that although aesthetic appreciation is related to the balance of order and complexity, the literature tends to be diffuse and contradictory. In fact, aesthetics is the most difficult to measure objectively because *"aesthetics is too nonrepresentational to be characterised using mathematical models"* [13]. Aesthetic measurement is a field of that belongs to psychology as explored in [15]. From our point of view, estimation based on user evaluation or models that learn to evaluate different aesthetic properties, with a pre-defined objective and setup in mind tends to be the get go solution.

Since image quality measurements most often derive from a subjective appreciation, we must have a deterministic and automated way of qualifying the quality of an image To tackle this problem, we made use of 3 distinct *no-reference* IQA tools, where *no-reference* means that the evaluation does not depend on a target or reference image to evaluate the quality of an input image.

PhotoILike (*PHIL*) is an IQA service provided by an external company, *omitted for blind review*, for evaluating real-estate pictures. This service is an AI service that is powered by massive surveys of real estate images as briefly explained in[26]. The service receives an image and returns a value from 1 to 10, where 1 means the worst commercial appeal and 10 is the best one. The calculated score is not solely based on the image's aesthetic but also on multiple features considered relevant for real-estate marketing. For instance, the baseline score of a pool picture is much higher than the bathroom's baseline.

Blind Image Spatial Quality Evaluator [20] (*BRISQUE*), is a *no-reference* IQA tool proposed by *A. Mittal* et. al. for image enhancement contexts[16, 31]. As opposed to the previous methods, *BRISQUE* is based on a set of classical feature extraction procedures that computes a collection of 36 features per image. This tool originally outputs a value between 0 and 100, where 0 represents the best quality and 100 the worst. However, for the outputs to be in concordance with the previously presented methods, the output was mapped to a 1 to 10 range, 1 being the worst score and 10 being the best.

Neural Image Assessment (NIMA), proposed by H. Talebi and P. Milanfar [9], is a *no-reference* IQA tool based on a deep Convolutional Neural Network (CNN). The paper highlights how the same architecture, trained with different datasets leads to state-of-the-art performance predicting both technical and aesthetic scores. As the paper states, technical judgment considers noise, blur, and compression artefacts, among other image features. On the other hand, the aesthetic evaluation aims to quantify the semantic level characteristics associated with images' beauty and related emotions. Both provided models predict final scores as an average distribution of scores between 1 and 10, where 1 means the worst score and 10 the best. Both models were used during the experiments and required each input image to have a resolution of 224 by 224 pixels. The NIMA aesthetic response was used for

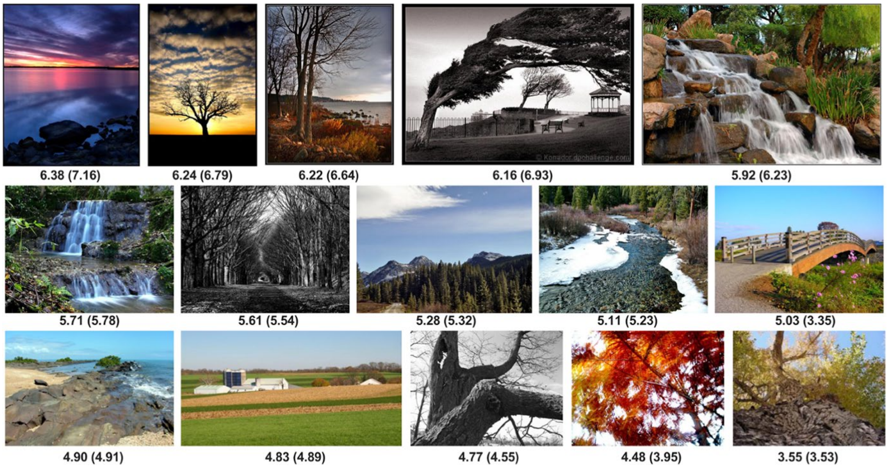
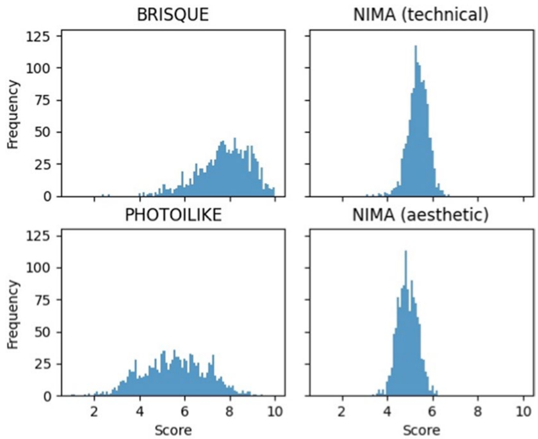


Fig. 3 A NIMA aesthetic sample of the training images and corresponding score given by the model for each image (image from [9])

Fig. 4 Graphical representation of the test dataset scores, computed by all four *no-reference* IQA tools used during the experiments



the evaluation of the individuals during the evolutionary process. Figure 3 shows typical images of the dataset used to train the NIMA aesthetic (i.e. photographs with high user ratings from the *dpchallenge*[2] website) (Fig. 4).

4.2 Test dataset

The dataset consisted of 12, 090 images obtained from real estate ads in Spain. To examine our results in an unbiased way, in an early stage, we separated 10%

[2] https://www.dpchallenge.com/

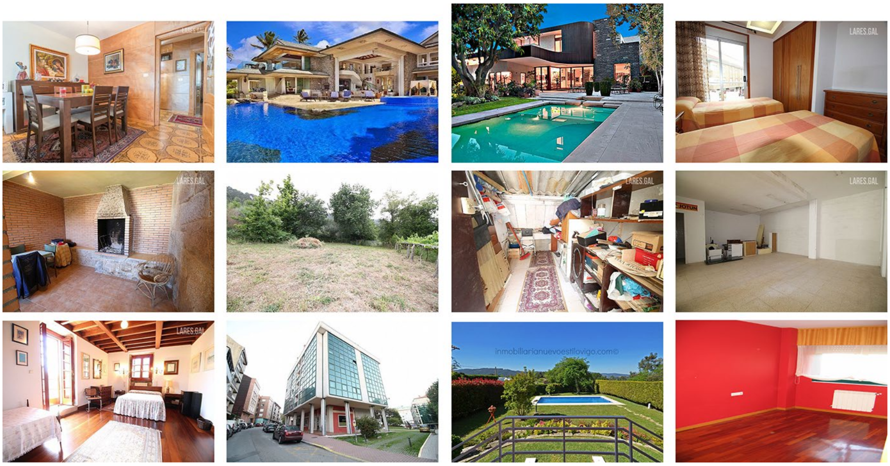


Fig. 5 Examples of images from the test dataset

Table 2 Summary of the GP configuration used during the experiments

Cross-over	One-point
Mutation	Sub Tree mutation - adds a tree with depth between 0 and 2
Selection	Tournament Size 3
Tree generation	Ramped half-and-half
Population size	80
Number of generations	150
Crossover probability	75%
Mutation probability	5%
Elite size	1
Max depth	10

of the images from the rest of the dataset and served as a training set for our approaches (i.e. to be selected along the evaluation's evolutionary process). Thus, 1, 209 images were included in the test set. Figure 5 presents the type of images that appeared, which included landscapes, house fronts, kitchens, bathrooms and garages. As a way of having a baseline for our measurements, we examined the original images of this test set, using all the IQA metrics presented in Sect. 4.1. Figure 4 includes four histograms, one histogram for each metric. It is noticeable that the *BRISQUE* tool was much more propitious to return the maximum possible score than any other metric. In contrast, the *NIMA* aesthetic model only ranged between medium scores, mostly between the 4 and 6 score range. *PHIL* evaluation shows that the test subset had a more spread-out range of scores, ranging from scores of 1 to 9.

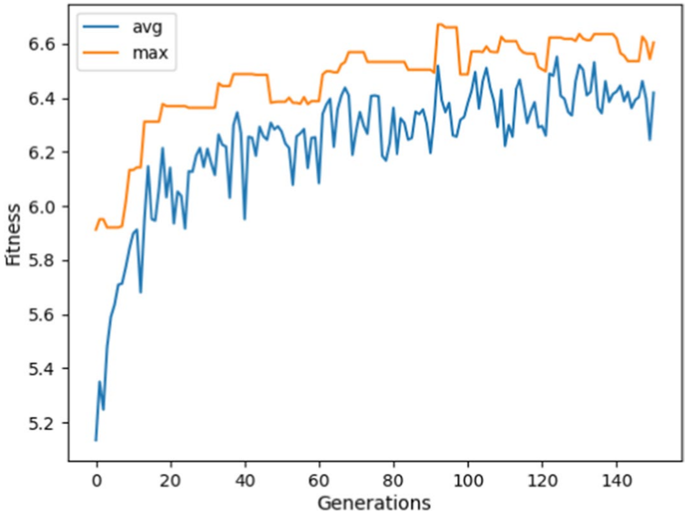


Fig. 6 Fitness evolution during 150 generations using the aesthetic model, with a depth of 10. The results are averages of 15 runs

4.3 Evolutionary setup

All the GP configuration used is summed up in Table 2, taking into account the standard GP operators and probabilities [1]. Decisions about the evolutionary setup were based on previous works that employed GP for general purpose image generation [18], generation of photorealistic face images [7] and the previous work with image filters [8]. The maximum depth was set to 10 to avoid the creation of really complex pipelines. We plan to test different parameters in future experiments.

As mentioned in Sect. 3, it was not the objective of these experiments to produce a solution that improves upon a specific image. Instead, we endeavoured in searching for a solution as generalist as possible. To achieve this, it was necessary to do a fundamental change to the typical GP evaluation. Each solution was evaluated on a set of 10 randomly selected images from the training subsetwas considered the fitness of the individual. In addition, to further prevent overfitting to a specific group of images, a new set of 10 images was selected in each generation. For this reason, it was expected a significant variation in fitness from one generation to another. In all performed experiments, the individual with the overall highest fitness was selected as the test subject for validation.

Defining a fitness function is one of the most crucial steps in building an evolutionary approach. In our case, we aimed to have a fitness function that evaluated each individual based on its visual aesthetic value. We selected the NIMA aesthetic metric as it produced results closer to the *state-of-the-art* on visual quality evaluation[9] (Fig. 6).

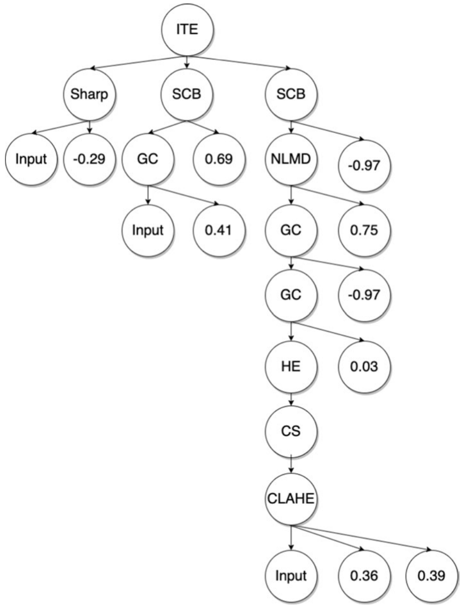


Fig. 7 Graphical representation of the best evolved individual. Input represents the original image, constant values are ephemeral constants, and filters are represented by their respective acronyms

Fig. 8 Execution of the best individual pipeline on the original (input) image on the leftmost side. The image undergoes transformation along the middle part of the evolved pipeline, which in this case means that a GC and SBC were applied to the image

5 Experiments in image enhancement

The experiments section is divided into three subsections: analysis of the evolved pipeline; experimental results in real estate image problem and; experiments and analysis with non-photorealistic effects that transform input images. Section 5.1 shows the evolved pipeline, together with the manual created one. In Sect. 5.2 we cover the experimental results in the context of image enhancement for the real estate image problem, from an analytical and image suitability perspective. Then, in Sect. 5.3, we analyse some of the best-generated pipelines, the different effects on different types of images and how some pipelines transform totally or partially the contents of the photos, in some cases producing a non-photorealistic effect.

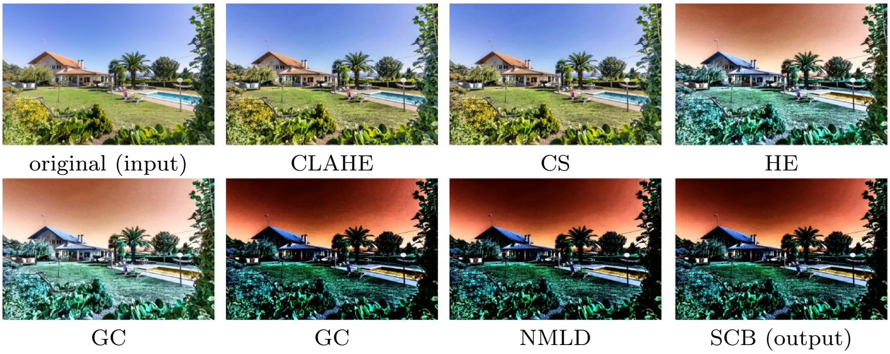


Fig. 9 Execution of the best individual pipeline on the original (input) image on the left-most side. The image has undergone the execution via the rightmost branch of the tree. The transformations order occurs from the top left to the bottom right in standard reading order

5.1 Evolved pipeline

The manual pipeline employs the following filters by order of application: $Contrast\ Balance\ (CB) -> CLAHE -> UnsharpMasking\ (UM) -> Non - localmeansdenoising\ (NLMD) -> ContrastStretching(CS)$.

On the other hand, the best individual obtained by ELAINE, described using the filter's acronyms from Sect. 2.1, is depicted in Fig. 7.

For the best individual evolved, an *if* which enabled for a different output depending on the condition of the input image x. In Fig. 8 the input image passed the condition $Sharp(x, -0.29)$ yielding the presented sequence of image filters. In contrast, Fig. 9 did not pass the condition, resulting in the application of the other side of the tree. As shown in these results, these examples demonstrate different outputs of the same individual. Further, the best individual transforms the image creating a non-photorealistic effect on the input image, arguably promoting aesthetics but in a way that, for the problem that we are attempting to solve, is not a suitable solution.

5.2 Experimental results

In this subsection, we explore the experiments' results in a sub-dataset of real estate images to enhance their quality. Starting with the evolutionary approach, Fig. 6 shows that we can maximise the fitness function and the iterations. It also demonstrates that the set of evaluation images changes at each generation, and the maximum value can oscillate even with elitism. Based on these results, the process creates pipeline solutions that enhance the aesthetic score of images according to the NIMA aesthetic output.

To further evaluate the pipeline solutions, we created a validation process where the best pipeline is applied to the test set, i.e. images not used in the evolutionary process. In this way, we can analyse the generalisation of the evolved pipelines. To establish a baseline approach, we conducted the same validation with (i) a manually

Table 3 Average (μ) and standard deviation (σ) of the original images from the test dataset using 4 *no-reference* metrics. All the metrics values range from 1 to 10 where 1 means the lowest quality and 10 highest quality

	NIMA A.	NIMA T.	BRISQUE	PHIL
Original—μ	4.91	5.36	7.77	5.56
Original—σ	0.45	0.43	1.15	1.41

Table 4 Average (μ) and standard deviation (σ) of the improvements compared to the input images. Original images, images from the dataset that serve as input to the different approaches

	NIMA A.	NIMA T.	BRISQUE	PHIL
Manual—μ	0.43	– 0.06	– 0.08	0.25
Manual—σ	0.31	0.26	1.86	0.64
One-Click—μ	0.16	**– 0.02**	– 1.19	0.39
One-Click—σ	0.25	0.23	0.69	0.69
Evolutionary—μ	**1.30**	– 0.55	**0.40**	**0.47**
Evolutionary—σ	0.45	0.43	1.10	0.96

Bold mark the best values in for each metric (column)

created pipeline, arranged based by one of the authors with huge design experience, with the default parameterisation of each filter, and (ii) One-Click, an external IE software.

The original scores according to the IQA metrics are presented in Table 3 and Fig. 4. The results of the manual pipeline, One-Click and evolved pipeline are presented in Table 4 and Fig. 10. For benefit of readability, NIMA models are abbreviated to their respective initials. A negative improvement score means degradation of the quality according to the respective metric.

The results reveal that the manual pipeline resulted in small alterations across all IQA scores, with the NIMA aesthetic response showing some improvements. On average, the manually defined pipeline improved the test dataset using NIMA and PHIL, and the other two metrics indicate that it worsened the images.

From the evaluation of the best-evolutionary pipeline, all the metrics showed greater enhancements when compared with both the manual pipeline and the external tool One-click. An exception is made for the NIMA technical, as some images tended to be worse than the original and the manual pipeline, averaging a −0.55 difference from the original and −0.49 from the manual pipeline.

One-click leveraged well the response of NIMA, yet the performance of BRISQUE was far worse than the manual approach. Such a difference is far more noticeable than the difference from the Evolutionary approach towards the NIMA technical. Overall, considering the initial distribution of the scores, we had significant improvements to the images' aesthetic component when using the response from the aesthetic NIMA model to assign fitness. Based on these results, it can be concluded that the best- evolved pipeline has better results than the manual pipeline and the external tool One-Click.

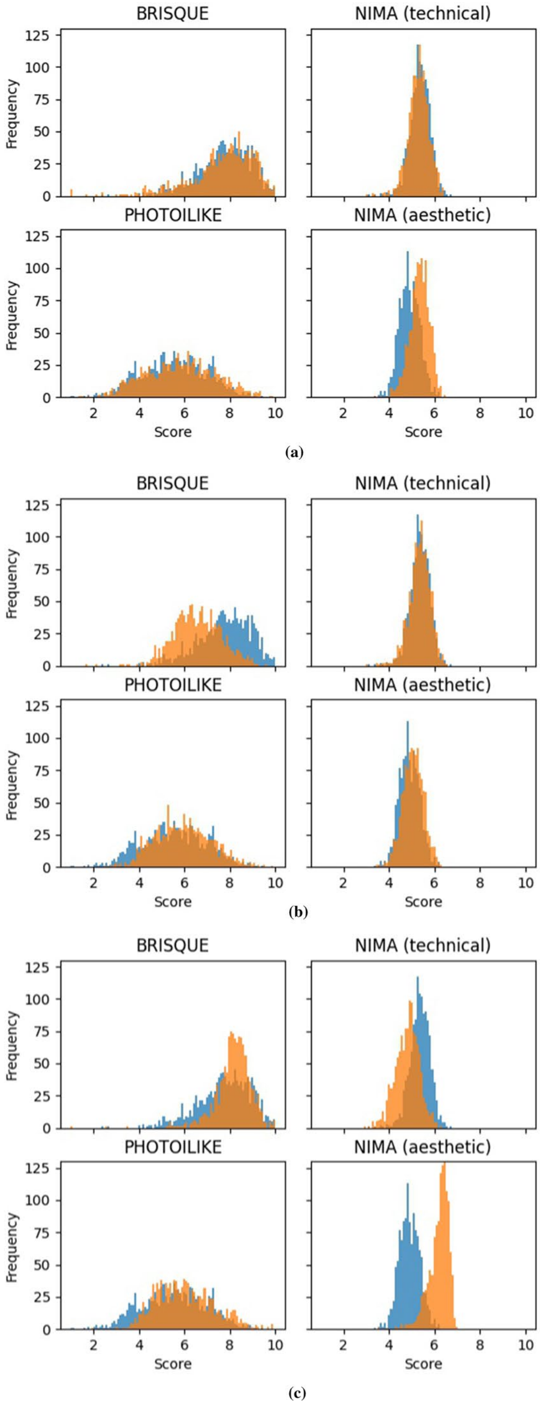


Fig. 10 Graphical comparison between: **a** The original test dataset (blue) and the results from the list of classical functions manually selected (orange); **b** Graphical comparison between the original test dataset (blue) and the results from the *One-Click* (orange) and; **c** Graphical comparison between the original test dataset (blue) and the results from using the best evolved solution using *NIMA* aesthetic as fitness, computed by all four *no-reference* IQA tools: NIMA aesthetic, PhotoIlike, NIMA technical and BRISQUE

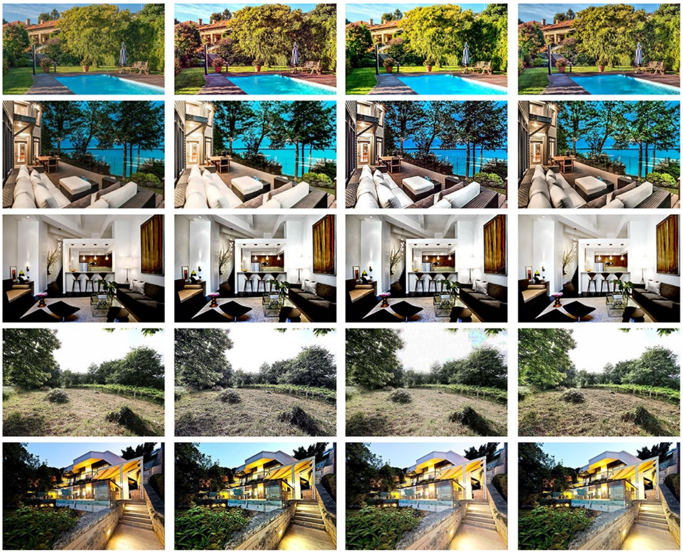


Fig. 11 Examples of images from the test set (1st column), processed using the manual pipeline (2nd column), one-click external tool (3rd column) and the evolved pipeline (4th column)

Figure 11 shows real estate images processed by all proposed pipelines to compare the result of original images, the manual pipeline, the One-Click tool and evolved pipeline. Compared to the manual pipeline and input images, all the presented *NIMA* images regarded metric improvements.

All the metric values range from 1 to 10 where 1 means the lowest quality and 10 the highest quality. The improvement is the subtraction of the original score from the resulting one.

In order to perform an external evaluation of the output of the system, we conducted a user study using Amazon Mechanical Turk, a tool that allows us to make online surveys of anonymous users. We evaluated two versions (original and improved) of 30 images. Each image was evaluated by 100 persons with no limitation of country, age or other considerations. Evaluations with less than .7 of correlation with the average score were deleted to avoid random answers. The "improved" image was selected 70% of the time. Figure 12 presents some of the images employed in the survey, along with their average scores.

As a negative point, some of the images were over processed. The *NIMA* aesthetic model produced fascinating results. In some details, these might resemble paintings rather than photorealistic pictures. Regarding the evaluation performed by the implemented models, it is noticeable that higher scores tended to be associated with over-edited and saturated images. We can extrapolate that improvements of the same magnitude as those in this test also lead to the extreme adulteration of the original images, even if these are considered good enough for the aesthetic model and *PHIL*.

For real estate marketing, based on these insights and results, we are considering incorporating the NIMA technical into the fitness function, along with some control

Fig. 12 Samples from the survey. Images on the left are the original images and on the right are the images after applying the best evolved pipeline. The score is bellow each image and ranges from 0 to 10

mechanism to prevent the output images from suffering many alterations compared with the input ones. The set of functions can also be more flexible and even improved with other filters or conditional operations. These lines of experimentation and research are already being pursued.

5.3 Image transformation

Besides the relevant results, we obtained pipelines where non-photorealistic renderings occurred. The results shown in Fig. 13 were unexpected but, at the same time, a pleasant surprise, indicating a moment of computational serendipity [22]. The overall goal was to create a filter pipeline that maximises aesthetics according to NIMA. As shown, we can maximise the response, and the results indicate that the images can be enhanced across the IQA tools that take aesthetics into account. More importantly, we are improving the PHIL metric, which estimates aesthetic value from a real-estate marketing perspective. Thus, the system tends to further evolve solutions that maximise the response of *PHIL*, which is the more closely related metric to the problem we are trying to solve. Moreover, a fascinating fact is that the NIMA aesthetic was not trained with paintings; even so, it

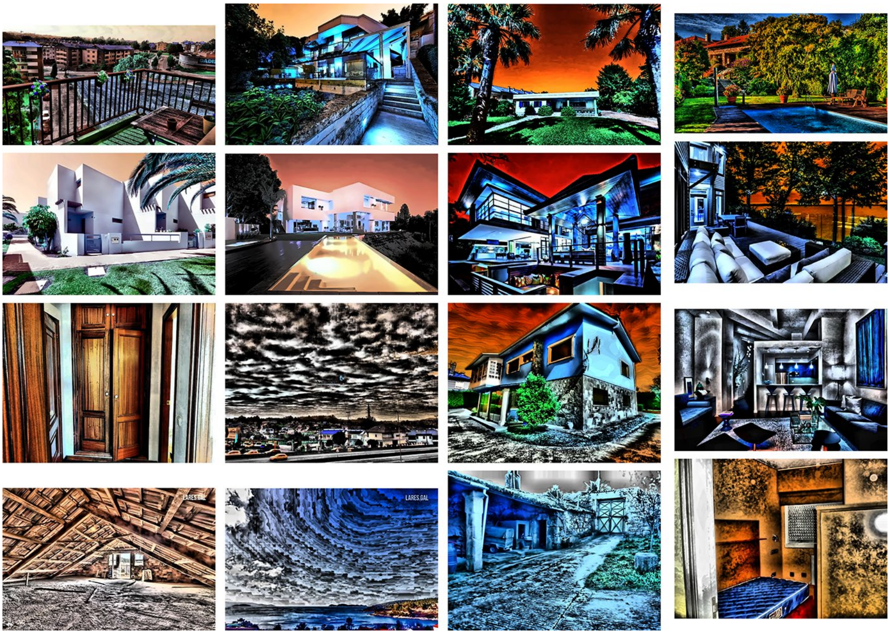


Fig. 13 Examples of images with non-photorealistic rendering effects

tended to score higher for the more non-realistic images, which tended to be more like paintings. This may be because the NIMA metric is based on DPChallenge images. DPChallenge images with positive scores tended to be highly processed and have high doses of saturation, which explains why exploring the metric to the limit generates images with extreme processing. The same applies to the PHIL metric. Artificially enhanced images were not included in the base dataset of the PHIL metric used, but only raw images of real advertisements. Subsequently, processed images were included in the training datasets of the PHIL metric to avoid this problem.

Based on these results, we tested evolved pipelines in another type of imagery to evaluate the effects and assess whether the results would render aesthetically pleasing images. We showcase the application of these same pipelines to different types of images: landscape photographic images (1st and 2nd rows of Fig. 14); abstract images (3rd and 4th row of Fig. 14); paintings (5th and 6th row of Fig. 14); and poster designs (7th and 8th row of Fig. 14). As shown in Figure, the pipelines are generic enough to alter the different types of images—computer images, paintings and raw photographs.

From a subjective standpoint, our approach might also have the potential for creating pipelines that lead to non-photorealistic aesthetic images, which can be useful, for example, for art and design purposes. It is important to note that, for the same filter, depending on the input image, different results can occur. Aesthetics-wise, compared to real-estate, art and design images might be even more

Fig. 14 Samples of images created with a selection of different evolved non-photorealistic pipelines. The original image is the first one on the left

subjective and thus difficult to judge automatically. Hence, such non-photorealistic pipelines might or not interest given creative thinkers.

6 Conclusions

In this work, we present an approach for automatic image enhancement using GP called ELAINE. The approach is instantiated in real estate marketing, to improve images of real estate which requires modular and adaptable solutions due to the diversity of images. We propose an approach that relies on a set of 7 filters from the literature that are linked to image enhancement and image quality assessment. The context of the problem indicates that the performed image enhancement should aim toward aesthetically pleasing images. We explored that characteristic by resorting to the aesthetic metric NIMA to evaluate individuals. After evolving pipeline solutions, we tested the best solution in a subset of test images, which were compared with the initial results and a manual pipeline, used as a baseline, as well as external image enhancement tools (One Click). The results show that the system can create filter application pipelines that improve the quality of the images in all the quality assessment metrics tested. The technical IQA tool metric was the only metric that suffered a negative effect. We argue that a mechanism to prevent the image from suffering many alterations could mitigate this effect.

During experimentation, as computational serendipity, the resulting images turned out to be non-photorealistic renderings. The system maximised the objective function and the validation metrics response while creating effects that manipulated the images until these became almost paintings. We moved to explore the application in another type of imagery and analysed the effects, showing the potential of this approach to create non-photorealistic rendering and transforming effects.

In future work, we plan to expand and alter the set of functions available in ELAINE, fine-tune evolutionary parameters, experiment with fitness assignment alternatives and change the evolution dataset. We also plan to incorporate machine learning approaches in the pipeline, to be used as filters for the input images. Furthermore, in the context of the problem at hand, the results suggest that we should not exaggerate and alter the images' features too much. We plan on doing a set of experiments using similarity metrics to control the pipelines and prevent heavy alterations on the inputs.

Furthermore, expanding the scope of the training images to create a more general-purpose image enhancer is also being pursued (i.e. using images not only from real estate). Lastly, more user-testing must be done to retrieve further insights into the visual results achieved using the developed pipelines.

Acknowledgements This work is funded by the Foundation for Science and Technology (FCT), I.P./MCTES through national funds (PIDDAC), within the scope of CISUC R &D Unit—UIDB /00326/2020 or project code UIDP/00326/2020 and under the grant SFRH/BD/ 143553/2019. This work is also funded by the INDITEX-UDC Program for predoctoral research stays through the Collaboration Agreement between the UDC and INDITEX for the internationalization of doctoral studies. Juan Romero received funding from Spanish Ministry of Universities for mobility stays of professors and researchers in foreign centres of higher education and research. Juan Romero and Adrian Carballal received funding with reference PID2020-118362RB-I00, from the State Program of R+D+i Oriented to the Challenges of the Society of the Spanish Ministry of Science, Innovation and Universities.

References

1. W. Banzhaf, F.D. Francone, R.E. Keller, P. Nordin, *Genetic programming: an introduction: on the automatic evolution of computer programs and its applications* (Morgan Kaufmann Publishers Inc., San Francisco, 1998)
2. S. Bazeille, I. Quidu, L. Jaulin, J.P. Malkasse, Automatic underwater image pre-processing. Proceedings of CMM'06 (2006)
3. Y. Bi, B. Xue, M. Zhang, Genetic programming with image-related operators and a flexible program structure for feature learning in image classification. IEEE Trans. Evol. Comput. **25**(1), 87–101 (2021). https://doi.org/10.1109/TEVC.2020.3002229
4. Y. Bi, B. Xue, M. Zhang, Genetic programming with image-related operators and a flexible program structure for feature learning in image classification. IEEE Trans. Evol. Comput. **25**(1), 87–101 (2021). https://doi.org/10.1109/TEVC.2020.3002229
5. A. Buades, B. Coll, J.M. Morel, Non-local means denoising. Image Process. Line **1**, 208–212 (2011). https://doi.org/10.5201/ipol.2011.bcm_nlm
6. S. Colton, P. Torres, Evolving approximate image filters. In: M. Giacobini, A. Brabazon, S. Cagnoni, G.A.D. Caro, A. Ekárt, A. Esparcia-Alcázar, M. Farooq, A. Fink, P. Machado, J. McCormack, M. O'Neill, F. Neri, M. Preuss, F. Rothlauf, E. Tarantino, S. Yang (eds.) Applications of Evolutionary Computing, EvoWorkshops 2009: EvoCOMNET, EvoENVIRONMENT, EvoFIN, EvoGAMES, EvoHOT, EvoIASP, EvoINTERACTION, EvoMUSART, EvoNUM, EvoSTOC, EvoTRANSLOG, Tübingen, Germany, April 15-17, 2009. Proceedings, Lecture Notes in Computer Science, vol. 5484, pp. 467–477. Springer (2009). https://doi.org/10.1007/978-3-642-01129-0_53
7. J. Correia, T. Martins, P. Machado, Evolutionary Data Augmentation in Deep Face Detection. In: GECCO 2019—Proceedings of the 2019 Genetic and Evolutionary Computation Conference. Prague, Czech Republic (2019)
8. J. Correia, L. Vieira, N. Rodriguez-Fernandez, J. Romero, P. Machado, Evolving image enhancement pipelines. In: J. Romero, T. Martins, N. Rodríguez-Fernández (eds.) Artificial Intelligence in Music, Sound, Art and Design—10th International Conference, EvoMUSART 2021, Held as Part of EvoStar 2021, Virtual Event, April 7-9, 2021, Proceedings, Lecture Notes in Computer Science, vol. 12693, pp. 82–97. Springer (2021). https://doi.org/10.1007/978-3-030-72914-1_6
9. H.T. Esfandarani, P. Milanfar, NIMA: neural image assessment. CoRR http://arxiv.org/abs/1709.05424 (2017)
10. F.A. Fortin, F.M. De Rainville, M.A. Gardner, M. Parizeau, C. Gagné, DEAP: Evolutionary algorithms made easy. J. Mach. Learn. Res. **13**, 2171–2175 (2012)
11. E.V. Geert, J. Wagemans, Order, complexity, and aesthetic appreciation. Psych. Aesthet., Creat. Arts **14**, 135–154 (2020)
12. D. Ghadiyaram, T. Goodall, L.K. Choi, A.C. Bovik, Perceptual image enhancement, in *Encyclopaedia Image Processing*. ed. by P.A. Laplante (CRC Press, Boca Raton, 2018)
13. L. He, F. Gao, W. Hou, L. Hao, Objective image quality assessment: a survey. Int. J. Computer Math. **91**(11), 2374–2388 (2014). https://doi.org/10.1080/00207160.2013.816415
14. J. Immerkær, Fast noise variance estimation. Comput. Vis. Image Underst. **64**(2), 300–302 (1996). https://doi.org/10.1006/cviu.1996.0060
15. C. Johnson, J. McCormack, I. Santos, J. Romero, Understanding aesthetics and fitness measures in evolutionary art systems. Complexity **2019**, 1–14 (2019). https://doi.org/10.1155/2019/3495962
16. J. Lim, M. Heo, C. Lee, C.S. Kim, Contrast enhancement of noisy low-light images based on structure-texture-noise decomposition. J. Visual Commun. Image Represent. **45**, 107–121 (2017). https://doi.org/10.1016/j.jvcir.2017.02.016. http://www.sciencedirect.com/science/article/pii/S1047320317300603
17. N. Limare, J.L. Lisani, J.M. Morel, A.B. Petro, C. Sbert, Simplest color balance. Image Process. On Line (2011). https://doi.org/10.5201/ipol.2011.llmps-scb

18. P. Machado, A. Cardoso, All the truth about nevar. Appl. Intell. **16**(2), 101–118 (2002). https://doi.org/10.1023/A:1013662402341
19. P. Machado, J. Romero, M. Nadal, A. Santos, J. Correia, A. Carballal, Computerized measures of visual complexity. Acta Psychologica **160**, 43–57 (2015). https://doi.org/10.1016/j.actpsy.2015.06.005. https://www.sciencedirect.com/science/article/pii/S0001691815300160
20. A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. **21**(12), 4695–4708 (2012)
21. C. Munteanu, A. Rosa, Evolutionary image enhancement with user behaviour modeling. ACM SIGAPP Appl. Comput. Rev. **9**, 87 (2000). https://doi.org/10.1145/372202.372352
22. A. Pease, S. Colton, R. Ramezani, J. Charnley, K. Reed, A discussion on serendipity in creative systems. In: M. Maher, T. Veale, R. Saunders, O. Bown (eds.), Proceedings of the 4th International Conference on Computational Creativity, ICCC 2013, pp. 64–71. University of Sydney, Faculty of Architecture, Design and Planning (2013). http://www.computationalcreativity.net/iccc2013/. Fourth International Conference on Computational Creativity, ICCC 2013 ; Conference date: 12-06-2013 Through 14-06-2013
23. J.L. Pech-Pacheco, G. Cristobal, J. Chamorro-Martinez, J. Fernandez-Valdivia, Diatom autofocusing in brightfield microscopy: a comparative study. In: Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, vol. 3, pp. 314–317 (2000)
24. E. Peli, Contrast in complex images. J. Opt. Soc. Am. A **7**(10), 2032–2040 (1990). https://doi.org/10.1364/JOSAA.7.002032
25. D. Rex Finley, Hsp color model—alternative to hsv (hsb) and hsl (2006). http://alienryderflex.com/hsp.html
26. N. Rodriguez-Fernandez, S. Alvarez-Gonzalez, I. Santos, A. Torrente-Patiño, A. Carballal, J. Romero, Validation of an aesthetic assessment system for commercial tasks. Entropy **24**(1), (2022). https://doi.org/10.3390/e24010103. https://www.mdpi.com/1099-4300/24/1/103
27. L. Rundo, A. Tangherloni, M. Nobile, C. Militello, D. Besozzi, G. Mauri, P. Cazzaniga, Medga: a novel evolutionary method for image enhancement in medical imaging systems. Expert Syst. Appl. **119**, 87 (2018). https://doi.org/10.1016/j.eswa.2018.11.013
28. J.C. Russ, *Image processing handbook*, 4th edn. (CRC Press Inc., USA, 2002)
29. T. Shan, S. Wang, X. Zhang, L. Jiao, Automatic image enhancement driven by evolution based on ridgelet frame in the presence of noise, in *Applications of evolutionary computing*. ed. by F. Rothlauf, J. Branke, S. Cagnoni, D.W. Corne, R. Drechsler, Y. Jin, P. Machado, E. Marchiori, J. Romero, G.D. Smith, G. Squillero (Springer Berlin Heidelberg, Berlin, 2005), pp.304–313
30. H. Talebi, P. Milanfar, Fast multi-layer laplacian enhancement. IEEE Trans. Comput. Imag. (2016). https://doi.org/10.1109/TCI.2016.2607142
31. G. Wang, L. Li, Q. Li, K. Gu, Z. Lu, J. Qian, Perceptual evaluation of single-image super-resolution reconstruction. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 3145–3149 (2017)
32. W. Wang, Z. Chen, X. Yuan, X. Wu, Adaptive image enhancement method for correcting low-illumination images. Inf. Sci. **496**, 25–41 (2019). https://doi.org/10.1016/j.ins.2019.05.015
33. C.Y. Wong, G. Jiang, M.A. Rahman, S. Liu, S.C.F. Lin, N. Kwok, H. Shi, Y.H. Yu, T. Wu, Histogram equalization and optimal profile compression based approach for colour image enhancement. J. Visual Commun. Image Represent. **38**, 802–813 (2016). https://doi.org/10.1016/j.jvcir.2016.04.019
34. S. Zhuo, X. Zhang, X. Miao, T. Sim, Enhancing low light images using near infrared flash images. Proceedings—International Conference on Image Processing, ICIP pp. 2537–2540 (2010). https://doi.org/10.1109/ICIP.2010.5652900

Authors and Affiliations

João Correia[1] · **Daniel Lopes**[1] · **Leonardo Vieira**[1] · **Nereida Rodriguez-Fernandez**[2] · **Adrian Carballal**[2] · **Juan Romero**[2] · **Penousal Machado**[1]

Daniel Lopes
dfl@dei.uc.pt

Leonardo Vieira
lmavieira@student.dei.uc.pt

Nereida Rodriguez-Fernandez
nereida.rodriguezf@udc.es

Adrian Carballal
adrian.carballal@udc.es

Juan Romero
jj@udc.es

Penousal Machado
machado@dei.uc.pt

1 Department of Informatics Engineering, Centre for Informatics and Systems of the University of Coimbra, University of Coimbra, Coimbra, Portugal

2 CITIC—Research Center of Information and Communication Technologies, University of A Coruña, Coruña, Spain