

Collaborative Interactive Evolution of Art in the Latent Space of Deep Generative Models

Ole Hall and Anil Yaman

Vrije Universiteit Amsterdam, 1081 HV Amsterdam, Netherlands
ole.moritz.hall@gmail.com, a.yaman@vu.nl

Abstract. Generative Adversarial Networks (GANs) have shown great success in generating high quality images and are thus used as one of the main approaches to generate art images. However, usually the image generation process involves sampling from the latent space of the learned art representations, allowing little control over the output. In this work, we first employ GANs that are trained to produce creative images using an architecture known as Creative Adversarial Networks (CANs), then, we employ an evolutionary approach to navigate within the latent space of the models to discover images. We use automatic aesthetic and collaborative interactive human evaluation metrics to assess the generated images. In the human interactive evaluation case, we propose a collaborative evaluation based on the assessments of several participants. Furthermore, we also experiment with an intelligent mutation operator that aims to improve the quality of the images through local search based on an aesthetic measure. We evaluate the effectiveness of this approach by comparing the results produced by the automatic and collaborative interactive evolution. The results show that the proposed approach can generate highly attractive art images when the evolution is guided by collaborative human feedback.

Keywords: Generative Adversarial Networks · Latent Variable Evolution · Interactive Evolutionary Computation · Collaborative Art

1 Introduction

Artificial intelligence (AI) based approaches to art generation have increased their popularity and received a great deal of attention due to available tools such as Artbreeder [34] and Nvidia Canvas [24]. A debate about the value of such art has developed at the latest since the auction of an AI generated artwork at the renowned auction house Christie’s [7]. Frequent critics doubt the creativity and novelty that can be generated by AI, while other voices postulate AI mainly as a potent tool of modern artists [6].

Deep generative models such as Generative Adversarial Networks (GANs) [14] play a major role in AI based art generation approaches. The basic idea of a GAN is to generate images that cannot be distinguished from real images. With this setup, however, doubts have been raised if artefacts generated in this way are

actually creative or merely attempting to emulate the training material [10]. On the other hand, the generation process offers a creative space, as the fake images can be generated by sampling from the latent space of learned art representations, providing infinite possibilities of potential images [6].

Due to this feature, GANs open up many possibilities for human-AI interaction in co-creative processes [6,15]. For example, they can inspire designers [30,43], and artists may spend hours generating different random images to find attractive or inspiring outcomes [35]. A more promising approach to discover art images is to apply (meta-)heuristic search algorithms such as evolutionary computing (EC) [9]. Evolutionary computing has been employed to discover the design space in GANs in different fields of application (e.g. [4,28,40]), but rarely in the field of art generation [29,34].

In this paper, we employ EC to navigate within the latent space of GANs trained on art images. Firstly, we use a specific GAN architecture, the Creative Adversarial Network (CAN) [10], to introduce novel art images. CANs can achieve this because the images generated imitate the real art distribution while deviating from established art styles through a modified loss function that penalises simple categorisation into an existing art style. Then, the vector representations of latent space variables are used to encode individuals in EC. We employ evolutionary operators such as crossover and mutation to generate new individuals from the existing ones. To determine the quality of art images (a.k.a. fitness in EC), we use two metrics: (1) automatic aesthetic, and (2) collaborative interactive human evaluation.

Automatic aesthetic evaluation is based on Neural Image Assessment (NIMA) [38], a Convolutional Neural Network trained on an annotated dataset for aesthetic visual analysis (AVA) [23]. We also used the automatic aesthetic evaluation metric as an intelligent mutation operator that aims to perform a local search [29] to improve the quality of the images and accelerate the evolutionary process similar to the approaches used in memetic algorithms [9].

Collaborative interactive human evaluation is based on the idea of interactive evolutionary computation (IEC) [37] where human evaluations replace the fitness evaluation step in EC. The IEC is particularly suitable in cases such as art generation where the measure of the quality is subjective [27]. However, in contrast to the classic IEC setup, here, we consider collaborative evaluations from several participants to account for their subjectivity.

To prevent the generation of very similar images especially within the same generation, we introduce diversity preservation mechanisms. If two similar images are detected, one is replaced by a new randomly generated image. This also intends to accelerate the exploration of latent space, as well as to avoid user fatigue from evaluating many similar images [9].

Our results show that this methodology is generally suitable for exploring the latent space of possible art images in a GAN, and it is able to create increasingly attractive images. It was shown that both the automatic as well as the collaborative evolution achieved an increasing fitness over the evolutionary process. In addition, we tested whether the local search leads to improvements beyond

a random level in the eyes of human participants, but this was not confirmed. Finally, we investigated whether the results obtained from both the automatic aesthetic and collaborative interactive evolution are in fact perceived as more attractive than randomly generated images. The findings indicate that human guidance is crucial for the evolution of art in order to achieve images that are perceived as more attractive.

2 Related Work

2.1 Generative Adversarial Networks

Introduced by Goodfellow et al. [14], GANs aim to generate artefacts that are indistinguishable from real images. For this purpose, two models are in a competing relationship, allowing for an unsupervised learning approach: the generator is trained to generate fake images that are indistinguishable from real ones, while the discriminator is trained to make the distinction by comparing the generated images with real ones. The training is therefore comparable to a min-max game for two networks. Both the generator and discriminator are designed as deep neural networks. Through mutual feedback, they constantly improve each other, eventually leading the generator to generate artefacts that are difficult to distinguish from real ones. These artefacts can be created from underlying latent variables by random sampling.

Despite impressive visual results, training GANs is considered difficult and partly unstable, and they were initially only able to generate low resolutions [18]. Therefore, many extensions and variants have been developed since then ([17], for an overview). A structural improvement for working with images was provided by Deep Convolutional Generative Adversarial Networks (DCGANs) [25], which form the basis of most of today's variants [17].

Another interesting extension is the CAN [10], which is building up on the DCGAN. Elgammal et al. argue that the classical architecture is merely emulating the training material, but that the reference to and influence by other artists' works is natural. They propose a new architecture, referring to a psychology-based theory from Martindale [21] that says that creative processes always try to evoke arousal, which can for example be achieved through stylistic breaks. At the same time, however, creative work does not want to let this arousal become too great in order to avoid negative reactions. In their architecture, the goal is to generate creative art by finding a balance between mimicking the real art distribution, while deviating from established styles. This should increase the novelty and creativity of the generated images, which was also supported by different experiments with human participants [10]. With regard to a collaborative interactive evolution of art images, this could be advantageous in that the participants are less likely to be guided in their evaluations by well-known representatives of different art styles. Therefore, the architecture of the CAN is used for the generative model in this work.

2.2 Latent Space Exploration

While theoretically every latent vector in a well-trained GAN leads to an output considered by the discriminator as an element of the target group, there is still a great variation between them. This diversity of possible outputs motivates the interest in exploring the latent space and opens up a creative space for human-AI interaction in co-creative processes. In their framework, Grabe et al. [15] make a distinction between four different interaction patterns in human-AI interaction with GANs such as curation, exploration, conditioning, and evolution.

By curation, they mean the most straightforward method, according to which control over the output is achieved through the selection of training data and the subsequent manual selection of generated artefacts. However, brute-forcing new artefacts and manual cherry-picking provides only a low level of control and allows only for random search [28], even if it is actually used this way in the everyday life of some artists [35]. At the same time, the model choice can also be understood as part of curation, for example the choice of the CAN architecture.

Although all four types could be subsumed under the umbrella term exploration, Grabe et al. use it to describe the iterative adaptation of generated artefacts, for example by moving a slider. One possibility is the interpolation between different images, for example CREA.blender [13] allows the exploration of space between different images. Another possibility is the conscious adaptation along either the latent vectors directly or semantic attributes represented in them. Examples of this are found in the alteration of facial features [33,42] as well as in adding, deleting, or altering aspects in images [2,3].

Conditioning refers to methods that allow people to determine desired aspects in advance, often referred to as conditional GAN. Examples are drawings or contour images that predefine the later output, which was used in the domain of fashion design [41] and in the creation of landscape paintings [24]. In addition, instructions can also be given in text form [26,43].

Latent Variable Evolution Another form of control is offered by evolutionary computing. Grabe et al. [15] refer under evolution only to explicitly interactive setups since in IEC, human feedback directly guides the process of computer-driven exploration. However, already the definition of a target of the evolutionary process as well as the specification of an automatically evaluating fitness function provide a certain degree of control.

Evolutionary computing have already been used several times in combination with GANs, both with automatic and interactive fitness evaluation metrics. The terms Latent Variable Evolution [5,31,40] and, in an interactive setup, Deep Interactive Evolution [4] were coined. The first steps in this direction were taken by Bontrager et al. [5], who applied an evolutionary search to find a latent vector in a GAN trained on real fingerprints that matches as many subjects as possible. Bontrager et al. [4] were also the first to propose an interactive setup for developing faces and shoes towards a target image, thereby demonstrating that GANs can work as a compact and robust genotype-to-phenotype mapping. Again in the field of fashion design, an interactive setup using a conditional GAN

with contour images was employed [41]. Zaltron et al. [42] chose an interactive setup with the additional possibility of fine-tuning the faces obtained during the process with the help of sliders. Also in [28] images were developed either by human feedback or by similarity to a target image. It was further shown that genetic algorithms are able to generate diverse sets of latent variables [11], which was also used to augment sparse training datasets [12]. In addition, game levels were developed both automatically [40] and interactively [31]. In a slightly different approach, Roziere et al. [29] used a (1+1) evolution strategy to find local optima in the neighbourhood of generated artefacts including artworks using automatic image quality evaluation metrics.

2.3 Evolutionary Art

Interactive Fitness Evaluation Evolutionary art can look back on a long history ([27], for an overview). From the beginning, the subjectivity of art has been a major challenge, since the definition of an objective fitness function is not trivial [20,22]. For this reason, in various creative domains such as games, music, and image generation often an IEC approach has been chosen [37]. One problem with IEC, however, is that humans can only evaluate a relatively small number of candidates in a reasonable amount of time and get tired after only a few generations, which is known as user fatigue. Therefore, numerous attempts have already been made to reduce it [37].

One solution to these problems can be provided by crowdsourcing based interactive evolution. In image generation, this approach has been applied by Picbreeder [32], where people can further evolve the evolved images of others online. This approach is also utilised in Artbreeder [34], a creative platform for the creation of GAN-based images, which was inspired by Picbreeder.

The IEC approach proposed in this work involves a collaborative approach differing from those already described in that the fitness is formed based on the ratings of several people. Even though only a small number of the same participants evaluated the images in this study, this approach seems well suited in such a subjective domain.

In general, it opens up interesting possibilities for taking different scores into account and measure the subjectivity. Expanding this approach to different users at different times could also help overcome the problem of user fatigue. An example of this is the Electric Sheep Project [8], ongoing since 1999, in which the fitness is determined by many different users around the world.

Automatic Fitness Evaluation Given the problems in IEC, there has been plenty of research on automatic aesthetic fitness evaluation metrics [20,22]. Using deep learning methods, great progress has been made for assessing images [38]. The availability of large annotated datasets for aesthetic visual analysis, such as AVA [23], which contains over 250,000 images rated by hobby photographers, have also contributed to these advances.

Even though most datasets consist of photographs, it has already been shown that automatic evaluation metrics can improve artworks for human viewers, at

least by using a technical image quality assessment [29]. Despite their less successful results in this way, we chose an aesthetic visual metric, since we assume that art does not only function through its technical quality. The aesthetic quality of the images is determined using Neural Image Assessment (NIMA) [38], which is based on an InceptionResNet-v2 [36] image classifier architecture and is trained on AVA [23].

3 Methods

3.1 Creative Adversarial Network

In this work, the generator part of a GAN is used as genotype-to-phenotype mapping. As explained in 2.1, the CAN [10] architecture is chosen for this. It aims to generate creative art that mimic the real art distribution, but at the same time deviate from established art styles. To achieve this, in addition to the classification of true and fake images, the discriminator has the further objective of assigning the true images to a certain art style. The generator, on the other hand, still has the goal of ensuring that the generated images are not recognised as fake, but in addition the discriminator should find it as difficult as possible to classify them into a particular art style. Further information and a block diagram illustrating this setup can be found in the original work [10].

Technical Details The design essentially follows the CAN [10] architecture, which in turn is based on the DCGAN [25] architecture. This consists of a series of strided convolutions for the discriminator and fractional strided convolutions for the generator. Each convolution is followed by a batch normalisation, except in the generator output layer and in the discriminator input layer. The activation function used in the discriminator is Leaky ReLu and in the generator ReLu, only in the generator output Tanh. A special feature of the CAN architecture is that the last strided convolution in the discriminator is followed by two heads. The first determines the probability of coming from the real image distribution using a fully connected layer. The second determines the probability of classification into the different art styles by means of three fully connected layers.

Since we showed the images on screens in 16:9 format, the original square format was converted accordingly. For this purpose, in the second, third, and fourth fractional strided convolution in the generator, the kernel sizes are adjusted from (4,4) to (2,4). The latent vectors z that are put under evolutionary control after training are of length 100 and drawn from a standard normal distribution. The exact architecture is thus:

Generator:

$$z \in \mathbb{R}^{100} \rightarrow 4 \times 4 \times 1024 \rightarrow 8 \times 6 \times 1024 \rightarrow 16 \times 10 \times 512 \rightarrow 32 \times 18 \times 256 \rightarrow 64 \times 36 \times 128 \rightarrow 128 \times 72 \times 64 \rightarrow 256 \times 144 \times 3 \text{ (final resolution)}$$

Discriminator:

$256 \times 144 \times 3 \rightarrow 128 \times 72 \times 32 \rightarrow 64 \times 36 \times 64 \rightarrow 32 \times 18 \times 128 \rightarrow 16 \times 9 \times 256 \rightarrow 8 \times 4 \times 512 \rightarrow 4 \times 2 \times 512$

head 1: $4 \times 2 \times 512 = 4096 \rightarrow 1$

head 2: $4 \times 2 \times 512 = 4096 \rightarrow 1024 \rightarrow 512 \rightarrow K$ (number of art styles)

We used the publicly available WikiArts dataset [39] as training data. It consists of 81,444 artworks from 27 different art styles. The images were normalised and resized to the appropriate resolution. Since this resolution is too low for display, we upsampled the images generated by a factor of eight for human evaluation, resulting in a resolution of 2048×1152 . The Laplacian Pyramid Super-Resolution Network (LapSRN) [19] was used for this purpose. The framework is available for experimentation¹.

3.2 Collaborative Interactive Evolution

After training, the generator is able to generate images from every possible latent vector z that follow the distribution of the training images but are difficult to classify into a specific art style. The first generation in the evolution consists of images resulting from randomly generated latent vectors. To not overwhelm the users, IEC classically uses small population sizes and a low number of generations [4,22]. In this work, we choose a population size of 15 and evolve it over 25 generations. In the following, the individual stages of the evolutionary algorithm are outlined. An overview of the algorithm is provided in Figure 1.

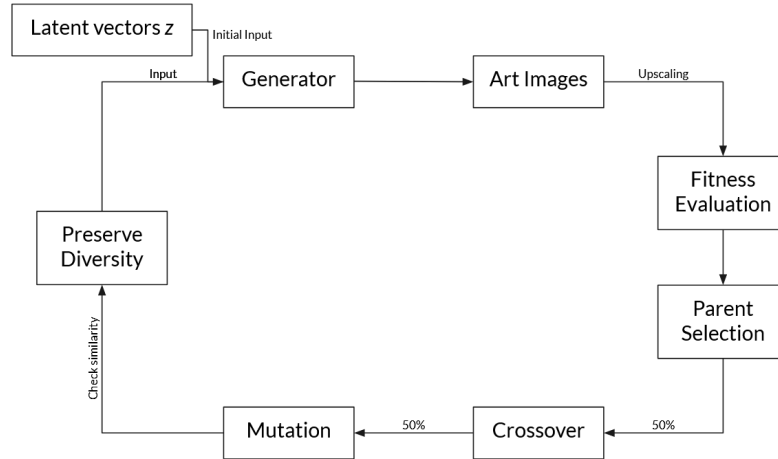


Fig. 1. Block diagram of the evolutionary algorithm.

¹ <https://github.com/OMHall/CollaborativeArt>

Fitness Evaluation In the collaborative interactive evolution, five participants rated each image independently on a scale of 1 to 10. For this purpose, a questionnaire with the upsampled images was sent to all participants every generation. They were advised to rate the images independently according to how much they like them, rather than comparing them. The average of these ratings then form the fitness of an image. An excerpt of the questionnaire can be found in the supplementary material in Figure 10.

In the automatic evolution, fitness is evaluated using an automatic aesthetic evaluation by NIMA [38], as outlined in 2.3. Starting from the same initial population and using the same algorithm otherwise, this allows for a comparison of the results.

Parent Selection Due to the collaborative evaluation, no direct selection of parents can be performed as is often the case in IEC [4,9,41]. Instead, stochastic universal sampling (SUS) [1] is used for parent selection, whereby 15 parents are selected. The next generation is formed from these through crossover and mutation.

Crossover Crossover is applied to two randomly chosen parents in 50% of the cases. As in other studies on IEC in combination with GANs, uniform crossover is chosen [4,41], whereby the probability of exchanging individual genes is set to 25%. The rationale behind this is that crossover should lead to interesting new images on the one hand, but on the other hand the selected parents should also be recognisable in order to reduce user frustration due to the loss of good solutions [9].

Mutation In order to explore the latent space quickly, the mutation probability is also set to 50%, as in [4]. A local search is implemented as mutation, which should both accelerate exploration and further increase the quality of the images in order to keep user fatigue low. This can be understood as intelligent mutation in the sense of a hybrid or memetic evolutionary algorithm [9].

The local search is implemented following Roziere et al. [29], who search for the best possible image in the neighbourhood of a latent vector z . For this, a $(1 + 1)$ evolution strategy [9] is used. In this strategy, there is initially one parent image, from which one offspring is created through mutation of the underlying latent vector z . Based on the results of [29], $1/\text{length of } z = 1/100$ is chosen as individual mutation rate. If a gene is mutated, the mutation comprises the addition of a random number drawn from a standard normal distribution. Survivor selection then takes place on the basis of deterministic elitist replacement, meaning that only the better evaluated image is kept. The quality of the images resulting as phenotypes is automatically evaluated using NIMA [38]. In trade-off between quality increase and preservation of diversity, each local search spans 100 generations, since too many generations tend to result in less preserved diversity [29].

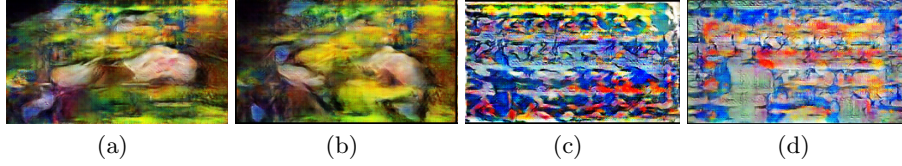


Fig. 2. Illustration of the *preserve diversity* metric. While the examples (a) and (b) would be judged as too similar, the examples (c) and (d) exhibit sufficient differences.

Preservation of Diversity Moreover, a metric is introduced that aims to keep the diversity in the population high. On the one hand, this has the goal of exploring the latent space quickly, since in IEC due to user fatigue usually only a small part of the search space is considered. On the other hand, this is intended to avoid user frustration, which can emerge after rating similar images over several generations and a feeling of being in a “blind alley” [9].

The metric chosen is the sum of the absolute distances between individual genes, which is not allowed to fall below a certain threshold. This threshold is set to 25, and examples of the effects of this metric can be found in Figure 2. After crossover and mutation, this metric is applied to the entire population. Given that two images are too similar, one is replaced by a random immigrant, i.e. an image consisting of a completely new random latent vector. Random immigrants are a way to explicitly increase diversity [16], and have also been used for this purpose in previous work on IEC in combination with GANs [4,41]. In order to increase the quality of the random immigrants, they are first undergoing a local search over 100 generations, as implemented as mutation.

3.3 Evaluation

The final evaluation by human participants of the results obtained with this framework consists of three parts. Each part is composed of pairwise comparisons, where two images are placed next to each other and participants have to decide which one they prefer. In the first part, 20 random images are compared before and after performing the local search to test its effect. In the second and third part, the results of the automatic and collaborative evolution are compared with random images as in the first generation. For this purpose, 10 images each were selected from the hall of fame, the overall best images during the evolutionary process, and randomly combined with the comparison images. Throughout the questionnaire, the order of the images is balanced to avoid position bias. An exemplary excerpt can be found in the supplementary material in Figure 11.

A total of $N = 31$ participants were recruited for the evaluation. Among them were 16 men, 13 women, and one person each who indicated a different gender or did not want to indicate their gender. The participants were between 20 and 87 years old ($M = 31.8, SD = 13.9$). In the evaluation, none of the participants involved in the collaborative evolution took part.

4 Results

4.1 Local Search

Figure 3(a) shows the fitness improvement of 20 random images that underwent local search for 100 generations as it was applied as mutation. These 20 images were also used to evaluate whether local search resulted in quality improvement for human participants.

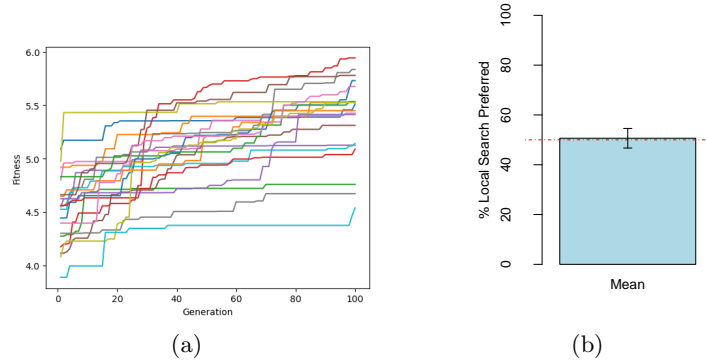


Fig. 3. (a) Fitness improvement of 20 random images over 100 generations of local search according to the automatic evaluation metric. (b) Proportion of participants preferring the local search results over the original image, averaged over all 20 images.

Figure 3(b) illustrates the proportion at which the local search was preferred over the original image, averaged over all 20 images. The mean value of 51% shows that the results are preferred only in slightly more than half of the cases, which corresponds to random level. Looking at the individual comparisons, besides many close decisions, local search led to clear improvements in some cases, but also to clear deteriorations in others. The most successful and least successful local searches are exemplified in Figure 4. Overall, it seems that sharper contours and more intense colours were perceived as improvements, while colour or composition changes received less approval.

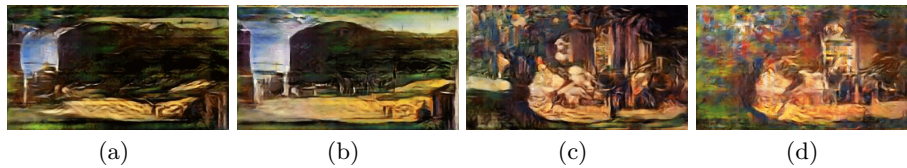


Fig. 4. Images (b) and (d) result from images (a) and (c) using the local search. (b) was preferred in 77% over (a), while (d) was preferred only in 10% over (c).

4.2 Automatic Evolution

The evolution based on the automatic aesthetic evaluation metric resulted in increasing fitness through the generations. This trend can be seen in Figure 5, which displays results averaged over five runs. The fitness increased particularly at the beginning and seemed to reach a plateau at the end. This trend is presumably supported by the fact that the mutation used also evolved the individual images with respect to the same aesthetic evaluation metric.

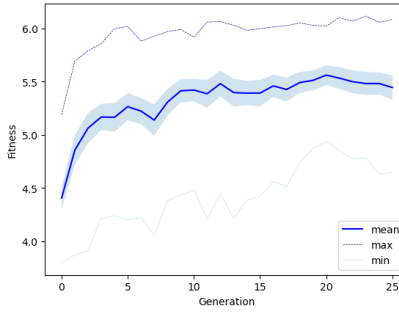


Fig. 5. Mean fitness through generations of the automatic evolution, with the shaded area representing the standard error. The results are averaged over five runs.

All images of the first and last generation are shown in the supplementary material in Figure 12, a visualisation of the entire evolution can be found online². A selection of the best images overall is displayed in Figure 6. It is apparent that the automatic evolution evolved in the direction of rather blurred images with less clear shapes. With the exception of the image in Figure 6(d), which has clearer structures and could be reminiscent of a landscape, diffuse contours predominate. Partly, a sky might be recognisable.

Evaluation The evaluation regarding the attractiveness of the images obtained in the automatic evolution compared to random images as in the first generation showed that human participants preferred the obtained images on average in 49% of the cases, which corresponds to random level. All individual comparisons and the resulting mean can be found in the supplementary material in Figure 13(a). Only two comparisons were clearly in favour of the automatic evolution results, while a majority of comparisons tended towards the random images.

4.3 Collaborative Evolution

The collaborative interactive evolution resulted in increasing fitness through the generations, too. This trend can be seen in Figure 7(a). In contrast to the automatic evolution, this increase was rather linear and kept rising until the end.

² Visualisation of the automatic evolution: https://youtu.be/JCRx3Ih_0hA

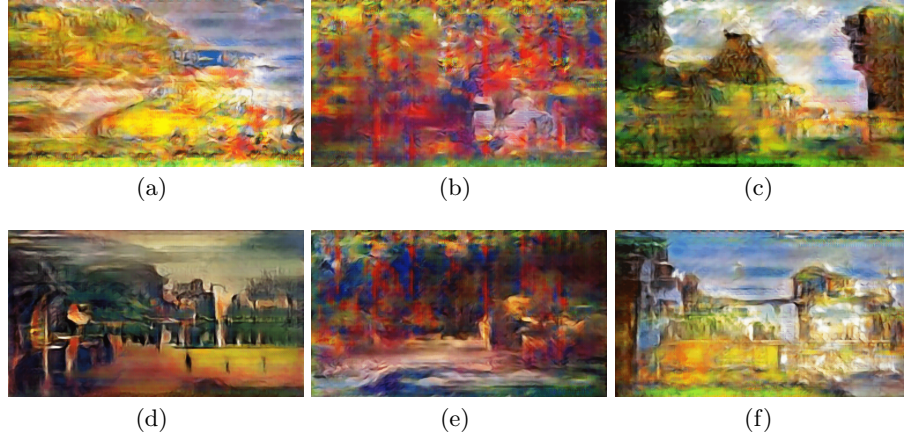


Fig. 6. A selection of the overall best images of the automatic evolution.

Figure 7(b) shows how the automatic aesthetic evaluation metric would have assessed the fitness of the collaborative interactive evolution. Interestingly, similar to the actual automatic evolution in Figure 5, the fitness is strongly increasing at the beginning and rather stagnant afterwards, which may also be due to the effect of the local search. Compared to the collaborative interactive evolution, there are similarities such as the local minima at generation 20, but the trends also reveal many differences.

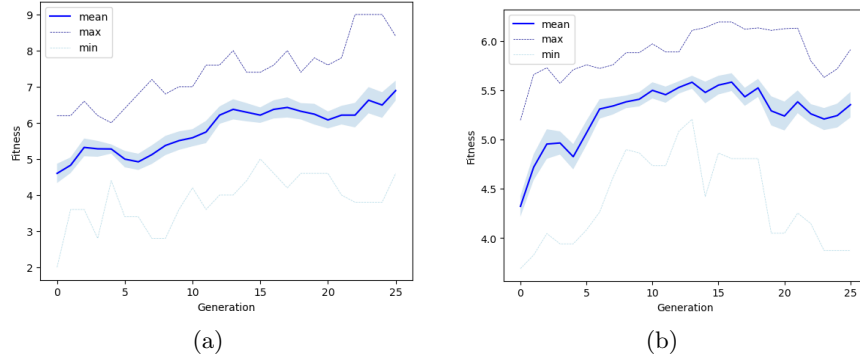


Fig. 7. (a) Mean fitness through generations of the collaborative interactive evolution. (b) The collaborative interactive evolution assessed by the automatic aesthetic evaluation metric. The shaded areas represent the standard error.

A total of 17 random immigrants were introduced into the population during the evolutionary process. Only in generation two, two immigrants were inserted,

otherwise at most one. Interestingly, the immigrants were rated above average in 14 out of 17 cases and were on average 0.67 above the generation means. This could indicate that novelty was perceived as positive by human participants throughout the evolutionary process.

The collaborative interactive approach further revealed how subjectively art images are perceived. The ratings differed significantly in some cases, with ranges across the entire scale (1-10) and standard deviations of up to $SD = 3.38$. For some images, however, the ratings were also highly similar. Figure 8 shows in (a) the image with the widest range in ratings and in (b) the image with the closest agreement in ratings. In fact, the image was the only one rated equally by all participants with an 8.

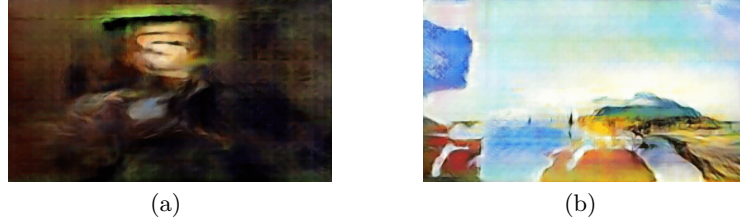


Fig. 8. (a) The image with the widest range in ratings (1-10). (b) The image with the closest agreement (all 8).

As with the automatic evolution, all images of the first and last generation are shown in the supplementary material in Figure 12, and a visualisation of the entire evolution can be found online³. A selection of the best images overall is displayed in Figure 9. In contrast to the automatic evolution, sharper contours prevail. The images are less blurry and the colours appear more diverse, also the images seem to differ more from each other. Many of the images evoke landscape-like associations or are reminiscent of abstract art.

Evaluation The evaluation regarding the attractiveness of the images obtained in the collaborative interactive evolution compared to random images as in the first generation showed that human participants preferred the obtained images on average in 60% of the cases. Averaged across all ten comparisons and all assessments, an exact binomial test revealed that the results of the collaborative evolution were significantly preferred over random images ($T = 185, p < .001$). All individual comparisons and the resulting mean can be found in the supplementary material in Figure 13(b). Some decisions were close and do not point clearly in one direction considering the standard errors. However, with descriptively eight decisions in favour of the results of the collaborative evolution and in at least three cases a strong preference, a clear tendency is recognisable.

³ Visualisation of the collaborative evolution: https://youtu.be/rG_pLiX_UFo

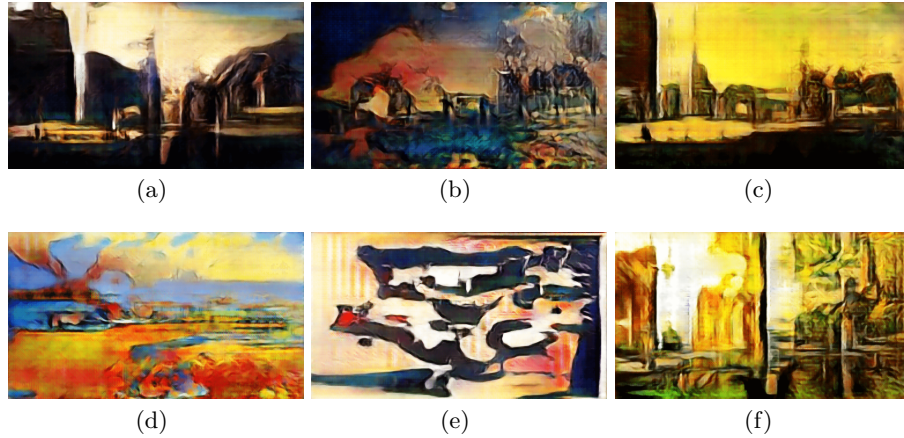


Fig. 9. A selection of the overall best images of the collaborative evolution.

5 Conclusion

In this work, we introduced a novel framework to explore the latent space of possible art images in a GAN using evolutionary computing. It was shown that the developed framework led to an increase in quality over the evolutionary process, using both a collaborative interactive and an automatic aesthetic evaluation metric. The evolutionary algorithm was hybrid, since a local search based on an automatic evaluation metric was incorporated as an intelligent mutation. However, the evaluation by human participants revealed that the local search did not lead to improvements beyond a random level. Furthermore, only the results of the collaborative interactive evolution, but not of the automatic aesthetic evolution were found to be significantly more attractive than randomly generated images. This highlights that automatic aesthetic evaluation of art is challenging and emphasises the importance of human guidance in the evolution of art.

It was demonstrated that the use of the generator part of a GAN as genotype-to-phenotype mapping offers a promising approach for the evolution of art. Throughout all generations, diverse images were generated that can be considered as art images, and the generated images increased in their attractiveness over time. While the already high quality of the randomly generated images makes it more difficult to create significantly more attractive ones in the evolutionary process, it benefits the ongoing interest of users and the prevention of user fatigue. Further, it allows techniques such as random immigrants to be used without leading to substantial fitness loss, and is thus also conducive to accelerated exploration of latent space. The introduced collaborative interactive approach indicated its usefulness due to the subjectivity of art and the positive outcomes, and opens up a multitude of future research possibilities.

References

1. Baker, J.E.: Reducing bias and inefficiency in the selection algorithm. In: Proceedings of the Second International Conference on Genetic Algorithms. pp. 14–21 (1987)
2. Bau, D., Strobel, H., Peebles, W., Wulff, J., Zhou, B., Zhu, J.Y., Torralba, A.: Semantic photo manipulation with a generative image prior. *ACM Transactions on Graphics* **38**, 1–11 (2019)
3. Bau, D., Zhu, J.Y., Strobel, H., Zhou, B., Tenenbaum, J.B., Freeman, W.T., Torralba, A.: GAN dissection: Visualizing and understanding generative adversarial networks. In: Proceedings of the International Conference on Learning Representations (ICLR). p. 19 (2019)
4. Bontrager, P., Lin, W., Togelius, J., Risi, S.: Deep interactive evolution. In: Computational Intelligence in Music, Sound, Art and Design: 7th International Conference, EvoMUSART 2018. pp. 267–282. Springer (2018)
5. Bontrager, P., Roy, A., Togelius, J., Memon, N., Ross, A.: Deepmasterprints: Generating masterprints for dictionary attacks via latent variable evolution. In: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). pp. 1–9 (2018)
6. Cetinic, E., She, J.: Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* **18**(2), 1–22 (2022)
7. Cohn, G.: AI art at Christie’s sells for \$432,500 (2018), <https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html>
8. Draves, S.: The electric sheep screen-saver: A case study in aesthetic evolution. In: Applications of Evolutionary Computing, LNCS 3449. Springer (2005)
9. Eiben, A.E., Smith, J.E.: Introduction to evolutionary computing. Berlin: Springer (2003)
10. Elgammal, A., Liu, B., Elhoseiny, M., Mazzone, M.: CAN: Creative adversarial networks, generating ”art” by learning about styles and deviating from style norms. *arXiv:1706.07068* (2017)
11. Fernandes, P., Correia, J., Machado, P.: Evolutionary latent space exploration of generative adversarial networks. In: Applications of Evolutionary Computation – 23rd European Conference, EvoApplications. pp. 595–609. Springer (2020)
12. Fernandes, P., Correia, J., Machado, P.: Towards latent space exploration for classifier improvement. In: 24th European Conference on Artificial Intelligence (ECAI 2020) – ADGN20: First workshop on Applied Deep Generative Networks (2020)
13. Gajdacz, M., Rafner, J., Langsford, S., Hjorth, A., Bergenholtz, C., Biskjaer, M.M., Noy, L., Risi, S., Sherson, J.: CREA.blender: A GAN based casual creator for creativity assessment. In: Proceedings of the International Conference on Computational Creativity ICC3. p. 5 (2021)
14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. *Advances in Neural Information Processing Systems* **27** (2014)
15. Grabe, I., Duque, M., Risi, S., Zhu, J.: Towards a framework for human-AI interaction patterns in co-creative GAN applications. In: Joint Proceedings of the ACM IUI Workshops (2022)
16. Grefenstette, J.J.: Genetic algorithms for changing environments. In: Proceedings of Parallel Problem Solving from Nature. pp. 137–144 (1992)

17. Gui, J., Sun, Z., Wen, Y., Tao, D., Ye, J.: A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Transactions on Knowledge and Data Engineering* **35**(4), 3313–3332 (2023)
18. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. In: *International Conference on Learning Representations* (2018)
19. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: *IEEE Conference on Computer Vision and Pattern Recognition* (2017)
20. Machado, P., Romero, J., Manaris, B.: Experiments in computational aesthetics: An iterative approach to stylistic change in evolutionary art. In: *The art of artificial evolution: A handbook on evolutionary art and music*, pp. 381–415. Springer (2008)
21. Martindale, C.: *The clockwork muse: The predictability of artistic change*. Basic Books (1990)
22. McCormack, J.: Facing the future: Evolutionary possibilities for human-machine creativity. In: *The art of artificial evolution: A handbook on evolutionary art and music*, pp. 417–451. Springer (2008)
23. Murray, N., Marchesotti, L., Perronnin, F.: AVA: A large-scale database for aesthetic visual analysis. In: *IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2408–2415 (2012)
24. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.: Semantic image synthesis with spatially-adaptive normalization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 2337–2346 (2019)
25. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv:1511.06434* (2015)
26. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative adversarial text to image synthesis. In: *ICML* (2016)
27. Romero, J.J., Machado, P. (eds.): *The art of artificial evolution: A handbook on evolutionary art and music*. Springer (2008)
28. Rozière, B., Riviere, M., Teytaud, O., Rapin, J., LeCun, Y., Couprie, C.: Inspirational adversarial image generation. *IEEE Transactions on Image Processing* **30**, 4036–4045 (2021)
29. Rozière, B., Teytaud, F., Hosu, V., Lin, H., Rapin, J., Zameshina, M., Teytaud, O.: EvolGAN: Evolutionary generative adversarial networks. In: *Proceedings of the Asian Conference on Computer Vision* (2020)
30. Sbaji, O., Elhoseiny, M., Bordes, A., LeCun, Y., Couprie, C.: DeSIGN: Design inspiration from generative networks. In: *ECCV workshop on Fashion, Art and Design* (2018)
31. Schrum, J., Gutierrez, J., Volz, V., Liu, J., Lucas, S., Risi, S.: Interactive evolution and exploration within latent level-design space of generative adversarial networks. In: *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*. pp. 148–156 (2020)
32. Secretan, J., Beato, N., D Ambrosio, D.B., Rodriguez, A., Campbell, A., Stanley, K.O.: Picbreeder: Evolving pictures collaboratively online. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 1759–1768 (2008)
33. Shen, Y., Gu, J., Tang, X., Zhou, B.: Interpreting the latent space of GANs for semantic face editing. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 9240–9249 (2020)
34. Simon, J.: *Artbreeder* (2018), <https://www.artbreeder.com>

35. Spratt, E.L.: Creation, curation, and classification: Mario Klingemann and Emily L. Spratt in conversation. *XRDS: Crossroads, The ACM Magazine for Students* (2018)
36. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 2818–2826 (2016)
37. Takagi, H.: Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE* **89**(9), 1275–1296 (2001)
38. Talebi, H., Milanfar, P.: NIMA: Neural image assessment. *IEEE Transactions on Image Processing* **27**(8), 3998–4011 (2018)
39. Tan, W.R., Chan, C.S., Aguirre, H., Tanaka, K.: Improved artgan for conditional synthesis of natural image and artwork. *IEEE Transactions on Image Processing* **28**(1), 394–409 (2019)
40. Volz, V., Schrum, J., Liu, J., Lucas, S.M., Smith, A., Risi, S.: Evolving mario levels in the latent space of a deep convolutional generative adversarial network. In: *Proceedings of the Genetic and Evolutionary Computation Conference*. p. 221–228 (2018)
41. Xin, C., Arakawa, K.: Object design system by interactive evolutionary computation using GAN with contour images. In: *Human Centred Intelligent Systems: Proceedings of KES-HCIS 2021 Conference*. pp. 66–75. Springer (2021)
42. Zaltron, N., Zurlo, L., Risi, S.: CG-GAN: An interactive evolutionary GAN-based approach for facial composite generation. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. pp. 2544–2551 (2020)
43. Zhu, S., Fidler, S., Urtasun, R., Lin, D., Loy, C.C.: Be your own prada: Fashion synthesis with structural coherence. In: *International Conference on Computer Vision ICCV* (2017)

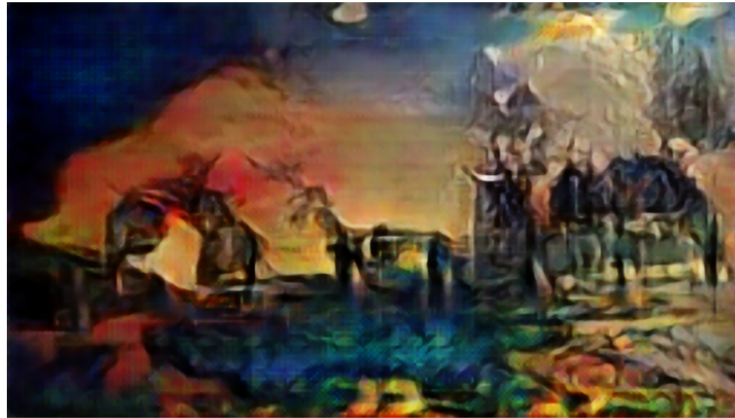
Supplementary Material

Interactive Evolution 25

Hello, thank you for participating in the collaborative evolution of art. Please rate the images below on a scale from 1 - 10, with 1 being the lowest and 10 being the highest. This is not a comparison of the images, please rate the images independently based on how much you like them.

Best,
Ole

Please rate image 1 *



1	2	3	4	5	6	7	8	9	10
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

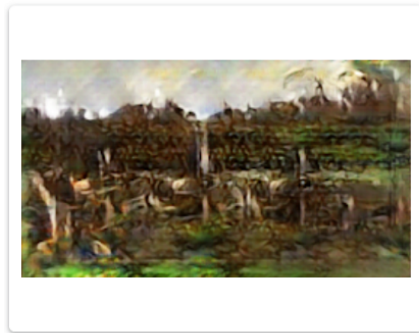
Fig. 10. Illustrative excerpt of the collaborative interactive evolution questionnaires.

Thesis survey

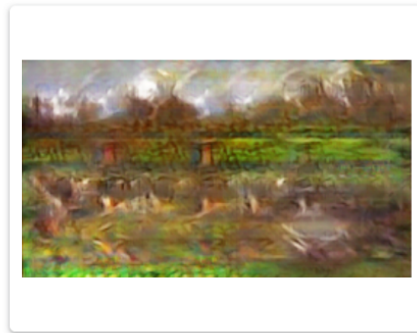
Hello, thank you for taking part in this study on comparing art images. The questionnaire consists of three parts. The first one briefly asks for some demographic information. The other two parts both contain 20 comparisons. In these, two art images are shown side by side and you are asked which of the two you like better.

Best,
Ole

Question 1: Please select the image you prefer *

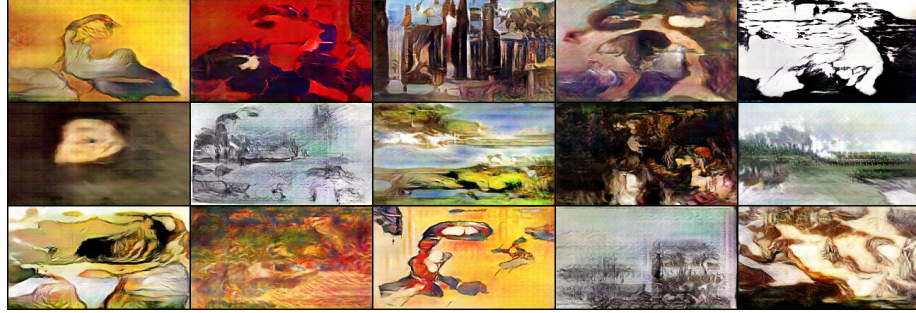


☐ Image A



☐ Image B

Fig. 11. Illustrative excerpt of the final evaluation questionnaire.



(a)



(b)

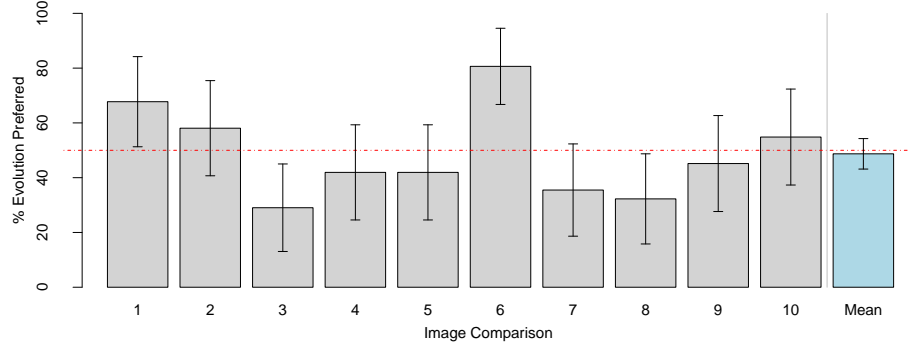


(c)

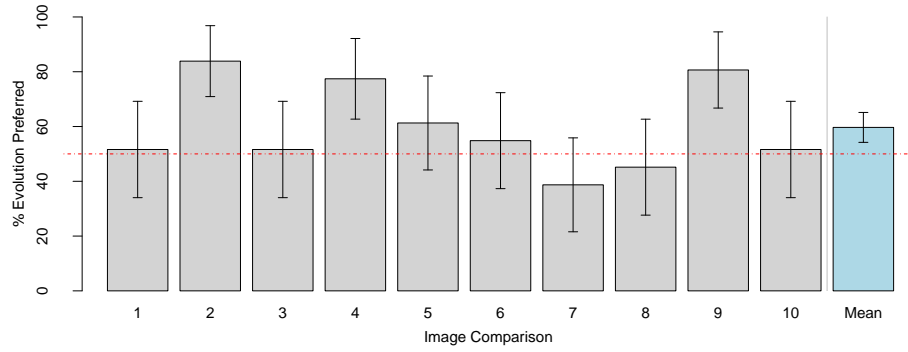
Fig. 12. (a) Population of random start images in the 1st generation, which served as starting point for both the automatic and the collaborative interactive evolution. (b) Population after 25 generations, evolved by the automatic aesthetic evaluation metric. (c) Population after 25 generations, evolved by the collaborative interactive evaluation. Visualisations of the entire evolutions can be found online.

→ Automatic evolution: https://youtu.be/JCRx3Ih_0hA

→ Collaborative evolution: https://youtu.be/rG_pLiX_UFo



(a)



(b)

Fig. 13. (a) Proportion of participants preferring the results of the automatic aesthetic evolution over random images. (b) Proportion of participants preferring the results of the collaborative interactive evolution over random images. Both for all 10 comparisons as well as averaged over all images. The bars represent the standard errors.