Generative System to Assist the Artist in the Choice of 3D Composition for a Still Life Painting

Guido Salimbeni , Frederic Fol Leymarie , William Latham.
Department of Computing, Goldsmiths, University of London, U.K.
Contacts: <g.salimbeni,ffl,w.latham@gold.ac.uk>

Abstract

We present a system built to generate an optimal arrangement of 3D models for aesthetic evaluation, with the aim to support an artist in their creative process. The novel architecture comprises four neural networks each used for a specific function, together with an evolutionary algorithm to generate samples. We explore how this system can automatically generate aesthetically pleasing content for use in the media and design industry, based on standards originally developed in master artworks. We demonstrate the effectiveness of our process in the context of paintings using a collection of images inspired by the work of the artist Giorgio Morandi[1]. Finally, we compare the results of our system with the results of a well-known Generative Adversarial Network (GAN).

1 Introduction and Related Work

By leveraging the features of a game engine framework, our system can analyse both the rendered image on the screen and obtain information on the position, depth and volume of the objects in the three-dimensional digital environment. We believe that these characteristics, coupled with artificial intelligence that elaborates prediction from that information, represent a valuable solution for the media and design industry that uses 3D models in the prototyping stage. In this article, we demonstrate the effectiveness of our method by undertaking a specific task that involves the artistic choice of pictorial compositions. Finding the optimal arrangement of visual elements for a painting is a manual and time-consuming activity that the painter pursues by applying general principles of design and personal aesthetic intuition. Our system uses a three-dimensional digital prototype of the objects to paint or render in a design and proposes a computational technique to assist the painter or designer in the selection of the most aesthetically pleasing composition for the work. In this communication, we illustrate how our system uses an evolutionary algorithm (EA) and four artificial neural networks(NNs) to automate and speed up the creative process of choice of composition. Finally, we compare the results of our system with those produced by a GAN currently used in commercial applications.

Several A.I. technologies have been applied in an attempt to automate the creative process, with significant and promising results [2, 7, 3]. In particular, [5] showed success in procedural design approaches that optimise designs for functional requirements. Kowaliw *et al.* [6] developed the EvoEco system that automatically detects and improves creative designs. Some research works on the design of fitness functions that can emulate human aesthetic preference by using machine learning technologies such as neural networks [1] and co-evolutionary algorithm [4]. However, the present work is the first to our knowledge to combine 3D model manipulation inside a game engine framework and a machine learning architecture to calculate the fitness function. Moreover, this work is the first, to our knowledge, to apply artificial intelligence in the selection and arrangement of 3D models for still-life paintings and designs.

[1] G. Morandi: Italian painter and printmaker — Bologna, 1890 – 1964.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

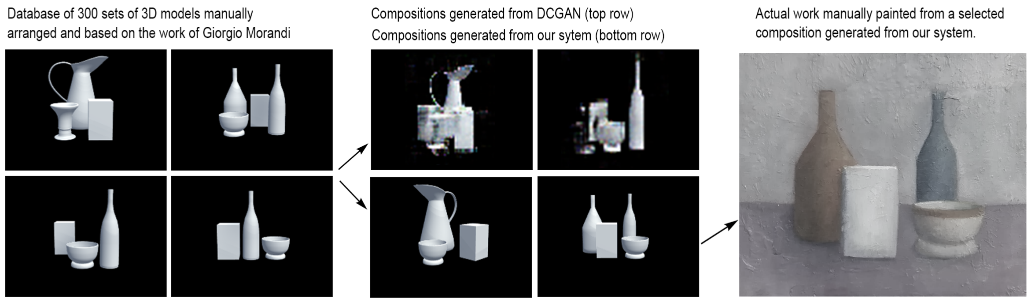

Figure 1: From an initial database to a final 3D composition and painting.

2 Proposed Generative System

Our system uses a Genetic Algorithm (GA) to generate a series of compositions of 3D models within Unity3D (www.unity.com) together with a automatic score to select the most visually balanced composition. This value is the output of an architecture made from four NNs trained with a database of numerical data and images each associated with an aesthetic judgement previously assigned by the artist. The database consists of 300 sets of 3D models arranged in space imitating the artworks of Giorgio Morandi. Through a process of crossover, mutation and selection based on the aesthetic value produced by the system independently of the user's intervention, the genetic algorithm generates compositions with a consistently better aesthetic evaluation and finally proposes an optimal composition layout for painting. The NN architecture is composed of 4 separate networks. The first is a fully connected NN that processes the front and top views of the 3D scene and assigns a positive or negative value to the composition by considering the volume and position of the 3D objects. The second is a CNN that uses the perspective main camera as input and calculates the similarity of the current composition in Unity3D with the images of the initial database and their assigned composition value. The third is another fully connected NN that takes the data directly from Unity3D and checks the symmetry of the 3D scene, any intersection of 3D models, the isolation of the objects and how much screen space they cover. These 3 networks each individually evaluate a specific aspect of the composition and then merge their outputs into a fourth NN that produces the final composition value between 0 and 1, which can be fed into the GA (architecture detailed in Appendix A).

3 Discussion

Our system represents a different approach to the generation of images that can be used by the painter to expand and expedite the creative decision of pictorial composition, but also by designers working in the media and design industry. In Figure 1, we present some images generated by our system and the outputs from a Generative Adversarial Network (DCGAN) [8]. Here, we list our conclusions regarding this comparison. Our application can generate new compositions in which the 3D objects do not intersect because the system is aware of the volume and position of the models in the three-dimensional space. The same ability is difficult to verify with the DCGAN. Moreover, the ability of our system to position the 3D model at a different point along the perspective axis is more evident than with DCGAN. The variety of composition solutions proposed by our system also includes the rotation of objects along their vertical axis. DCGAN working only with 2D images cannot recognize the 3D attributes of the objects and so is not able to rotate them to produce greater variety in composition solutions. Our system can create new compositions with 3D models not included in the original database since the criterion for aesthetic evaluation is less dependent of the object's type and shape than with DCGAN. However, DCGAN can generate new images quickly, while the current state of our system requires the generation of several potential solutions before selection. We intend to add a Reinforcement Learning solution to the system to speed up the selection process and to experiment in other fields including design, game content generation and architecture.

References

[1] Shumeet Baluja, Dean Pomerleau, and Todd Jochem. Towards automated artificial evolution for computer-generated images. *Connection Science*, 6(2-3):325–354, 1994.

[2] Carlos M Fernandes, Antonio M Mora, Juan J Merelo, and Agostinho C Rosa. Kants: A stigmergic ant algorithm for cluster analysis and swarm art. *IEEE transactions on cybernetics*, 44(6):843–856, 2013.

[3] Mario García-Valdez, Juan J Merelo, Leonardo Trujillo, Francisco Fernández-de Vega, José C Romero, and Alejandra Mancilla. Evospace-i: a framework for interactive evolutionary algorithms. In *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*, pages 1301–1308. ACM, 2013.

[4] Gary R Greenfield. Co-evolutionary methods in evolutionary art. In *The Art of Artificial Evolution*, pages 357–380. Springer, 2008.

[5] Rubaiat Habib Kazi, Tovi Grossman, Hyunmin Cheong, Ali Hashemi, and George W Fitzmaurice. Dreamsketch: Early stage 3d design explorations with sketching and generative design. In *UIST*, volume 14, pages 401–414, 2017.

[6] Taras Kowaliw, Alan Dorin, and Jon McCormack. Promoting creative design in interactive evolutionary computation. *IEEE transactions on evolutionary computation*, 16(4):523, 2012.

[7] Chien-Hung Liu and Chuan-Kang Ting. Evolutionary composition using music theory and charts. In *2013 IEEE Symposium on Computational Intelligence for Creativity and Affective Computing (CICAC)*, pages 63–70. IEEE, 2013.

[8] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

Ethical Considerations

Automated systems that replace people in decision-making require ethical considerations about the methodology and the actual capacity left to the person for evaluating the decisions. This aspect becomes even more critical when the automatic mechanism is produced by a machine learning algorithm, which is often difficult to interpret, so the well-known black box name. Furthermore, the ability of neural networks to generalise beyond the data acquired during the learning process is often limited to the characteristics of the initial database. Therefore, it becomes essential to define the characteristics of the training database accurately so that the end-user is aware of the real potential of the machine learning application. We acknowledge, for example, the impossibility of our system to obtain a system of production of universal aesthetic values also because the initial database is limited to compositions selected by the user. We think that declaring the genuine ability of the application should be clearly stated for any machine learning implementation. We do not believe that our system is a tool that does replaces the critical evaluation of the user. Rather, it works like a sketchbook that proposes images, giving the user the possibility of using them as they are, correcting them or taking inspiration for new aesthetic solutions. Even if technological advancement could calculate a universal aesthetic score, the machine would hardly have the purely human capacity to subvert the schemes of the past and for this generate a new aesthetic. The ethical risk, in this case, would be the potential flattening of the concept of beauty towards a common and standardised value. This risk would be even more relevant in commercial, economic sectors, where the desire to reach a valid and generalisable result quickly and efficiently could tend to flatten the creative novelty.

A1. System's Architecture

In figure 2 we show the functional diagram of the entire system. The generation of the composition inside Unity3D is achieved using a genetic algorithm. The set of 4 NNs and their inputs values are used to generate an overall score for the fitness calculation.

The NNs are implemented with the Python programming language, using the Keras machine learning library with the Tensorflow backend. After training, the NNs are converted into bytes formats (which store the weights and architecture of the trained models) that Unity3D can use to make predictions. The system uses a plugin called "TensorFlowSharp" to facilitate the integration of machine learning models developed in Tensorflow with the Unity3D API written in C-Sharp. Each of the four NN is compiled using the categorical cross-entropy loss function and produces an output value in the range from 0 to 1.

The first (from the top-left in fig.2) NN takes as input the combined images of the front and top camera views of the scene, as obtained from Unity3D. It is composed of three fully connected layers with 128 neurons each and a final fully connected layer with two neurons and a Softmax activation function.

The second (from the top-left) NN takes as input the image from the main camera view of the scene from Unity3D. It is a CNN composed of one convolutional layer with 32 units, followed by a three convolutional layers of 64 units and three MaxPooling layers, a fully connected layer with 64 neurons and a final fully connected layer with two neurons and a Softmax activation function.

The third (from the top-left) NN takes as input four scores generated directly from Unity3D which account for: symmetry, visual unity, visual isolation and visual balance. This NN is composed of two fully connected layers with 32 neurons and 16 neurons each and a final fully connected layer with two neurons and a Softmax activation function.

These three networks join into a final fourth NN that takes their outputs as inputs. It is composed of two fully connected layers with 32 neurons and 16 neurons each and a final fully connected layer with two neurons and a Softmax activation function.

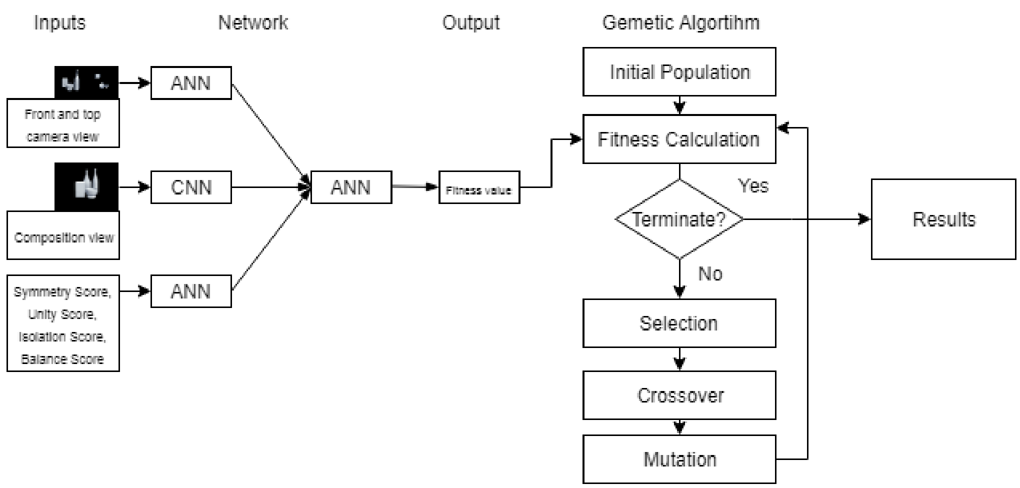

Figure 2: Architecture

A2. Compositions generated from our system and from DCGAN

In figure 3, we show more images of compositions generated from our system using the genetic algorithm illustrated in this paper. In figure 4, we show a collection of compositions produced from the DCGAN architecture.

Figure 3: Samples of compositions generated from our system

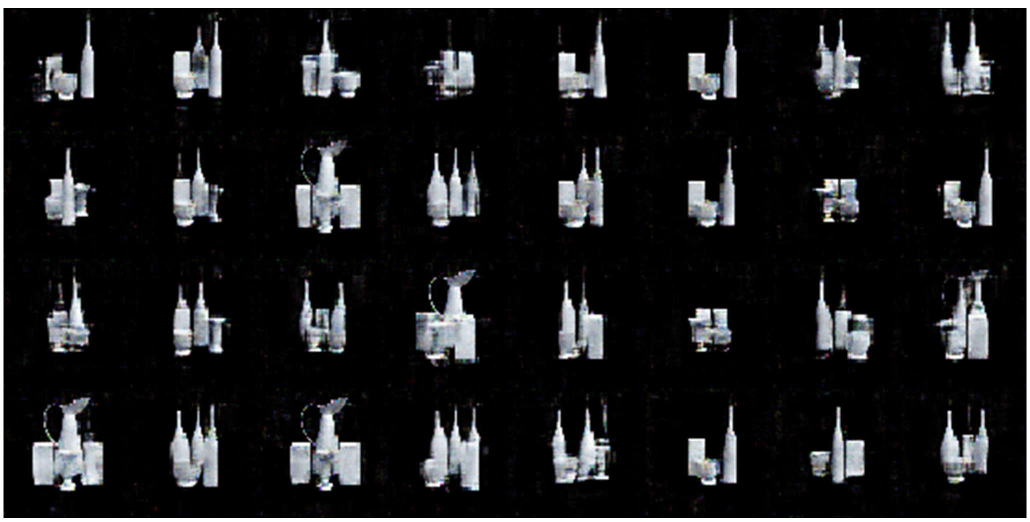

Figure 4: Samples of compositions generated from DCGAN