Neural Networks and Digital Arts: Some Reflections

Rômulo Augusto Vieira Costa*, **Flávio Luiz Schiavoni**

Arts Lab in Interfaces, Computers, and Everything Else - ALICE, Federal University of São João del-Rei, UFSJ,
São João del-Rei, Minas Gerais, Brazil

*Corresponding author: Rômulo Vieira, Federal University of São João del-Rei, Email: romulo_vieira96@yahoo.com.br

ABSTRACT: The constant advancement in the area of machine learning has unified some areas that until then did not communicate, such as the area of computing with the arts in general. With the emergence of digital art, people have become increasingly interested in the development of expressive techniques and algorithms for creating works of art, whether in the form of music, image, aesthetic artifacts, or even combinations of these forms, usually being applied in an interactive technology installation. Due to their high diversity of creation and complexity during processing, neural networks have been used to create these digital works, which present results that are difficult to reproduce by human hand and are usually presented in museums, conferences, or even at auctions, being sold at high prices. The fact that these works are gaining more and more recognition in the art scene, ended up raising some questions about authenticity and art. In this way, this work aims to address the historical context regarding the advancement of the area of machine learning, addressing the evolution of neural networks in this field, about what art would be and who would be the artist responsible for digital work, given that despite After performing a good part of the creation process, the computer does not perform the entire process, becoming dependent on the programmer, who in turn is responsible for defining parameters, techniques and, above all, concepts that will attribute all the aesthetic value to the work. From this point of view and the growing interest in the generation of art via computers, the present work presents applied research around neural network techniques and how they can be applied in artistic practice, either generating visual elements or generating sound elements. Finally, some perspectives for the future are presented and how this area can evolve even further.

KEYWORDS Computer music, Generational art, Neural networks

1. Introduction

Natural language processing uses processes from artificial intelligence, linguistics, and statistics to naturally analyze and represent the occurrence of texts or other levels of human language in computers. Although it was used in encryption and code translation systems during World War II, it was only in 1957 that the idea of generative grammar began to gain strength, thanks to the studies of Noah Chomsky. Today, this relationship between AI and linguistics can be seen in speech recognition, in the retrieval of transcribed information, summarizing and machine translation [1].

The application of artificial intelligence to text, image, and audio files, started to gain strength in the second half of the 20th century. The relationship between computers and art arose from the need to represent a piece of art digitally. Several initiatives helped in this process, among them, the AI, which allowed the manipulation and representation of this data. However, the first personal computers were unable to process this data as WAVE and BMP directly, using some other strategies, such as the symbolic representation of music and images. In Section 2, the research methodologies are presented, highlighting its literary review, understanding of the context and analysis of neural network techniques. Section 3 shows the beginning of the relationship between art and artificial intelligence.

The evolution of Computers and the possibility to add neural networks could be combined to add more sensibility to computers. Cameras became digital eyes and microphones became digital ears. Computer vision, in turn, became the area of signal processing computing that seeks to build systems capable of obtaining information through images. Focusing on machine listening, Computer music is an area of interdisciplinary study that also involves signal processing computing, encompassing concepts from computer science, electrical engineering, and of course, music. These fields of study investigate methods, techniques, and algorithms for processing and generating sound and images. Since its inception, computer music and computer vision have been strongly related to artificial intelligence, as presented in Section 4.

With the increase of computational power, some special Neural Networks were created to signal processing. Some of these NN, like CNN, made it simple to process signals, like image and/or sound, because they need fewer preprocessing stages and can process signals directly. Section 5

present some of these models.

However, since computers could directly process signals and generate signals, how can we think about artwork authorship? Looking forward to this discussion, Section 6 presents some inconclusive topics on this subject. In the end, Section 7 presents final remarks and future works.

1.1. Literature Survey

Research on the relationship between technology and art (whether visual or sound) was based on the works A (very) brief history of artificial intelligence [2], Artificial intelligence and the arts: Toward computational creativity [3], Machine learning for artists [4] and History of LISP [5].

The concepts about Neural Networks were guided by works Learning features from music audio with deep belief networks [6] and Extracting knowledge from artificial neural networks [7], while those about Convolutional Neural Network were based on Gradient-based learning applied to document recognition [8], Early Diabetes Discovery From Tongue Images [9], Intelligent Diabetes Detection System Based On Tongue Datasets [10] and A study on convolutional neural networks and their application in pedestrian detection [11].

The concepts of Generative adversarial Network go through Introduction to degenerative adversarial networks (GANs– Generative Adversarial Networks) [12], An introduction to generative adversarial networks [13] and Enhanced super-resolution generative adversarial networks [14], while Autoencoder and Variational Autoencoder are guided by works Transforming auto-encoders [15], Introduction to Autoencoders [16], An Introduction to variational Autoencoders [17] and Collaborative variational autoencoder for recommender systems [18].

2. Methodology

The central question of the work involves an approximation between neural network techniques and artistic creation. As it involves several theories, methods, and techniques of computer science, it is applied research. In addition, such research has a cyclical character, based on the following steps:

- **Understanding the Context and Rationale of the Problem**

    Systematic Analysis of Literature: the investigation of techniques, state-of-the-art, and justifications for this work go through selective bibliographic research in materials published in books, journals, and proceedings of international congresses that deal with the main theme (Convolutional Neural Network, Generative Adversarial Network, Autoencoder, and Variational Autoencoder;

    Historic evolution: from the literature review, a timeline is exposed demonstrating the historical evolution of artistic creation mediated by technology;

- **Analysis of Neural Network Techniques:** stage where techniques such as brainstorming and brainwriting are used to expand the repertoire of alternatives for artistic creation mediated by artificial intelligence. This makes the research present an exploratory character and seeks to understand the behavior, particularities, and motivations of these processes;

- **Research Rating:** with the entire functional system, tests will be carried out with users, whether experts or laymen, to validate the experimentation environment. There will also be analyzes of the answers to the questionnaires to identify problems and possible improvements.

3. The Early days - Generational Art and Symbolic data

In its history, artificial intelligence has always permeated other fields of study, such as philosophy, engineering, and even literature. Since Homer wrote about automated "tripods", even a mechanical trumpet created by Ludwig van Beethoven and the works of Jules Verne and Isaac Asimov, AI and computer systems have also become important in the design of art [2].

Concerning the use of computer systems for the creation of plastic arts, the first works date from 1950, when Franke and Laposky, without knowing about the other's experiment, used oscilloscopes to generate photographs. However, the rise of this mode of art production was in the 1980s, when personal computers and video game consoles began to improve graphics software [19]. Among the works worth mentioning are AARON [3], a pioneering robotic system capable of painting an image of reference drawings and Painting Fool [20], a software created by Simon Colton, capable of giving its artistic interpretation of the images found on the Web.

The machine learning area has also been used to create art. Between the years 1970 and 1990, Myron Krueger carried out several studies in the area of augmented reality, creating several interactive installations, which used computational techniques for artistic creation, reinforcing the relationship between the computer and art [4]. Such applications ended up inspiring a whole generation of artists/scientists who started to see the computer as a very promising tool for creating art.

Some tools played a fundamental role in this process, namely: Lisp and Prolog. Lisp is a programming language that allows you to use mathematical functions as a structure for elementary data. One of its main advantages is the ability to treat software as data, thus enabling it to serve as an input in another application. It is present in document processing, hypermedia, graphics and animation, and natural language processing [5].

An example of how LISP can be used in the Art field with symbolic music representation is the software OpenMusic [21]. This software is a visual programming environment, dedicated to computer-assisted composition and music analysis. Its main feature is the easy programming of visual modules, consisting of the connection of boxes with information that presents some logical relationship, from

a musical point of view. Thus, it allows artistic projects to be built on the computer through the expressive power of a programming language [22].

Initially, OpenMusic's visual programming tools were used to generate or process musical objects, such as scores or other symbolic data. Recent developments have extended this approach, allowing the creation of scores containing their own functional or algorithmic structure, in the form of constructed visual programs and also the processing of real-time audio.

Lisp also assists in programming languages that are used to create plugins that increase the functionality of software geared to art. Some languages to be highlighted are Nyquist and Scheme. The first one is focused on the synthesis and sound analysis and serves as a basis for plugins present in Audacity. The second supports functional and procedural programming and builds plugins for Gimp, a tool used to create and edit vector and raster images.

Prolog focuses more on describing facts and their occurrences, rather than describing a sequence of steps to be followed by the computer to solve a problem, just like in traditional programming languages. Thus, this technique provides better communication between human-computer interfaces [23].

These classic techniques are good to process discrete data, like numbers and strings but were not used at that time to process streams of data, like WAV files and bitmap images. At this time it was not easy to think about signal processing and creation using large data, like audio and images. For this reason, in the early days, it was used to combine artificial intelligence with symbolic data representation in music and visual art, such as the MIDI[1] and the SVG[2] format.

4. Neural Networks reaching Arts and Sound Processing

With the advancement of the computer, in addition to the processing of symbolic data, several audios, images or videos have also been processed, expanding the possibilities for artistic creations. These surveys gave rise to several data pre-processing techniques, such as convolution, feature extractors, filters, and audio descriptors, providing support for the rise of the multidisciplinary area called Music Information Retrieval (MIR) that involves areas as Machine Learning and Musical Computing.

Audio is stored in the time domain, which ends up bringing little or no information about its content beyond its respective amplitude (in the time domain) and envelope. One of the ways to get information from audio files is extracting characteristics from an audio in the frequency domain, using for it the audio descriptors. These descriptors can act between two domains, being the time domain, also called basic descriptors, involving the processing of the sampled audio signals, or the frequency domain, using

spectral descriptors, such as the Mel Frequency Cepstral Coefficients [6].

As they are implemented with convolutions, in the form of mathematical equations such as the Discrete Fourier Transform (DFT) and the Fast Fourier Transform (FFT), the audio descriptors end up being part of the data pre-processing step, followed by the application of some machine learning algorithm. In this way, it is possible to use non-convolutional neural networks in conjunction with audio descriptors.

From these descriptors, the audio is reduced to a simplified structure, without loss of information, which can be defined according to the descriptor used. Currently, there is a lot of research in relation to the development of technological tools for the extraction of audio descriptors, both in the form of programming libraries and graphic software or extensions to existing tools aimed at extracting and visualizing the characteristics of audio.

Among the existing tools, we have the Aubio [26] and LibXtract [27] libraries, frameworks like MARSYAS [28], jAudio [29], CLAM Music Annotator [30], jMIR [31] and Sonic Visualiser [32], and the toolbox for MatLab MIRtoolBox [33].

Consequently, more complex databases started to emerge, including, for example, historical images, scientific articles, recordings of artistic performances, among other types of data. In this way, the data started to be stored in different formats such as MP3, AIFF, and WAV for audio and JPEG, RAW and BMP for images.

Currently, regarding audio as data to be processed, there are several databases for different purposes, such as dataset's dedicated to voice recognition tasks such as AudioSet [34] and Common Voice [35]; musical dataset's like Million Song Dataset [36] and BallRoom [37]; or even more specific datasets for classifying instrument tones [38].

On one hand, we had the emergence of artistic datasets and on the other hand, the possibility to process this data with neural networks. However, in addition to the complexity of the data, the entry of neural networks was still unidimensional, which led to several studies on efficient ways to reduce the space of data representation without loss of information.

With neural networks, which use different weights and connections under one architecture, processing has become more complex and the types of data to be processed have been expanded, enabling the creation of several artistic applications. It is worth mentioning that, until then, only symbolic data were used under Production Rules or Decision Trees algorithms, such as CN2 and C4.5, mentioned in [7]. These techniques had the advantage that all the knowledge processed and generated is something understandable for the human being, however, the data used must be symbolic.

In machine learning, several statistical and data mining algorithms are used, focused on extracting characteristics and information from a given database. That is, from a

---

[1]The MIDI protocol was created in 1983 by a consortium of electronic instrument manufacturers, with the aim of standardizing communication between them and also allowing connection to computers. It is important to note that this protocol does not transmit sound, but music information, such as the pitch of a specific note, the time of a song, its pitch, etc. In general terms, it is the equivalent of a score for computers [24].

[2]SVG, is an open file format, which represents drawings and graphics in a vector way. Its main feature is the fact that it maintains the quality of the file when it is enlarged and is present in the representation of vector geometric shapes, raster, and bitmap images and text. Although it did not cause a major revolution on the Web after its introduction, this type of format is present in the most diverse publishing software, both proprietary and free software [25].

certain amount of data, it became possible to train the computer to classify something or even generate new data automatically. In this way, increasingly larger databases have emerged and are being made available to the whole community, such as MIDI libraries containing various artistic performances performed around the world.

5. Special NN models to signal processing

Another way to process digital signals, especially audio, is through the use of convolutional neural networks. As the name implies, this type of network performs convolution and data processing within its structure, causing information to be extracted without the use of descriptors. Thus, these networks can directly receive the file and use less computational resources to learn what kind of processing can be used to extract data from the input value.

5.1. CNN - Convolutional Neural Network

Around 1988, inspired by biological processes [39], Yann LeCun and collaborators [8] created the so-called Convolutional Neural Networks. Initially, the proposal was for these networks to be focused on the processing of one-dimensional or two-dimensional data structures, making CNN's very promising for the processing of images and sounds. It is worth mentioning that the images are composed of pixels that form a 2D structure, but the audio can be treated either as a one-dimensional structure if we analyze the variations of sound waves over time, or two-dimensional if we analyze their respective spectrogram, in which case, he would be treated as an image.

A CNN is a variation of the Multilayer Perceptron Network, but focused on a type of processing quite similar to those of computer vision, enabling the application of filters under the data to be processed, maintaining the relationship between their respective pixels [11]. As the name implies, CNN's operate mainly under a mathematical operation called convolution, which is described as:

$$st \quad x * wt \quad {}_a \quad xawt - a \tag{1}$$

Convolution is a linear operator that is based on the calculation of two functions: the kernel function, also called a filter, described as w; and the sum of the products generated by the function x, referring to the processed data. Regarding the architecture of a CNN, they can be classified as a combination of several layers, which can be classified as Convolutional layers; Pooling layers; and the fully connected layers [40].

The convolutional layers are composed of several neurons, which are connected, according to a certain weight, to a set of pixels provided by the previous layer. These neurons are responsible for applying filters under the captured pixels, through kernel functions, and the result of this processing is then propagated to the next layer.

In this step, the kernel is convoluted together with the input provided, sliding the window, which in this case will move twice, generating two outputs. The variable stride concerns how many pixels will be skipped since the window

will not always move from pixel to pixel, generating a size 2 entry for the next layer, as shown in Figure 1. In addition, it is not always possible for windows to convolve to the end. To do this, exist a technique called *zero-padding* where zeros are added at the edges allowing convolution under all pixels. And after convolution, it is common to apply a nonlinear activation function, like those used in common neural networks.
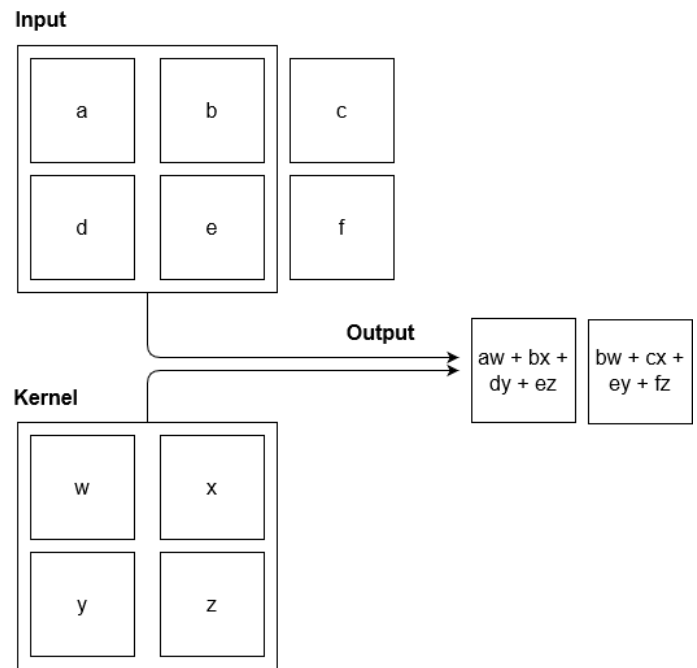

Figure 1: Kernel function example.

An interesting aspect about CNN is the possibility of using more sophisticated filters when compared to the classics used for image processing, which are generally two-dimensional. That is, it is possible to apply a filter that has height, width, and even depth, optimizing the extraction of information provided by the training data.

After the convolution with the activation function, we have the *pooling* layer, responsible for reducing the size of the data. For example, given an entry with 2x2 windows, a value is selected to represent it through a function, which is usually the function called *maxpooling* which selects the maximum value within a rectangular neighborhood.

Currently, there are other more sophisticated architectures such as LeNet-5 [8] which has two convolution layers followed by *pooling* and one more convolution layer or GoogLeNet [41] which has five convolution layers in a row of *pooling* [11].

Finally, after the *pooling* layer, for classification problems, usually one more layer is used, fully connected. This layer is then responsible for making the decisions that the network will take, given the outputs received by the previous layers. In this layer, machine learning algorithms or regressions are generally applied, generating an output called *score*, which will define the classification result.

5.2. GAN - Generative Adversarial Network

First of all, it is important to understand the discriminative and generative algorithms. A discriminative model tries to classify the data from the entry of its label or category, being concerned only with this relationship. The generative model is the opposite of this process. Instead of predicting a label from input data, it tries to predict the resource that led to that label. In mathematical terms, the discriminative algorithm is concerned with the relationship between x and y inputs, while GAN investigates how x value is obtained from an input y [12].

From a competition between these two models, the Generative Adversarial Network emerges. The generative network is responsible for receiving noise as input and generating data samples. The discriminating network receives samples from the generator and the training data and must be able to distinguish what is noise and what is real data. These networks play against each other continuously, where the generative model is learning to produce increasingly realistic samples, and the discriminator is learning to improve more and more in the distinction of data. The purpose of all this is to make the generated samples indistinguishable from the actual data [13].

Training involves presenting samples of the training dataset until acceptable accuracy is achieved. The generator trains based on its ability to "trick" the discriminator. Typically, the generator is seeded with random input that is sampled from a predefined latent space (for example, a multivariate normal distribution). Subsequently, the candidates synthesized by the generator are evaluated by the discriminator. Backpropagation is applied to both networks, so that the generator produces better images, while the discriminator becomes more adept at signaling synthetic images. The generator is typically a deconvolutional neural network and the discriminator is a convolutional neural network [12].

Although originally proposed as a form of a generative model for unsupervised learning, GANs have also proved useful for semi-supervised learning, fully supervised learning and reinforcement learning [14].

Consequently, the potential of GANs allows it to mimic any data distribution, creating images from scratch, using all possible combinations of pixels in a figure, in addition to songs, poetry, etc. This type of network can be found in fashion and advertising, simulating models, makeup and products [42]; in science, helping to improve astronomical images [43]; and digital art, contributing to the modeling of characters and scenarios [14].

5.3. AE - Autoencoder

Autoencoder is a type of unsupervised neural network that uses backpropagation to copy the input and display it as an output, to learn a new representation of the input data. In mathematical terms, he learns an approximation of the identity function, generating a function $f(x')$ similar to $f(x)$ [15].

This type of network is divided into two parts: the encoder, responsible for encoding the input data, represented by the function $f(x)$, and the decoder, a function $g(x)$ that reconstructs that data [15].

Because it generates a copy of the input data, the autoencoder must be able to generate useful properties. This can be achieved through the so-called incomplete autoencoder, which creates restrictions on the copy task and forces the network to learn the most important features of a given data [16].

A characteristic that differentiates this method from the others is its ability to offer several outputs that enhance the model. One is to add a sparse condition to the weights, which reduces the size of the latent vector. Another way of improvement is to allow the encoder and decoder to be deep neural networks. So, instead of trying to find linear transformations, the network will be concerned with finding nonlinear information [16].

Currently its common application in noise removal systems, dimensionality reduction in data visualization and in the acquisition of semantic meaning in words. In music, it is widely used for compression and reconstruction of musical information, in addition to being able to be used as a denoising filter, common in music production.

5.4. VAE - Variational Autoencoder

Variational Autoencoder, as the name suggests, has some relationship with Autoencoder, such as the architecture composed of an encoder and a decoder, trained to minimize the reconstruction error between the initial data. In a nutshell, VAE is an Autoencoder whose encodings are regulated during training, ensuring that the latent space has good properties for generating new data. The term "variational", in turn, comes from the relationship between regularization and the methods of variational inference in statics [17, 18].

However, there are also significant differences, the main one being the fact that Autoencoders are deterministic discriminative models, while VAEs are generative models. This happens because of the different mathematical formulations addressed by both. Variational Autoencoder is directed probabilistic graphical models (DPGM), approximated by a neural network, that tries to simulate how the data is generated, and from that, understand the underlying causal relationships. Therefore, instead of creating an encoder that generates a single value to describe each latent state attribute, an encoder is formulated to describe a probability distribution for each attribute.

In contrast to the most common uses of neural networks as regression or classification, VAEs are powerful generating models. This is because it is allowed to generate a new random output, similar to the training data, in addition to changing or exploring variations in the data more frequently, following a specific and desired direction. This is where VAEs work better than any other method currently available.

The framework of variational autoencoders (VAEs) provides a principled method for jointly learning deep latent-variable models and corresponding inference models using stochastic gradient descent. The framework has a wide array of applications from generative modeling, semi-supervised learning to representation learning and inference models.

Among the applications for this tool are the generation of fake human faces; recognition of manuscripts; interstellar photographs; and the production of purely synthetic music.

When dealing with music specifically, MusicVAE [44] emerges, where its main characteristic is to be hierarchical in the learning of latent spaces for musical scores. The tool uses a combination of recurrent neural networks (RNN) and VAE that encodes a musical sequence into a latent vector, which can later be decoded into a musical sequence. As latent vectors are regularized to be similar to a standard normal distribution, it is also possible to sample from the sequence distribution, generating realistic music based on a random combination of qualities [45, 46].

This is possible through the interpolation of the music, that is, through the mixing of different musical sequences, or through the vector arithmetic of attributes, which adds certain features to the music. As a creative tool, the goal is to provide an intuitive palette with which a creator can explore and manipulate the elements of an artistic work [47].

6. Looking forward: Some perspectives and inconclusive topics

From a technical point of view, the use of artificial intelligence has helped in various ways the production of art. In music, examples range from the classification of the musical genre through the combination of information contexts [48] or through the classification by relational algorithms [49], to the creation of hierarchical systems for recommending music [50] and intelligent systems that propose audio plugins to assist in musical production [51].

In addition to also occupying space in the artistic area and making several transformations in this field, it also covered the use of different technologies to create a (no more) new type of art, called digital art. With the emergence of several discussions between researchers and artists about this artistic mode, digital art is increasingly present in several places, such as museums (Uncanny Mirror, Museum of Modern Art and Barbican Center), auction houses (Christie's and Sotheby's), or even at certain scientific events.

A very successful work that unites artificial intelligence with the artistic world is "ToTa Machina", developed by Katia Wille. This work is an installation that captures the emotions of the audience using facial recognition techniques and, through robotics, has visual projections that move according to the audience. Through facial stimuli, the relationship between the human being and the machine ends up becoming even shorter, reinforcing engagement and appreciation under the work [52]. An interesting fact that reinforces even more the promise of technological installations, is that during the presentations of the work, the public stayed a long time in the installation, where the children were curious to know more about the structure of the installation and the adults interacted creating different affections for obtaining different results.

Another more current example of neural networks aimed at artistic production is the work "Edmond de Belamy", created in 2018 by the Obvious group, formed by three researchers from the University of Montreal who used a GAN neural network, explained in subsection 5.2, to elab-

orate this art work [53]. The work consists of a set of 11 portraits generated by the neural network, portraying members of the fictional Belamy family. This work was very successful, being auctioned for US$432.500,00, reinforcing the idea that the computer is capable of being a tool for artistic productions [54].

The inclusion of this work caused several discussions involving both artists and researchers in the field of AI who wondered about the possibility of algorithms by themselves, being artists. Mark Riedl, associate professor of machine learning and AI at the Georgia Institute of Technology, classifies the algorithms as "very complicated brushes with many mathematical parameters, which make it possible to create an effect that would be difficult to obtain otherwise".

It is worth mentioning that GAN's are networks that need a large amount of data to obtain a good result and all the knowledge acquired is then used to generate new results. That is, the final product is directly linked to a long process of data selection, mathematical parameters, and selection of the results obtained. In this way, we cannot say that the machine is fully responsible for the work generated since the researcher himself became an artist because he was responsible for all the elaboration of the parameters and the architecture used during the creative process. Then the question arises as to who the real artist would be, the algorithm for being trained for months to generate a certain work or the researcher carrying out various tests and adjustments to obtain a certain result?

Under this question, like any work of art, we must analyze the entire creation process and not just the final product. Thus, it is worth mentioning the thinking of three contemporary artists: Anna Ridler, a British artist who used a GAN relating the volatility of the cryptocurrency market with Tulip Fever, who understands AI as a tool, as it allows specific creations, and also as process, since the artist needs to label objects and model the data to be used; Gauthier Vernier, french artist who participates in the Obvious group, mentioned above, who has a less human thinking in the sense of who is responsible for art, providing credit to the machine and signing the work with the algorithm's own formula, because according to the author, it was him even who created the work; and Mario Klingemann, a German artist pioneering the use of artificial intelligence, says that what the machine does to create a work of art is basically imitating human aspects, and under current technology, it ends up failing to try to interpret subconscious and emotional aspects that many sometimes difficult to quantify [55].

Some discussions precede the question about who the artist would be in a work developed by AI, such as the definition of whether the result obtained is an artistic work or not. There are people with more futuristic thinking like the gallery owner Luisa Strina who defends the idea that to have quality, art must not follow trends. She further states "Used with conceptual and aesthetic coherence, new technologies open a door for research that has not yet been done in art" [52]. Neural networks are new techniques if analyzed in the artistic field, and because they are not fully automated, since the programmer needs to adjust it for certain results, they end up opening space for the insertion of

aesthetic and conceptual concepts during the creation stage, making sure that the final result can be directly related to the artist/programmer's intention. Consequently, neural networks end up having a great potential for the creation of several quality artworks, bringing us one more question raised by several artists. But concerning art creation, will machine replace artists in the future?

Heretofore, It is worth mentioning three thoughts: Anna Ridler states that people pay a large amount of money for simple and minimalist work, but if we analyze works in which there is a lot of repetition, perhaps the AI will optimize these processes, facilitating the creation and perhaps making the work cheapest; Gauthier Vernier states that the machine will not replace man, given that the human takes care of much of the creative process and it is he who provides purpose for what was created, that is, the machine is not capable of creating a work of art by itself, human/machine collaboration is necessary; finally, Mario Klingemann addresses the idea that for artists who perform repetitive works, AI has great potential to replace them in the future [55].

Finally, in this section were presented thoughts on the definition of what would be art, in addition to the presentation of some works that involve the use of neural networks for artistic creations. However, the fact is that, defining or not, digital art has evolved more and more, in addition to generating both financial and academic results. In this way, it is interesting to think not only about new ways or techniques for creating innovative artistic works but also about what type of art people are seeing, what tool is used or even what type of music is created due to the concept of what it is popular.

7. Final Remarks

Maybe, the digital arts are as old as the computer, and until then, several artistic works have been created, in the form of interactive web pages, pictures painted by machines, installations that capture the interaction of the public to synthesize images and sounds, among others. The discussion about what is art is quite old and pertinent, however, willingly or not, artists are creating digital artworks and also a public that admires this type of work.

This market has grown considerably in recent years, which makes the creation of digital works very promising, in addition to increasingly engaging researchers to develop techniques that provide to the programmer, more autonomy under a neural network, bringing benefits and innovations not only under the arts but also machine learning in general.

Taking into account the point of view that work, to be artistic, must have an entire aesthetic concept, which involves both the creative process and the final result, the artistic definition under work is exclusively up to its creator. As previously mentioned, when using a neural network, the programmer needs to define the entire network structure, referring to the number and type of layers, parameters to be used, and even what would be the type of input.

In this way, we can affirm that the programmer is essential for the creative process and that perhaps the work itself is not art, but rather the entire creation process involving the choices and decisions that the programmer had to carry out for the development of the work. It may be possible that in the future, digital artists will start to be recognized not for their works, but for all the creative processes created for the creation of art, causing auction houses, for example, to sell algorithms and not just concrete works.

The choice of algorithms and techniques throughout this text is due to the fact that they are widely disseminated in the academic community, in addition to highlighting that even the oldest technologies can be used in the relationship between art and AI.

Having exposed this long-standing relationship and contribution of Artificial Intelligence in art, future strategies should be considered. The technological advance of the latest techniques has provided the emergence of new algorithms and software that support the artistic practice mediated by computers, in addition to making this process cheaper and popular. In addition, fields such as STEAM (Science, Technology, Engineering, Arts and Mathematics) began to gain more space, being an excellent tool to explore both areas.

Acknowledgment

References

[1] E. D. Liddy, "Natural language processing", *Encyclopedia of Library and Information Science*, 2001.

[2] B. Buchanan, "A (very) brief history of artificial intelligence.", *AI Magazine*, vol. 26, pp. 53–60, 2005.

[3] R. L. de Mántaras, "Artificial intelligence and the arts: Toward computational creativity", https://www.bbvaopenmind.com/en/articles/artificial-intelligence-and-the-arts-toward-computational-creativity/, 2020-05-17.

[4] G. Kogan, "Machine learning for artists", https://medium.com/@genekogan/machine-learning-for-artists-e93d20fdb097#.xc9qk7ca7, 2020-05-17.

[5] J. Mccarthy, "History of lisp", *ACM SIGPLAN Notices*, vol. 13, 1996, doi:10.1145/960118.808387.

[6] P. Hamel, D. Eck, "Learning features from music audio with deep belief networks.", "ISMIR", vol. 10, pp. 339–344, Utrecht, The Netherlands, 2010.

[7] E. Martineli, "Extração de conhecimento de redes neurais artificiais.", Ph.D. thesis, Universidade de São Paulo, 1999.

[8] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[9] S. Naveed, G. G, L. S, "Early Diabetes Discovery From Tongue Images", *The Computer Journal*, 2020, doi:10.1093/comjnl/bxaa022, bxaa022.

[10] S. Naveed, G. Mohan, "Intelligent diabetes detection system based on tongue datasets", *Current Medical Imaging Reviews*, vol. 14, 2018, doi:10.2174/1573405614666181009133414.

[11] A. C. G. Vargas, A. Paes, C. N. Vasconcelos, "Um estudo sobre redes neurais convolucionais e sua aplicação em detecção de pedestres", "Proceedings of the XXIX Conference on Graphics, Patterns and Images", vol. 1, 2016.

[12] D. S. Academy, "Introdução às redes adversárias generativas (gans – generative adversarial networks)", http://deeplearningbook.com.br/introducao-as-redes-adversarias-generativas-gans-generative-adversarial-networks/, 2018, 2020-05-18.

[13] J. Glover, "An introduction to generative adversarial networks", https://blog.aylien.com/introduction-generative-adversarial-networks-code-tensorflow/, 2018, 2020-05-18.

[14] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, X. Tang, "Esrgan: Enhanced super-resolution generative adversarial networks", "Proceedings of the European Conference on Computer Vision (ECCV)", pp. 0–0, 2018.

[15] G. Hinton, A. Krizhevsky, S. Wang, "Transforming auto-encoders", "21st International Conference on Artificial Neural Networks", vol. 6791, pp. 44–51, 2011.

[16] J. Jordan, "Introduction to autoencoders", https://www.jeremyjordan.me/autoencoders/, 2018, 2020-05-18.

[17] D. Kingma, M. Welling, "An introduction to variational autoencoders", *Foundations and Trends® in Machine Learning*, vol. 12, pp. 307–392, 2019, doi:10.1561/2200000056.

[18] X. Li, J. She, "Collaborative variational autoencoder for recommender systems", "23rd ACM SIGKDD International Conference", pp. 305–314, 2017, doi:10.1145/3097983.3098077.

[19] M. Kauw-A-Tjoe, "Generation of abstract geometric art based on exact aesthetics, gestalt theory and graphic design principles.", *Journal of Computational and Applied Mathematics - J COMPUT APPL MATH*, 2005.

[20] A. Organization, "Timeline of ai art", https://aiartists.org/ai-timeline-art, 2020-05-17.

[21] G. Assayag, C. Rueda, M. Laurson, C. Agon, O. Delerue, "Computer-assisted composition at ircam: From patchwork to openmusic", *Computer Music Journal*, vol. 23, no. 3, pp. 59–72, 1999.

[22] J. Bresson, C. Agon, "Visual programming and music score generation with openmusic", "2011 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)", pp. 247–248, IEEE, 2011, doi:10.1109/VLHCC.2011.6070415.

[23] A. Colmerauer, P. Roussel, "The birth of prolog", "History of programming languages—II", pp. 331–367, 1996, doi:10.1145/234286.1057820.

[24] E. M. Miletto, L. L. Costalonga, L. V. Flores, E. F. Fritsch, M. S. Pimenta, R. M. Vicari, "Introdução à computação musical", "IV Congresso Brasileiro de Computação", sn, 2004.

[25] J. D. Eisenberg, *Svg Essentials*, O'Reilly Media, 2014.

[26] P. M. Brossier, "Automatic annotation of musical audio for interactive applications", Ph.D. thesis, University of London, 2006.

[27] J. Bullock, U. Conservatoire, "Libxtract: a lightweight library for audio feature extraction.", "ICMC", 2007.

[28] G. Tzanetakis, P. Cook, "Marsyas: A framework for audio analysis", *Organised sound*, vol. 4, no. 3, pp. 169–175, 2000.

[29] C. McKay, R. Fiebrink, D. McEnnis, B. Li, I. Fujinaga, "Ace: A framework for optimizing music classification.", "ISMIR", pp. 42–49, 2005.

[30] X. Amatriain, J. Massaguer, D. Garcia, I. Mosquera, "The clam annotator: A cross-platform audio descriptors editing tool.", "ISMIR", pp. 426–429, 2005.

[31] C. McKay, I. Fujinaga, "jmir: Tools for automatic music classification.", "ICMC", 2009.

[32] C. Cannam, C. Landone, M. Sandler, "Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files", "Proceedings of the 18th ACM international conference on Multimedia", pp. 1467–1468, ACM, 2010.

[33] O. Lartillot, P. Toiviainen, T. Eerola, "A matlab toolbox for music information retrieval", "Data analysis, machine learning and applications", pp. 261–268, Springer, 2008.

[34] J. F. Gemmeke, "Audioset", .

[35] M. Team, "Common voice", .

[36] L. Rosa, "Million song dataset", .

[37] F. Gouyon, "Ball room", .

[38] N. Team, "Nsynth dataset", .

[39] M. Matsugu, K. Mori, Y. Mitari, Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network", *Neural Networks*, vol. 16, no. 5-6, pp. 555–559, 2003.

[40] H. R. Guimarães, "Recuperação de informações musicais: Uma abordagem utilizando deep learning", 2018.

[41] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions", "Proceedings of the IEEE conference on computer vision and pattern recognition", pp. 1–9, 2015.

[42] C. Wong, "The rise of ai supermodels", https://www.cdotrends.com/story/14300/rise-ai-supermodels, 2020-05-18.

[43] K. Schawinski, C. Zhang, H. Zhang, L. Fowler, G. K. Santhanam, "Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit", *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 467, p. slx008, 2017, doi:10.1093/mnrasl/slx008.

[44] A. Roberts, J. Engel, C. Raffel, I. Simon, C. Hawthorne, "Music-vae: Creating a palette for musical scores with machine learning", https://magenta.tensorflow.org/music-vae, 2020-04-21.

[45] W. Sharber, "Musicvae: A tool for creating music with neural networks", https://medium.com/@wvsharber/musicvae-a-tool-for-creating-music-with-neural-networks-db0f4b84a698, 2020-04-20.

[46] M. Dinculescu, J. Engel, A. Roberts, eds., *MidiMe: Personalizing a MusicVAE model with user data*, 2019.

[47] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, D. Eck, "A hierarchical latent vector model for learning long-term structure in music", *International Conference on Machine Learning*, 2018.

[48] M. Domingues, S. Rezende, "The impact of context-aware recommender systems on music in the long tail", "Brazilian Conference on Intelligent Systems, BRACIS 2013", pp. 119–124, 2013, doi:10.1109/BRACIS.2013.28.

[49] J. Valverde-Rebaza, A. Soriano Vargas, L. Berton, M. C. Oliveira, A. Lopes, "Music genre classification using traditional and relational approaches", "Brazilian Conference on Intelligent Systems, BRACIS 2013", 2014, doi:10.1109/BRACIS.2014.54.

[50] I. Nunes, G. Leite, D. Figueiredo, "A contextual hierarchical graph model for generating random sequences of objects with application to music playlists", "Brazilian Conference on Intelligent Systems, BRACIS 2013", 2019.

[51] P. Moura da Silva, C. Mattos, A. Júnior, "Audio plugin recommendation systems for music production", "Brazilian Conference on Intelligent Systems, BRACIS 2013", 2019, doi:10.1109/BRACIS.2019.00152.

[52] B. Calais, "A inteligência artificial invade o mundo da arte", https://forbes.com.br/forbeslife/2020/01/a-inteligencia-artificial-invade-o-mundo-da-arte/, 2020, 2020-05-25.

[53] D. Kaufman, "Dá pra fazer arte com inteligência artificial?", https://epocanegocios.globo.com/colunas/IAgora/noticia/2019/08/da-pra-fazer-arte-com-inteligencia-artificial.html, 2020, 2020-05-25.

[54] J. Pearson, "Uma obra de arte gerada por ia foi vendida num leilão por us$ 432 mil", https://www.vice.com/pt_br/article/43ez3b/uma-obra-de-arte-gerada-por-ia-foi-vendida-num-leilao-por-usdollar-432-mil, 2020, 2020-05-25.

[55] F. Gonzaga, M. Fernandes, "Inteligência artificial é o próximo passo na evolução da arte?", https://arte.estadao.com.br/focas/estadaoqr/materia/inteligencia-artificial-e-o-proximo-passo-na-evolucao-da-arte, 2020, 2020-05-25.

computer networks. Laureate with first place in the "Wikipedia Challenge", at the 18th Brazilian Symposium on Computer Music, acting in the dissemination of the theme. Flamengo fan and full-time anti-fascist, he practices music and skateboarding in his spare time.

FLÁVIO LUIZ SCHIAVONI Graduated in Computer Science from UEM (1999), specialization in Web Development from UEM (2004), Master in Computer Science from UEM (2007) and Ph.D. in Computer Science from IME/USP (2013).

He is a professor at the Federal University of São João del-Rei and serves as professor in the Graduate Program in Computer Science (PPGCC) and in the Interdepartmental Interdisciplinary Graduate Program in Arts, Urbanities and Sustainability (PIPAUS) of the same institution where coordinates ALICE (Arts Lab in Interface, Computers, and Else) and the Orchidea Collaborative Art Group (Orchestra of Ideas).

RÔMULO AUGUSTO VIEIRA COSTA Master in Computer Science from the Federal University of São João del-Rei, he also holds a bachelor's degree in Control and Automation Engineering. Research focused on the Internet of Musical Things (IoMusT), new interfaces for musical expression, participatory art, and