Sarah Pires Pérez

Improving Art Style Classification with Synthetic Images from Self-Attention Generative Adversarial Networks

São Paulo

2022

Sarah Pires Pérez

Improving Art Style Classification with Synthetic Images from Self-Attention Generative Adversarial Networks

Revised Version

Master thesis presented to the Escola Politécnica da Universidade de São Paulo to obtain the Master of Science degree.

Concentration Area:
Computer Engineering

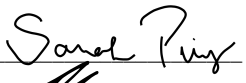
Supervisor:
Prof Dr Fábio Gagliardi Cozman

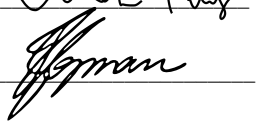
São Paulo

2022

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, 06 de julho de 2022

Assinatura do autor: 

Assinatura do orientador: 

Catalogação-na-publicação

Pérez, Sarah
Improving Art Style Classification with Synthetic Images from Self Attention Generative Adversarial Networks / S. Pérez -- versão corr. -- São Paulo, 2022.
67 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Computação e Sistemas Digitais.

1.Neural Network 2.Image Recognition 3.Intelligent Agents I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Computação e Sistemas Digitais II.t.

"It is only by drawing often, drawing everything, drawing incessantly, that one fine day you discover, to your surprise, that you have rendered something in its true character."

Camille Pissarro

ABSTRACT

Art is the means by which humanity has always expressed itself, as art offers a record of humanity's feelings, its ways of life and its conception of the world. Although we are fortunate to have a vast store of cultural wealth from past generations, the sheer number of artworks has become an obstacle to their categorization into styles. This research explores a strategy that maximizes the performance of style classifiers applied to works of art. Automatically classifying artworks into styles is quite challenging due to the relative lack of tagged data and the complexity of the class definitions. This complexity is manifested by the fact that some image augmentation techniques not only do not improve performance but may also degrade performance. We propose to resort to Adversary Generating Networks (GANs). Originally, GANs set out to create images capable of deceiving the human eye and making us believe that generated images are true images. The proposal here is not to create art, but rather to use this architecture as a data augmentation tool. To assess the impact of using GANs on image augmentation, we have studied performance improvements over EfficientNet B0, a state-of-the-art image classifier. In addition, we present a Class-by-Class Performance Analysis that can be useful in the study of other high-complexity image datasets.

Key-words: Computer Vision, Generative Adversarial Networks, Image Classification

RESUMO

Arte é o meio pelo qual a humanidade sempre usou para se expressar, tornando-a um registro de seus sentimentos, seus modos de vida e sua concepção de mundo. No entanto, embora tenhamos a sorte de ter uma vasta riqueza cultural proveniente de várias gerações, a quantidade de obras de arte tornou-se um impedimento para sua categorização em estilos. Esta pesquisa se propõe a estudar uma estratégia para maximizar o desempenho dos classificadores de estilo em obras de arte. A classificação automática das obras de arte em seus estilos é bastante desafiadora devido à relativa falta de dados rotulados e à complexidade das classes envolvidas. Essa complexidade é refletida no fato que algumas técnicas de augmentação de imagens não só não agregam ao desempenho do modelo mas também podem degradar seu desempenho. Por isso, introduzimos neste trabalho o estudo de Redes Adversárias Geradoras (GANs). Originalmente, as GANs foram propostas para criar imagens capazes de enganar o olho humano e nos fazer acreditar que as imagens geradas são imagens verdadeiras. Essa pesquisa não se propõe a criar arte, mas pretende usar essa arquitetura como uma ferramenta de ampliação de dados. Para avaliar o impacto do uso de GANs na augmentação das imagens, treinamos a EfficientNet B0 para verificar a melhoria no desempenho do EfficientNet B0, um classificador de última geração. Além disso, apresentamos a Análise de Desempenho de Classe por Classe, que deve ser útil no estudo de outros conjuntos de imagens de alta complexidade.

Palavras-chave: Visão Computacional, Redes Adversárias Geradoras, Classificação de Imagens

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS OR ABBREVIATIONS

CNN	*Convolutional Neural Network*
GAN	*Generative Adversarial Network*
Conv	*Convolutional layer*
Pool	*Pooling layer*
FC	*Fully Connected layer*
GeoAug	*Geometric Augmentation*
CCPA	*Class-by-Class Performance Analysis*
EffB0	*EfficientNet B0*
ILSVRC	*ImageNet Large Scale Visual Recognition Challenge*
GPU	*Graphics Processing Unit*

CONTENTS

1 INTRODUCTION

When Frida Kahlo says “I paint flowers so they will not die.”, she highlights the perpetuity of art. In it, life, thought, and feeling are diligently recorded in order to convey the truths of the human condition that transcend the capacity of words to describe. In efforts to understand artistic records throughout history, one of the most important issues has to do with the age of an artwork. If researchers cannot determine the age of a monument, they cannot place the work in its historical context. To remedy this, scholars rely on four types of evidence: physical, documentary, internal and stylistic.

Physical evidence refers to the age of the material and the employed technique, making it possible to define the earliest possible date and the latest possible date of creation. Documentary evidence refers to writings that in the past could have served as a service contract between the artist and their patron. Internal evidence is found by analyzing the integral elements of the artwork, ranging from a depiction of a famous person to a hairstyle from a specific era. Stylistic evidence refers to the artist’s specific way of creating art (GREEN et al., 2011). The latter is the best-known way of categorizing art, which gives a great amount of information about their ideas and intentions for academic purpose. Furthermore, art style classification goes beyond the domain of art scholars and serves as an important guide for beginning art aficionados and the general public within museums and galleries, not only providing context but also for recommendation tools to improve a museum visitor’s experience.

Despite its importance, the categorization of art style became an almost impossible task given the vast cultural richness we have inherited from our ancestors. In order to prevent this quantity from being an impediment to the appreciation of art in an organized way, this research strives to maximize the performance of the art style classification. Furthermore, given the complexity of the challenge, the search for solutions to the challenge can lead to unexpected solutions that could be used in other contexts.

Art style classification falls within image classification in Computer Vision. Automatic artwork style classification was initially approached with pattern recognition techniques and machine learning algorithms (SHAMIR et al., 2010; ARORA; ELGAMMAL, 2012). The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) enabled the development of deep learning techniques for image classification (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). As a result, not only did the specific architectures for image classification evolve, but models trained in the competition became the starting point for classifiers for other domains, including the art domain. Another important milestone for artwork

studies within Computer Vision was the creation of the WikiArt dataset.[1] This dataset offers a rich source of labeled artwork, and it has become a standard dataset for art style classification. Figure 1 shows a sample of the WikiArt dataset.

Analyzing the literature on this problem, we see that the complexity of the art dataset has been mentioned as a challenge to overcome. Beyond the challenge of class imbalance, style classification is further complicated by aesthetic diversity, a problem that is not usually met in most image classification models. Some artworks, such as abstract expressionist paintings (Figure 1a), have neither traditional subjects nor even settings. The "subject" is the artwork itself – its colors, textures, composition, and size. On the other hand, Romanticism (Figure 1n) emerged in Europe in the 18th century and its paintings reflect the context of the industrial revolution and the Enlightenment, which was an intellectual and philosophical movement based on reason.

Many efforts have been made in order to maximize the information obtained from artwork images, in approaches based on data augmentation. However, none of these efforts have relied on Generative Adversarial Networks (GANs). Augmentation based on GANs aims to generate label preserving images in the classifier domain; it is a novel way, proposed here, to add information to a classifier's training and thus to improve its performance.

In short, this research aims to maximize the performance of art style classifier. To this end, we present a study of the use of GANs in image augmentation, in order to add more information about styles in the training phase of a classifier. As a result of this study, we present the Class-by-Class Performance Analysis strategy that aims to organize the knowledge developed in this research and to allow extrapolation to other contexts in which data from a high complexity domain is also involved. These contributions have been described in the following article:

- Pérez S.P., Cozman F.G. (2021) **How to Generate Synthetic Paintings to Improve Art Style Classification**. In: Britto A., Valdivia Delgado K. (eds) Intelligent Systems. BRACIS 2021. Lecture Notes in Computer Science, vol 13074. Springer, Cham.

In Chapter 2, we provide the required background knowledge, followed by related work regarding image classification and Generative Adversarial Networks in Chapter 3. Then, in Chapter 4, we discuss our contributions, followed by our experiments and results in Chapter 5. Finally, in Chapter 6 we present our conclusions, future work and timeline.

[1] ⟨https://www.wikiart.org/⟩

(a) Abstract Expressionism

(b) Art Nouveau

(c) Baroque

(d) Color Painting Field

(e) Cubism

(f) Early Renaissance

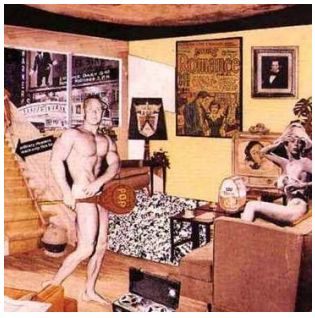
(g) Expressionism

(h) Impressionism

(i) Minimalism

(j) Naive Art

(k) Northern Renaissance



(l) Pop Art

(m) Realism

(n) Romanticism

(o) Ukiyo-e

Figure 1 – An example of each of the art movements in the WikiArt dataset.

2 BACKGROUND

This chapter provides necessary knowledge to understand this research. It starts with an introductory explanation of the artistic styles covered in this work. Subsequently, we summarize the techniques behind modern algorithms of image classification, with special attention to the classification network used in this research, the EfficientNet B0. Afterward, we present some important concepts in image oversampling and image augmentation. The theory behind how GANs work is also presented. Finally, the concepts of Self-Attention and Wasserstein with Gradient Penalty loss function are shown and its relation to GANs is explained.

2.1 Art Style

According to Green et al. (2011), art style is an artist's distinctive manner of producing an object. Once the artist's style has been defined, a very challenging task in itself, one seeks to fit his work into predefined styles. An Art Style can be specific to a span of years – also known as Period Style – such as Early Renaissance. However, there have been many periods when stylistic variations were linked to geography, and that leads these styles to be taken as Regional Styles. Also, a Personal Style may be contemplated, since many artists end up developing a unique method of producing artistic elements. Furthermore, an artist's Personal Style may change dramatically during their life time and the artworks of some artists end up distributed through different styles.

The following sections cover the artistic movements of interest to this research (in alphabetic order). In addition to the aforementioned reference, explanations about the context and characteristics of the movements are based on The Art Story Guide.[1]

Abstract Expressionism

Originated in New York, Abstract Expressionism (1943-1965) was an artistic movement that reflected a post-war mood filled with anxiety and trauma. Visually, it can be identified by color-centered paintings with abstract forms. These abstract forms may be originated from the dripping technique of Jackson Pollock. The canvas is huge to match the size of the statement, filled with profound emotions and universal themes originated from Surrealism.

[1] ⟨https://www.theartstory.org/⟩

Art Nouveau

Art Nouveau (1890 - 1905) expressed the desire to modernize design and to evolve traditional art that took painting and sculpture as superior to craft-based decorative arts. Artists chose to emphasize linear contours and shapes that highlight elegance, also to help narrow the gap between what as considered fine art and applied art. Colors took a secondary role and they were restricted to muted green, browns, yellows and blues.

Baroque

Baroque (1584 - 1723) has its history deeply connected with the Protestant Reformation. As the Protestant Reformation criticized the Catholic Church, the latter used art as propaganda to its importance and grandeur within society in order to awaken religious fervor. Therefore, it is not surprising that Baroque is completely dedicated to the religious theme. The religious representation were easy to understand and they should impress the audience with a complex use of light and rich ornaments to induce the feeling of something elevated and sacred.

Color Field Painting

Similar to Abstract Expressionism, Color Field Painting (1940 - 1960) had color as the main character, avoiding the suggestion of form or mass standing out in the canvas. Clyfford Still painted fields of colors in opposition: light and dark, that the painter would reference as “life and death merging in fearful union”.

Cubism

Cubism (1907 - 1922) completely changed the Renaissance depiction of space, using multiple vantage points to fracture images into geometric forms. The evolution of the style is the result of the partnership between Picasso and Braque, making it difficult to distinguish their work over time. They focused in the genres of portraiture and still life with earth tones and muted gray.

Early Renaissance

Early Renaissance (1401 - 1490) is the art developed in Italy during the XV century. The art was mostly influenced by the Humanist philosophy, in which the individual is considered the center and their relation with God does not belong to the Church. Technically, this style presented the one point linear perspective, which indicated the importance given to knowledge of mathematics and architecture.

Expressionism

Born in Germany, Expressionism (1905 - 1933) was a response to a feeling of lack of authenticity and spiritually. Distorted form, strong colors and exaggeratedly executed brushstrokes were considered a way of expressing the widespread anxiety and the feelings of the overwhelmed individual dealing with urbanization and capitalism.

Impressionism

Impressionism (1862 - 1892) is ground-breaking art movement for the modern painting, due to its disruptive change in thematic and technique. Painters were now focused on painting mundane subjects and their impressions of it, with looser brushwork and lighter colors. Furthermore, they sought to portray the world exactly as it was perceived by them, depicting imperfections, but, at the same time, not necessarily portraying reality.

Minimalism

Minimalism (1960-1969) can be defined as the denial of expression. These painters aimed to distance themselves from the Abstract Expressionism excesses and the perception of references or metaphors of any kind. Geometric forms were used not only to force the audience to confront arrangements and scale, but also to break down traditional notions of sculpture and to erase distinctions between painting and sculpture.

Naïve Art

Naïve Art (1890 - 1945) carried the "noble savage" idea, in which the primitive man is considered good and it is t he society that turns him into a corrupted being. Besides the innovative aesthetic, it is important to highlight the critique of modernization of the western society.

Northern Renaissance

Northern Renaissance (1430 - 1580) was also influenced by the Protestant Reformation, leading the painters to disdain grand idealizations of the Catholic Church. Jan van Eyck learned the technique of linear perspective of the the Italian Renaissance, however he applied to realistic and everyday figurative characters in his paintings. While art in Italy was, at that time, for the rich and powerful, art in northern Europe was for the bourgeois.

Pop Art

Following the popularity of Abstract Expressionism, Pop Art (1950 - 1970) introduced the art mixed with popular imagery. These artists aimed to transform the images of

advertisement and cartoons into high art, which earned them the criticism of being in agreement with capitalism.

Realism

Realism (1840s - 1880s) is considered the first modern movement in art, in which painters replaced the idealistic images and literary concepts of traditional art with real-life events, giving to outcasts similar weight as to grand history paintings and allegories. Their choice to bring everyday life into their canvases was an early manifestation of the avant-garde desire to merge art and life; their rejection of pictorial techniques, like perspective, prefigured the many 20th-century definitions and redefinitions of modernism.

Romanticism

Romanticism (1780 - 1830) follows the idealism of the French Revolution, as artists felt that they lacked freedom. The culprit for this feeling? The Enlightenment and its quest for reason were considered an affront to artists. As a form of direct combat, scenes of protests and revolts were painted. Others, more indirectly, embraced the connection between the individual and their deepest feelings instead of any kind of reason. Furthermore, the historical events of class struggles spawned nationalist sentiments that inspired artworks emphasizing folklore, tradition and local nature.

Ukiyo-e

Ukiyo-e (1672 - 1880s) is the famous Japanese painting and woodcut style for the Edo period, aiming at a representation of the leisure districts of cities. The name of this style translates to “images of the floating world”, an ancient Buddhist term describing the transience of human life and the ephemeral nature of the material world. These idyllic narratives not only document the leisure activities and atmosphere of the time, but also portray the decidedly Japanese aesthetic of beauty, nature and spirituality.

2.2 Image classification

An image can be viewed as a matrix. The first neural network architectures demanded that, when an image was the input, the image should be transformed to a flat tensor so as to be accepted as input. Therefore, notions of spatial distribution were lost, and the quality of image classifiers was penalized. In Convolutional Neural Networks (CNNs), each neuron of a network receives input from a region of the previous layer, while a neuron of a fully connected layer receives input from every element of the previous layer. Figure 2

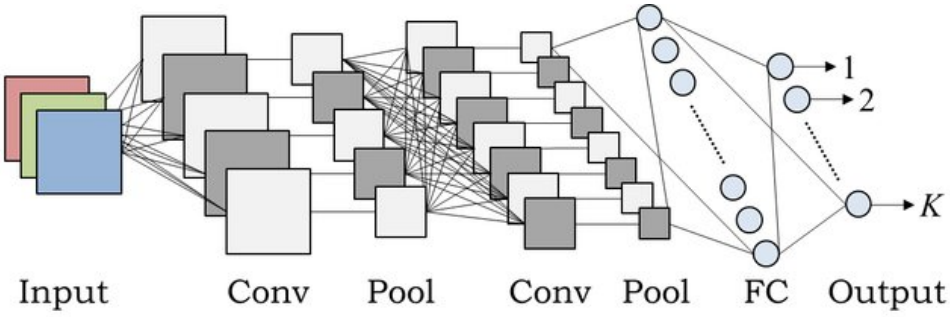


Figure 2 – An example of a CNN (HIDAKA; KURITA, 2017).

shows an architecture with convolutional layers. Only the last layer is fully connected so as to act as a classification layer.

The history of CNNs is deeply connected with the ImageNet Large Scale Visual Recognition Challenge, an annual competition in which the main goal is to correctly classify and detect objects and scenes. The winner of 2012 was Krizhevsky, Sutskever & Hinton (2012) with AlexNet, a network with eight convolutional layers and three fully connected layers. Since then, several larger and more complex architectures have been created to win the ILSVRC. For this research, we chose an architecture from the same family of the winner of the 2019 ILSVRC, the EfficientNet B0. Besides its good performance, it is a small network; it has 5.3 million parameter versus the 66 million parameters from the EfficientNet B7, the winner of the 2019 ILSVRC. It was very important to choose a small network for training agility, which provided the possibility of creating multiple experiments in a short period of time.

The EfficientNet family is an innovative way of thinking about CNNs building. The ResNet networks (HE et al., 2016) can be scaled down (ResNet-18) or up (ResNet-200) by adjusting network depth, and the Inception networks (SZEGEDY et al., 2015; SZEGEDY et al., 2016) can increase parallel convolution modules and change their width. Instead the EfficientNet family has a scaling method that uniformly scales all dimensions of depth, width and resolution (Figure 3) by using a compound coefficient (TAN; LE, 2019). The compound coefficient ϕ captures the following parameters, where $\alpha \geq 1, \beta \geq 1, \gamma \geq 1$ and $\alpha\beta\gamma \approx 2$:

$$\begin{aligned} \text{depth} &: d = \alpha\phi, \\ \text{width} &: w = \beta\phi, \\ \text{resolution} &: r = \gamma\phi. \end{aligned} \tag{2.1}$$

Tan *et al.* carried out a Neural Architecture Search (NAS) to create the EfficientNet B0, so as to maximize accuracy and to minimize computational cost (TAN et al., 2019). All EfficientNet models were then scaled from the baseline EfficientNet B0 using a different compound coefficient ϕ (Equation (2.1)).

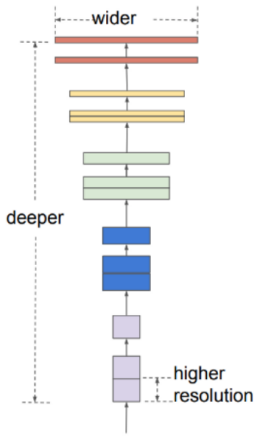


Figure 3 – Representation of compound scaling (TAN; LE, 2019).

Table 1 – Summary of the composition of the EfficientNet B0.

Stages	Operators	Resolution	# of channels	# of layers
1	Conv3x3	224x224	32	1
2	MBConv1, k3x3	112x112	16	1
3	MBConv6, k3x3	112x112	24	2
4	MBConv6, k5x5	56x56	40	2
5	MBConv6, k3x3	28x28	80	3
6	MBConv6, k5x5	14x14	112	3
7	MBConv6, k5x5	14x14	192	4
8	MBConv6, k3x3	7x7	320	1
9	Conv1x1 & Pooling & FC	7x7	1280	1

Table 1 summarizes the stages in the EfficientNet B0 architecture. Stages 2-8 are built with blocks of MBConv (SANDLER et al., 2018), which are combined with a Squeeze-and-Excitation optimization (HU; SHEN; SUN, 2018). A MBConv denotes mobile inverted bottleneck convolution and is a combination of operations as follows. All blocks have a combination of Expansion, a Deepwise Convolution and a Squeeze-and-Excitation phases. Figure 4 was designed based on the architecture described in Tan & Le (2019) and shows one of the most common combinations of these operations.

2.3 Image Oversampling

Image Oversampling is a common strategy for dealing with inbalanced datasets. In order to compare the results obtained in this research with Image Augmentation, the classifier is also trained with Weighted Class Oversampling. The Weighted Class Oversampling is oversampling technique to enlarge a dataset at the probability of occurrence of each class.

Many others techniques of oversampling rely on statistical data from each class to sample the images that best represent each class. However, due to its complexity, it is

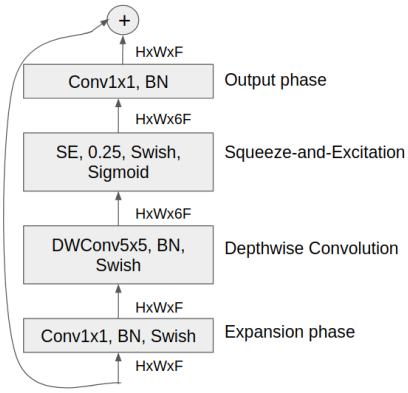


Figure 4 – MBConv6, k5x5. BN denotes Batch Normalization. Swish and Sigmoid are activation functions. 0.25 is the Squeeze-and-Excitation ratio. Source: Prepared by the author (2021).

another challenging task engineering features for painting images.

2.4 Image augmentation

Image augmentation encompasses a suite of techniques that enhance the size and quality of training datasets so that better deep learning models can be built using them (SHORTEN; KHOSHGOFTAAR, 2019). These techniques help deep learning models to better generalize not only in cases of overfitting but also in case of class imbalance.

It is also important to highlight that the use of image augmentation requires a careful analysis of the dataset domain. Some changes can result in non-label-preserving transformations. A classic example is the MNIST handwritten number dataset and the possibility that the image of the number 6 ends up being transformed into a 9 with the vertical flip, a very common geometric transformation.

In this section, we discuss the most common image augmentation techniques (geometric transformations, color space transformations) as well as other classical transformations and image augmentation based on deep learning.

2.4.1 Geometric transformations

A geometric transformation is any transformation that changes size, position and/or orientation of the image. Some examples of this type of transformation:

- **Resize:** is the rescaling of an image.
- **Flipping:** is the horizontal or vertical rotation of an image.
- **Cropping:** is a cutout of a sector of the original image.

- **Rotation:** is the rotation around the axis perpendicular to the image.
- **Translation:** is the displacement of an image horizontally and/or vertically.

Usually, these methods are used in combination: once an image was cropped, the resizing technique is used in order to correct the input image size. All these geometric transformations are expected not only to increase the diversity of the images, but also to improve the generalization power of the model against changes in position, orientation and size of the target to be detected. These transformations are quite valuable in the art domain, as it is possible to maintain the essence of art.

2.4.2 Color space transformations

This type of transformation involves any change to the color of the image, which makes the color space transformation a very broad category. The most common examples of color space transformations are cited below:

- **Color jittering:** it randomly changes the contrast, brightness, and saturation of an image.
- **Histogram matching:** it manipulates the pixels of an input image so that its histogram matches the histogram of the reference image.
- **Histogram equalization:** it adjusts the contrast of an image by modifying the intensity distribution of the histogram. When the image has a narrow range of intensities values, this technique has a large impact on contrast.

Color transformations may discard important color information and thus are not always label-preserving. For example, when decreasing the pixel values of an image to simulate a darker environment, it may become impossible to see the objects in the image (SHORTEN; KHOSHGOFTAAR, 2019). In the context of art, colors are meaningful and altering them without care can result in a non-label-preserving transformation. The importance of colors is perceived in the painters' own speech; Picasso paints sadness a grayish blue and Edouard Manet believes the true color of the atmosphere is violet (KASTAN; FARTHING, 2018). And Van Gogh was adamant in saying "There is no blue without yellow and without orange".[2]

2.4.3 Other classical transformations

Here are other techniques that are very important in image augmentation:

[2] ⟨http://www.webexhibits.org/vangogh/letter/18/B06.htm⟩

- **Noise injection:** is the injection a matrix of random values. A very common type of noise derives from Gaussian distributions.
- **Kernel filters:** is a filter that may increase or decrease the sharpness of the image. A Gaussian filter will result in a more blurred image and a high contrast vertical or horizontal edge filter will result in a sharper image along the edges .
- **Random erasing:** is the deletion of some random values of one or more channels of an image. It was inspired by the mechanisms of dropout regularization (once the network cannot see the whole image, it is harder for it to overfit).
- **Mixing images:** is the patching of two images in a blended new image.

Kernel filters, in the context of CNNs, are not very insightful, as the filters have a very similar interaction with the images as to the internal mechanisms of CNNs. Also, the technique of mixing images has obtained good results but it is not clear why. One possible explanation could be the increase in examples of low-level features, such as lines and borders.

2.4.4 Image augmentation based on deep learning

Here are some image augmentation techniques based on deep neural networks:

- **Feature space augmentation:** is a type of data augmentation done in the feature space of an image. This technique introduced by DeVries & Taylor (2017) implies that an autoincoder structure is able to generate a feature space that generalizes the class well and than the feature space is modified to generate artificial images. Once the feature space is changed, another decoder structure is used to generate the generated images.
- **Adversarial training:** is a framework for using two or more networks in which their loss functions encode disputes between them. It is commonly used for creating images to be misclassified: it finds the minimum possible noise injection needed to cause a misclassification with high confidence. In the context of data augmentation, this technique helps to exploit the weaknesses of the classification model by acting as a search engine.
- **GAN-based Data Augmentation:** is the use of generative networks for up-sampling a training dataset. This technique has been widely used in the medical field, due to the limited access to data or simply because it is a condition that is rare to occur. Just to name a few examples, it has already been used for simulating lung nodules (HAN et al., 2019), ECG (YILDIRIM et al., 2018), liver lesions (FRIDADAR et al., 2018), chromosomes (WU et al., 2018b), skins lesions (QIN et al., 2020)

and Covid-19 results (ELDEEN; KHALIFA, 2020; WAHEED et al., 2020). In Suh et al. (2021), a pipeline is presented as the classification enhancement generative adversarial network (CEGAN). It is composed of three independent networks - a discriminator, a generator and a classifier - and the classifier loss is included in the GAN training process to reduce ambiguity between classes.

The GAN-based Data Augmentation was the chosen strategy for this research because of its ability to handle and create complex images as opposed to the other aforementioned techniques. The next section explains in detail all the relevant theory.

2.5 Generative Adversarial Networks

Generative Adversarial Networks offer a deep learning architecture that generates synthetic images (GOODFELLOW; BENGIO; COURVILLE, 2016). A GAN involves two models:

- Generator model: is a function G that takes a fixed-length random vector as input z, uses $\theta^{(G)}$ parameters and has an output that is a generated image in the trained domain. This vector z is also called a latent space or a vector space comprised of latent variables.
- Discriminator: is a function D whose input x is an example of image of the domain (that could be real sample or a generated sample) and uses $\theta^{(D)}$ parameters; its output is the binary class label of real or fake.

Figure 5 shows the training dynamic of a GAN. The two models interact with each other during training following a game theoretic scenario in which the generator network must compete against an adversary. The generator network produces initially random samples. Its adversary, the discriminating network, tries to distinguish between samples taken from training data and samples taken from the generator. From the feedback of the discriminating network, the generating network learns to draw images that may confuse the discriminating network. This competition is translated into the loss functions J of the training of these networks.

$$J^{(D)}(\theta^{(D)}, \theta^{(G)}) = -\frac{1}{2}\mathbb{E}_{x\sim p_{data}} logD(x) - \frac{1}{2}\mathbb{E}_z logs(1 - D(G(z))), \tag{2.2}$$

$$J^{(G)}(\theta^{(D)}, \theta^{(G)}) = -J^{(D)}(\theta^{(D)}, \theta^{(G)}). \tag{2.3}$$

The discriminator wishes to minimize Equation (2.2) and it can only control $\theta^{(D)}$. The generator wishes to minimize Equation (2.3) and it can only control $\theta^{(G)}$. Because

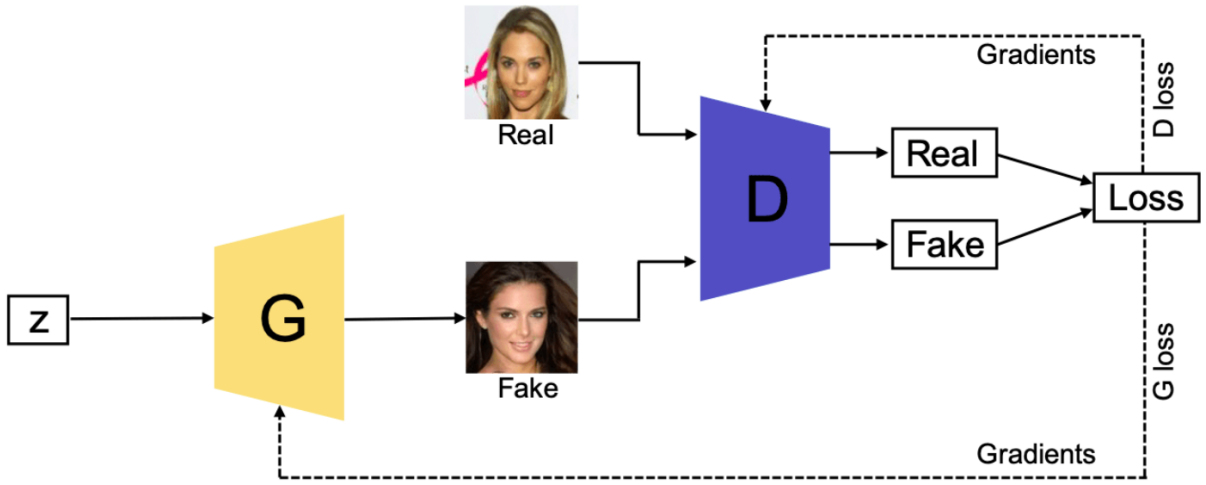


Figure 5 – GANs' architecture (WANG; SHE; WARD, 2021).

each player's cost depends on the other player's parameters, but each player cannot control the other player's parameters, this scenario should not be confused with an optimization, but should be interpreted as a game as mentioned above. The solution to this game is a Nash equilibrium. In this context, a Nash equilibrium is a tuple $(\theta^{(D)}, \theta^{(G)})$ that is a local minimum of $J^{(D)}$ with respect to $\theta^{(D)}$ and a local minimum of $J^{(G)}$ with respect to $\theta^{(G)}$.

2.6 GAN challenges

According to Wang, She & Ward (2021), the generation of synthetic images with GANs faces three major challenges:

- **image quality**: it refers to generated images that can actually confuse the observer to whether the image is real or fake;
- **image diversity**: this captures whether the generator can create a variety of images of the real images domain, avoiding the mode collapse;
- **training stability**: this is linked to whether the architecture and the training setup is able to avoid that the model training does not loose control, a situation known as vanishing gradient. The vanishing gradient problem can also be defined as when the error signals flowing backwards are subsequently smaller until it vanishes and model stops learning as it is unable to change its weights.

The evolution of the architecture of GANs is a sprawling topic, specially because of their variety of purposes that goes beyond creating synthetic images but also image to image transfer, image super resolution, image completion and text-to-image generation. Since the purpose of the GANs in this work is to generate images from a random distribution, we will focus on references that address this challenge.

In the work of Mirza & Osindero (2014), an extra parameter was introduced in the GAN's architecture, making its generator model capable of creating images according to its class label. Chen et al. (2016) developed a GAN architecture able to learn disentangled representations in an unsupervised manner. They introduced a representation learning algorithm called Information Maximizing Generative Adversarial Networks (InfoGAN), in which an information-regularized minimax game is used in order to train a multi class generative model without the label information.

The work of Che et al. (2017) focused on solving one of the GAN deficiencies by reason of highly unstable training and sensitivity to hyper-parameters, which makes it prone to miss modes. The idea is to use an encoded version of the real images to produce the latent variable z, different from earlier works, in which a random distribution was used as input to the generative model.

Zhang et al. (2019) work was the first to introduce the concept of self-attention in GANs (SAGAN). The convolutional architecture processes information better in local neighborhoods and it has no mechanism to deal with long distance dependencies. The Self-Attention mechanism has the effect of enabling both the generator and the discriminator to deal with widely separated spatial regions. Besides the Self-Attention module, this work uses the spectral normalization technique used in the cGAN (MIYATO et al., 2018) in both generator and discriminator models and the two-timescale update rule (TTUR) (HEUSEL et al., 2017) for training stability.

Brock, Donahue & Simonyan (2019) designed a new GAN architecture based on SAGAN with the aim of creating bigger images. This work explores different latent variables z and presents the truncation trick, which involves re-sampling the latent variable z with values which have a magnitude above a chosen threshold. It leads to improvement in individual sample quality at the cost of reduction in overall sample variety. This truncation trick provides a trade-off between image quality or fidelity and image variety. The main features of the BigGAN are: the Self-attention module, class information via class-conditional batch normalization and updating the discriminator twice as much as the generator.

The work of Daras et al. (2020) also is based on the SAGAN construction. The authors replaced the dense attention layer for a novel module of sparse attention patterns for two-dimensional grid. They based the latter in the information theoretic framework of Information Flow Graphs. This assesses how information can be transmitted over multiple steps and still preserve two-dimensional locality. This architecture is considered the current state-of-the-art.

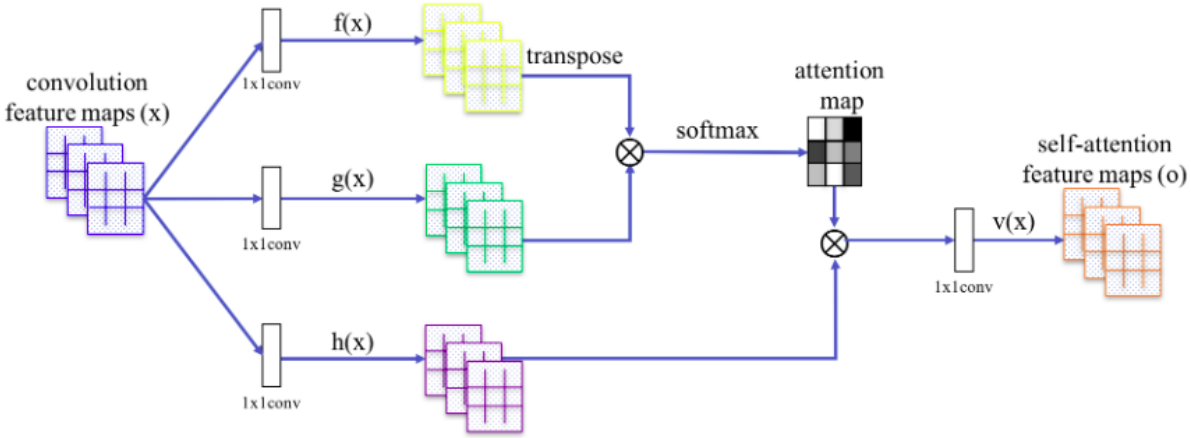


Figure 6 – The self-attention mechanism (ZHANG et al., 2019).

2.7 Self-attention mechanism

The attention mechanism was developed to solve the problem of forgetfulness in Sequence to Sequence (seq2seq) models dealing with the language challenge of machine translation (BAHDANAU; CHO; BENGIO, 2015). In the context of text, each word embedding receives an attention weight that depends on the position of the next word in the translated sentence. Consequently, the model has more information about the importance of each word in the sentence. This architecture was originally composed of recurrent or convolutional neural networks working as encoder and as decoder with the attention mechanism. Vaswani et al. (2017) proposed Transformers, an architecture that dispenses with the use of recurrences and convolutions entirely and is based only on an attention mechanism known as self-attention. The concept of self-attention derived from the intra-attention of Cheng, Dong & Lapata (2016), in which it calculates the answer at a position in a sequence once all positions with the same sequence have been checked.

The self-attention mechanism lets inputs to interact with each other and to find out to which part of the input they should pay more attention. This mechanism was also translated to the computer vision task, specially in the context of GANs. Zhang et al. (2019) presented the Self-Attention Generative Adversarial Network (SAGAN), a generative adversarial network architecture with an attention module. SAGAN has the attention module on both the generator and the discriminator, which are trained in an alternating fashion by minimizing the hinge version of the adversarial loss.

Figure 6 shows in detail the mechanism behind the self-attention mechanism. Transformers are used to create the key, query and value:

$$f(x) = W_f x, \qquad g(x) = W_g x, \qquad h(x) = W_h x.$$

The attention map is created after applying a softmax to the dot product of the key and the query (Equation (2.4)). Another dot product is taken between the attention map and

the value; the attention map is applied to the value in order to create a self-attention map o_j (Equation (2.5)):

$$\alpha_{i,j} = \text{softmax}(f(x_i)^\top g(x_j)), \tag{2.4}$$

$$o_j = W_v \left(\sum_{i=1}^{N} \alpha_{i,j} h(x_i) \right). \tag{2.5}$$

The self-attention map ensures that the model learns long range dependencies within an image, which is a deficiency in convolutional neural networks.

2.8 Loss functions for GANs

A loss function directs a deep learning model during training on how to find its best parameter values. It is not surprising that in the case of GANs the loss function is very important, as it is responsible for training two models simultaneously.

The original loss function of Goodfellow (2016) incurs in both mode collapse and vanishing gradient (WANG; SHE; WARD, 2021). Arjovsky, Chintala & Bottou (2017) studied these problems and adapted the Earth-Mover (EM) distance to work as a loss function. The Earth-Mover (EM) distance or Wasserstein-1 is defined as:

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \prod(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\gamma}[||x - y||], \tag{2.6}$$

where $\prod(\mathbb{P}_r, \mathbb{P}_g)$ denotes the set of all joint distributions γ whose marginals are respectively $\mathbb{P}_r$ and $\mathbb{P}_g$. Intuitively, $\gamma(x, y)$ indicates how much "mass" needs to be transported from x to y in order to transform the distribution $\mathbb{P}_r$ into the distribution $\mathbb{P}_g$. The EM distance then is the "cost" of the optimal transport plan. Translating this into the image context, P_r is the probability distribution of a set of real images and P_g is the probability distribution of generated images. Equation (2.7) presents the loss function introduced by Arjovsky, Chintala & Bottou (2017):

$$L = \sup_{||f||_L \leq 1} \mathop{\mathbb{E}}_{\tilde{x}\sim\mathbb{P}_g} [D(\tilde{x})] - \mathop{\mathbb{E}}_{x\sim\mathbb{P}_r} [D(x)], \tag{2.7}$$

where the supremun is over all the 1-Lipschitz functins. By implementing the Wasserstein loss function, it is necessary to ensure that the discriminator is in the space of the 1-Lipschitz functions. This means the norm of the gradients should be as most 1. In order to guarantee this constraint, one can use the weight clipping technique, which meant limiting the weights to an interval after each gradient update. The authors highlighted that enforcing the 1-Lipschitz constraint with the weight clipping was a problematic way of doing it as it limits the learning ability: if the clipping parameter is small, it may lead to vanishing gradients; on the other hand, if the clipping parameter is large, it can

take longer for the weights to reach the limit which means that it is harder to train the discriminator (ARJOVSKY; CHINTALA; BOTTOU, 2017). The solution came with the work of Gulrajani et al. (2017) with the addition of a Gradient Penalty:

$$L = \mathop{\mathbb{E}}_{\tilde{x}\sim\mathbb{P}_g}[D(\tilde{x})] - \mathop{\mathbb{E}}_{x\sim\mathbb{P}_r}[D(x)] + \lambda \mathop{\mathbb{E}}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}\left[(||\nabla_{\hat{x}} D(\hat{x})||_2 - 1)^2\right]. \tag{2.8}$$

The Gradient Penalty is a regularization term. The $\hat{x}$ is the interpolated resulting image distribution between the real image distribution and generated image distribution. The interpolated image's gradients are then regularized. This loss function is maximized by the discriminator and minimized by the generator. The λ is the gradient penalty coefficient and has its default value of 10 from the original work.

Besides that, the work of Wu et al. (2018a) proposed a loss function called Wasserstein divergence, which is a relaxed version of the original Wasserstein loss function and it does not require the k-Lipschitz constraint; and Adler & Lunz (2018) generalized the theory of WGAN with gradient penalty to Banach spaces.

3 RELATED WORK

In this chapter, we explore relevant related work in artwork classification. We also present the Generative Adversarial Networks (GANs) and their use in the art domain.

3.1 Artwork classification

The task of automatically classifying artwork has been extensively studied since 2010. Following Image Processing classic techniques, initially the research on artwork classification was focused on feature engineering for machine learning algorithms (SHAMIR et al., 2010; ARORA; ELGAMMAL, 2012).

The success of Krizhevsky, Sutskever & Hinton (2012) in general image classification strongly affected strategies in artwork classification. They used the images of the ImageNet Challenge to train a deep convolutional neural network for object classification. With a top 5 error of 15.3% in the ImageNet Challenge 2012, CNN scored an error that was more than 10.8 percentage points below the second position. This work not only proved the superior performance of CNNs, but also their trained weights with the ImageNet dataset became an important resource for fine-tuning models for other purposes.

In subsequent works in the field of artwork classification, the studies focused on testing the superior performance of CNNs compared to resource extraction techniques and verifying the difference in performance of model training from scratch and with the model weights trained with the ImageNet dataset as initial weights (KARAYEV et al., 2014; BAR; LEVY; WOLF, 2015; CONDOROVICI; FLOREA; VERTAN, 2015; TAN et al., 2016; FLOREA; TOCA; GIESEKE, 2017). It was shown that mid-level features derived from the ImageNet object datasets are generic for art recognition and they were superior to any kind of attempt to generate hand-tuned features.

As architectures other than the AlexNet from Krizhevsky, Sutskever & Hinton (2012) emerged, research in the context of artwork classification was directly benefited. In the ILSVRC 2015, the Residual Neural Network, a new family of deeper convolutional networks, won the challenge (HE et al., 2016). In Lecoutre, Negrevergne & Yger (2017), both the AlexNet and the ResNet50 architectures were used for artistic movement classification. They also used ImageNet pretrained models as initial weights and started training only the last layer until they retrained the whole network. They found that the best result was reached when approximately 20% of the layers were retrained.

Cetinic, Lipic & Grgic (2018a) expanded the tasks involving artwork classification beyond style, genre, artist and time period classification. They approached the classification

of the nationality of the painting. Their results showed that scene recognition and sentiment classification yields better results than fine-tuning networks pre-trained only for object recognition. They conjectured that the semantic correlation between the former domains could be inherent in the CNN weights. In later work, they explored the memorability of artwork within the deep learning architecture. They concluded that abstract styles tend to be more memorable than figurative and symmetry does not correlate with memorability (CETINIC; LIPIC; GRGIC, 2018b).

Chu & Wu (2018) study the style classification task focusing on describing image texture with deep learning using the VGG-19, the 2014 ILSVRC winner. They investigated the intra-layer and inter-layer correlations in order to create deep features for style classification. Chen & Yang (2019) presented an adaptive cross-layer correlation for artwork classification, in which it adaptively weights features in different spatial locations based on similarity in a VGG-16, a smaller version of the previous architecture. On the other side, the work of Elgammal et al. (2018) analyzed the learned representations of a fine-tuned ResNet-152, the biggest ResNet in that moment. It was shown that some of the style patterns designed by art critic Heinrich Wölfflin (1864-1945) correlates with the PCA decomposition of these learned representations.

In Rodriguez, Lech & Pirogova (2019), five image patches of painting were used for training and weights for each patch were optimized in order to improve accuracy of the final model. Sandoval, Pirogova & Lech (2019) also worked with image patches, but in a two-stage deep learning approach, in which these five patches are trained independently at a first step. At the second stage, the outcome of these patches are fused to a second shallow neural network for the final decision. And Bianco et al. (2019) develop a multibranch approach for exploiting the painting and the crops at different resolutions. These crops are extracted with a Spatial Transformer Network trained to identify the most discriminative subregions of paintings.

The Inception-V3 network was the first runner up in ILSVRC 2015. The work of Zhu et al. (2019) not only trained a Inception V3 for classifying nine artistic movements, but also used Grad-CAM heat map for visualizing the areas of the images the model was focusing for class prediction.

More recently, Zhong, Huang & Xiao (2020) presented a two-channel dual path network and two inputs are used: the RGB image and four-directional gray-level co-occurrence matrix for detecting the brush stroke information. Instead of comparing performance between manually made features and features originating from CNNs, they developed a Dual Path Network to combine the outputs. In order to obtain good results, the authors needed to train this architecture from scratch with the ImageNet dataset, and only then perform the fine-tuning for the art style classification domain, in accordance with the other works mentioned here.

It is clear that a great deal of effort has been made to improve art style classification. However, this is a challenge that many have approached differently: different datasets, different classes and addressing a different number of classes. Thus, even though some works consider themselves state of the art, we believe that a comparison between them is not fair.

3.2 Generative Adversarial Networks and Art

In the art domain, generative adversarial networks have been developed aiming not only to create art that would convince an audience to it truthfulness but also exploring their possible creative power. The variety of GANs range from a generator for image style transfer – also known as the CycleGAN (ZHU et al., 2017) –, an Image-to-Image translator from art to real images (TOMEI et al., 2019; GAO; TIAN; QI, 2020) to a model specialized in creating Chinese landscape (XUE, 2021).

Some works aimed to create artwork using the WikiArt dataset. Tan et al. (2017) created the ArtGAN, a conditional GAN that allows the backpropagation of the label information of the genre that the generated artwork should belong. Elgammal et al. (2017) developed the Creative Adversarial Network aiming to creatively generate artwork by maximizing deviation from established styles and minimizing deviation from art distribution. Due to its goal of creating original art, these authors found a way to encourage the generator "to be creative", which was to penalize it any time that it was too easy for the Discriminator to identify the synthetic image as being art from a certain style.

4 IMPROVING CLASSIFICATION RESULTS WITH GANS

In this chapter we present the main contribution of our work. These are, namely, the adaptation of the SAGAN training for better learning to capture art images of the WikiArt dataset with the Wasserstein loss function and gradient penalty; and the creation of a strategy that we here call Class-by-Class Performance Analysis to improve the performance of classification models.

4.1 Training GANs Architecture

The original article of the SAGAN architecture employed hinge loss function. For our dataset, we experienced frequent mode collapse and vanishing gradient with this loss function. We suspect that this is due the diversity of each of the classes of our dataset. For this reason, we used the Wasserstein with Gradient Penalty (Wasserstein-GP) loss function so as to have better control of the values for feedback. In our experiments, the interpolation between a batch of real images and fake images was enforced and the gradient norm of its output was limited at 1. The value of the penalty coefficient (λ) was 10, following the original paper (GULRAJANI et al., 2017).

For this research, we had, as a limiting element, the hardware of two GeForce GTX 1080 Ti (12 GB) GPUs. Therefore, we trained with a batch size of 32 images, unlike SAGAN's original training batch size of 256 images. The process of training a generator of images of size 128x128 ran for for 200.000 epochs (approximately 34 hours). Each class was trained independently, using the equivalent training dataset of the classifier, in order to assure that any data leakage could not occur between the generated images and the test dataset. The optimizer setup follows the original SAGAN article: Adam optimizer with $\beta_1 = 0$ and $\beta_2 = 0.9$. The learning rate is constant but specific for each model: for the discriminator is 0.0004 and for the generator is 0.0001 (ZHANG et al., 2019).

4.2 Training EfficientNet B0 Architecture

In order to train the EfficientNet B0 for our purpose, we analyzed in the literature the best way of training with a relatively small and imbalanced dataset. Several previous proposals on art style classification achieved best results with ImageNet pretrained weights as initial weights (KARAYEV et al., 2014; BAR; LEVY; WOLF, 2015; CONDOROVICI; FLOREA; VERTAN, 2015; TAN et al., 2016; FLOREA; TOCA; GIESEKE, 2017) and

most of this same literature shows that better results where obtained when only unfreezing some layers. We could not find a detailed description of training for EfficientNet; we trained the EfficientNet architecture in four steps:

1. initiate training with only the last block unfrozen (that is, only parameters of this block are set to be trained);
2. once training is done, the next block is unfrozen and training starts again;
3. step 2 is repeated until the training does not present any improvement in the validation dataset;
4. once we learn up to which block to unfreeze to have the best model training, this block becomes our reference, which we call block N. Since training up to block N brought the best result, apply the following strategy to all experiments: we train with all blocks thawed up to block N-1, block N and block N+1.

This last step was adopted to reduce training time. For all training, the Stochastic Gradient Descent optimizer was used with decay 0.9 and momentum 0.9; initial learning rate of 0.01 with decay after the fifth epoch ($lr = lr * e - 0.1$). Images were resized to 224x244 and batch size was 32 images.

4.3 Class-by-Class Performance Analysis

We introduce a strategy that we refer to as Class-by-Class Performance Analysis. This strategy was developed empirically, guided by trial and error experiments. First, to start up learning, a baseline model is trained without image augmentation techniques so as to have a reference of our improvements. Then, the steps required to implement CCPA are:

1. **Train a model exploring the benefits of geometric transformations:** with due care, geometric transformations are welcome in the context of art paintings. A very common type of geometric transformation would be the vertical flip. For abstract images, this would not be much of a problem, but there are many landscape and people paintings and the vertical flip in these contexts would not make sense. That is why we trained a model only using horizontal flips with 50% occurrence probability, besides some random rotations with not very prominent angulation – around $[-10, +10]$ degrees – and central and random crops.
2. **Analyze the trained model with geometric transformations:** once we find the best model under this training condition, the class-by-class should be performed in order to inspect the size and the performance of each class.

3. **Train GAN models for each class belonging to the group of the lowest performing classes:** in order to focus our efforts in the classes in which the dataset provides the lowest amount of information, we only trained GAN models for creating synthetic images of those low performing classes.

4. **Train EffB0 models with generated images:** in this step, each EffB0 model is trained with the generated images from one of the low performing classes considered in steps 2 and 3.

To successfully complete step 4, the initial question is: what would be the adequate quantity of artificial images to maximize the information obtained from the artificial dataset? In order to ensure a careful choice of the number of generated images to be added, we implemented two strategies to guide this decision that is directly linked to the number of images belonging to classes with low performance. These classes may or may not contain a small number of images. Hence, we suggest two strategies when we sample either:

- **low quantity classes:** add a multiple of the number of original images;
- **high quantity classes:** add a fraction of the number of original images.

Here the concern was to prevent the up-sampling of classes with large volumes from ending up further unbalancing the distribution of images by class. Hence, we define that a class is considered low quantity when its volume does not exceed 20% of the volume of the largest class. Thus, we ensure that the low quantity class technique is not at risk of adding a disproportionate amount of images. Once this is exceeded, the class is treated as a high quantity class.

The Class-by-Class Performance Analysis is better understood by reading later Section 5.2, where all the step are illustrated with our results.

5 EXPERIMENTS

In this chapter we describe our experiments. Initially, we present the artwork dataset used in all experiments. Afterward, the results related to the classification task are presented following the steps of the Class-by-Class Performance Analysis. It is also presented the results for the classification task with weighted class, a traditional technique for data balancing. Finally, we discuss the results obtained from trained generators with sample of synthetic images for visual inspection.

5.1 The WikiArt Dataset

We used the Wikiart dataset, which is an online encyclopedia of visual arts commonly used for the art style classification task. For our research, we used the version that was discussed by Elgammal et al. (2018), from which we derived the follow combination of art movements that were greatly correlated:

- New Realism and Contemporary Realism were added to Realism;
- Action Painting was added to Abstract Expressionism;
- Synthetic Cubism and Analytical Cubism were added to Cubism.

Figures 1a-1o depict some of the most significant examples of each class. A total of 63,659 images are available there; 10% of them were used for testing and 10% of the remaining dataset was used for validation. The training volumetry of the image distribution is presented in the last column of Table 2. The training dataset is used both in classifier training and in GAN training. This restriction was necessary in order to guarantee that any data leakage could occur between the generated images and the test dataset. All classifiers were trained five times with different stratified samples for training and validation. Ideally, each of the classes would present around 4,250 images, but the class imbalance is apparent with the quantity of images varying between 10,566 and 940 images.

5.2 Classification results

The result of the EfficientNet B0 baseline trained model is shown in Table 10 (first line). In the second line, it is presented the EfficientNet B0 model trained with geometric augmentation. The performance for each class of the latter is shown in Figure 7. From this analysis, we verified a very unusual behavior: the Ukiyo-e movement – the Japanese style –

Table 2 – Dataset used in experiments.

Art movement	Total of images	Training images
Abstract Expressionism	2,783	2,283
Art Nouveau	4,292	3,442
Baroque	4,241	3,448
Color Field Painting	1,615	1,308
Cubism	2,417	1,942
Early Renaissance	1,391	1,134
Expressionism	6,720	5,457
Impressionism	13,060	10,566
Minimalism	1,258	1,009
Naïve Art	2,340	1,917
Northern Renaissance	2,552	2,084
Pop Art	1,460	1,205
Realism	11,400	9,188
Romanticism	6,963	5,640
Ukiyo-e	1,167	940

had the least amount of images and the best f1-score. On the other hand, when verifying the amount of images of the classes that obtained the worst performance, we found the Pop Art class, with the worst performance and few images (1205), but it is soon followed by the Expressionism and Romanticism classes, both with more than 5,000 images in training. With this, we reinforce the idea that we had intuitively at the beginning of the research that this dataset is quite complex to the point where the performance of the classes is not correlated with the amount of training images.

With this first analysis as a guide, we performed the EfficientNet B0 retraining by adding images generated from only one class at a time and for the class with low performance and low number of real images we applied the low quantity classes sampling and low classes performance and high quantity of real images, the high quantity classes sampling. Finally, we compiled the main classification results of this research.

5.2.1 Sampling low quantity classes

The analysis of the trained EffB0 shown in Figure 7 points out the Pop Art class as fourth lowest class in terms of quantity of images, so the low quantity class sampling strategy was applied. The results are shown in Table 3. The classifier that obtained the best accuracy score is the same as that obtained the best f1-score for the Pop Art class.

5.2.2 Sampling high quantity classes

Analyzing the others culprits of deterioration of the classifier's performance, we examined the Expressionism and the Romanticism art movements. Both classes have a similar

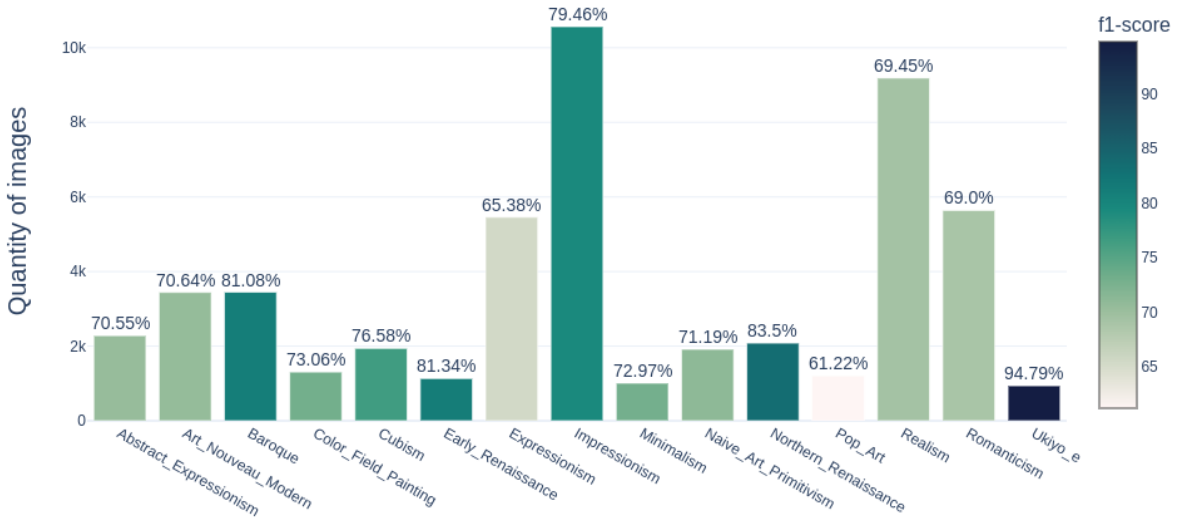


Figure 7 – Analysis of the trained EfficientNet B0 with geometric augmentation. Source: Prepared by the author (2021).

Table 3 – Summary of experiments with added synthetic Pop Art images.

Synthetic images	Class f1-score (%)	Classifier accuracy (%)
1205 (1x)	62.80 ± 2.21	73.98 ± 0.11
2410 (2x)	62.52 ± 1.63	73.60 ± 0.16
3615 (3x)	**63.26 ± 2.56**	**74.05 ± 0.19**
4820 (4x)	62.67 ± 2.09	73.87 ± 0.25

Table 4 – Summary of experiments with added synthetic Expressionist images.

Synthetic images	Class f1-score (%)	Classifier accuracy (%)
682 (1/8)	65.67 ± 1.24	73.74 ± 0.29
1364 (1/4)	**67.02 ± 0.43**	**74.31 ± 0.15**
2728 (1/2)	65.01 ± 0.83	73.83 ± 0.09
4092 (3/4)	65.89 ± 0.10	74.18 ± 0.10

behavior to that of our reference model: they have many more images than the average of 4250 images that would be expected for each classes – more than four times Pop Art's image quantity – but low performance. It is interesting to observe that for both classes, the quantity of images that generated the best results was 1/4 of its original training data quantity. We did not go so far as to point out that this would be a rule, but we believe that there are two possible explanations: we found the limitation of the mode diversity of the trained GAN; or the classes by coincidence were only able to bring this amount of information or at least an approximate amount, since in some cases the measurements are not that different considering the margin of error.

Table 5 – Summary of experiments with added synthetic Romantic images.

Synthetic images	Class f1-score (%)	Classifier accuracy (%)
705 (1/8)	69.60 ± 0.34	73.77 ± 0.34
1410 (1/4)	**70.32 ± 0.82**	**74.12 ± 0.44**
2820 (1/2)	69.43 ± 0.55	73.90 ± 0.27
4230 (3/4)	70.13 ± 0.84	74.04 ± 0.30

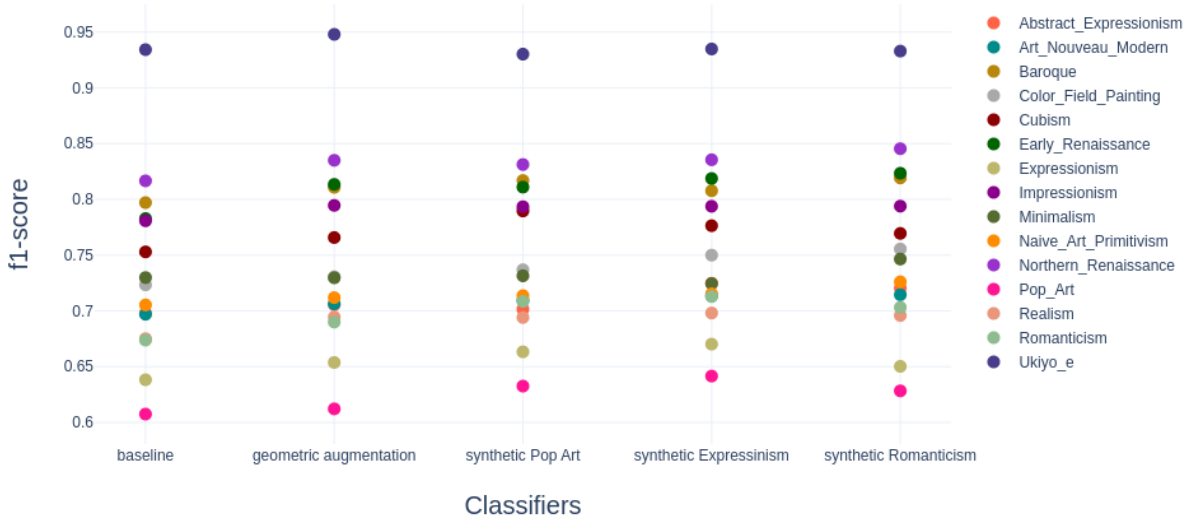


Figure 8 – The evolution of performance for each class during experiments. Source: Prepared by the author (2021).

5.2.3 Impact analysis of the synthetic images

Figure 8 shows how each class performed in the best experiment for each class (Pop Art, Expressionism and Romanticism). Looking at the top, it is clear that the addition of synthetic images caused the Ukiyo-e class to lose performance. At the same time, the performance of the Impressionistic and Realistic class were almost unaltered through all experiments. We believe that this behavior is related to the high amount of images in both classes. None of the experiments caused the classes with generated images to reach a similar amount. However, several performance improvements were observed when analyzing the other classes, as described below. Tables 6, 7, 8 e 9 show for which class the classifier mostly predicted wrongly the bottom-5 performing classes for each model.

Impact of Pop Art synthetic images

Analyzing the graph in Figure 8, it is possible to see that not only the Pop Art class had its performance improved against the model with only geometric transformations, but also Cubism and Color Field Painting. This behavior indicates the indirect benefit of improving

Table 6 – Bottom-5 performing classes for the model with only geometric transformations

Model classification	Model's most mistaken class
Pop Art	Expressionism
Expressionism	Realism
Romanticism	Realism
Realism	Impressionism
Abstract Expressionism	Color Field Painting

Table 7 – Bottom-5 performing classes for the model with Pop Art synthetic images

Model classification	Model's most mistaken class
Pop Art	Abstract Expressionism
Expressionism	Realism
Realism	Impressionism
Abstract Expressionism	Color Field Painting
Romanticism	Realism

the quality of the Pop Art class. Comparing Tables 6 and 7, we notice that the addition of synthetic Pop Art images has worsened the differentiation between Pop Art and Abstract Expressionism.

Impact of Expressionist synthetic images

Figure 8 shows that generated Expressionist images helped even more than generated Pop Art images for improving the Pop Art class performance. The Color Field Painting class was also benefited by these new information. Analyzing Tables 6 and 8, it is notable that adding the generated images were not enough for the classifier to stop confusing Expressionist images as being mostly Realist images. It is speculated that perhaps the generation of realistic images could add value in this context.

Impact of Romantic synthetic images

By including Romantic images, Figure 8 shows that not only the Romantic class had its performance improved against the model with only geometric transformations, but also Northern and Early Renaissance. By visual inspection of Figure 1, these classes are the once most visually close. Tables 6 and 9 show the same behavior of adding synthetic Expressionist images: the confusion with Realist images is still an issue.

5.2.4 Summary of results

Finally, Table 10 shows all the relevant results of each EfficientNet B0 model trained in this research. The strategy of Class-by-Class Performance Analysis allowed us to improve the classifier accuracy by more than 2%. The experiment with the weighted class

Table 8 – Bottom-5 performing classes for the model with Expressionist synthetic images

Model classification	Model's most mistaken class
Pop Art	Art Nouveau
Expressionism	Realism
Realism	Impressionism
Romanticism	Realism
Art Nouveau	Expressionism

Table 9 – Bottom-5 performing classes for the model with Romantic synthetic images

Model classification	Model's most mistaken class
Pop Art	Cubism
Expressionism	Realism
Realism	Impressionism
Romanticism	Realism
Art Nouveau	Expressionism

Table 10 – EfficientNet B0 trained models.

Experiment	Precision (%)	Recall (%)	Accuracy (%)
Baseline	73.78 ± 0.43	73.30 ± 0.38	72.21 ± 0.26
GeoAug	75.24 ± 0.52	74.39 ± 0.47	73.59 ± 0.18
GeoAug + weighted class	74.93 ± 0.40	74.89 ± 0.58	73.23 ± 0.47
GeoAug + synthetic Pop Art	75.32 ± 0.17	75.18 ± 0.47	74.05 ± 0.19
GeoAug + synthetic Expressionism	**75.93 ± 0.34**	**75.33 ± 0.13**	**74.31 ± 0.15**
GeoAug + synthetic Romanticism	75.92 ± 0.66	75.39 ± 0.56	74.12 ± 0.44

approach is also added. It shows that the attempt of balancing equally the classes ended up degenerating the classifier performance.

5.3 Generated image analysis

The generation of images that resemble artwork in this research is restricted to the objective of improving the performance of classifiers. However, it is interesting to observe how close the generated images were to some true examples of the artistic movement. Figures 3, 4 and 5 show a sample of the images generated in our experiments. To help in understanding the quality of these generated images, real images of corresponding artistic styles were added. By visual inspection, it is noticeable that the generated images retain general properties of the styles. We observe that for the three styles presented here, the models are competent in choosing the color palette, even capable of creating grayscale and colored images, as seen in Figures 4 and 5. It is also noticeable that the models do not perform well in defining shapes, however we were pleased as we understood since the beginning of

this research that image augmentation with GANs would help us in the context of color space transformation.

Also, the same GAN training configuration was used to train the other classes of the WikiArt dataset. Extending the training to the other classes was important to better understand the real image quality range of the training setup set for the GAN used in this research.

In addition to the already observed ability to maintain the correct color combination, the GAN models were able to generate almost perfect images of the Minimalist and the Color Field Painting artistic movements (Figures 19 and 15). By visual inspection, it can be seen that these are the styles with less complex shapes. Abstract Expressionist style and Cubist style (Figures 12 and 16) were also well portrayed by the GAN model. They have a more complex pattern of format than the latter, but they are a more repetitive pattern, especially when compared to specific forms of human representation in other art movements.

As a result of the generation of images from other classes with a lot of human representation in their subject, it became quite clear the low performance of the GAN model to generate human figures when compared to landscapes, for example (Figures 14, 17, 21 and 22). We are able to identify the human figures generated by the GAN models only because of the characteristic silhouette and the choice of colors close to skin color.

The GAN model trained to generate Ukiyo-e style images (Figure 23) does not stand out for performance. It is possible to see the distinction of the images – that is, the generated images are clearly inspired by the Japanese style – but the outline of this style is too complex for the model to absorb.

Finally, we noticed a good performance in the images of artistic movements in which black-and-white image style are present. It is interesting to note that this GAN architecture can have great potential for become a generator of sketch images or simply images of a more color restricted domain.

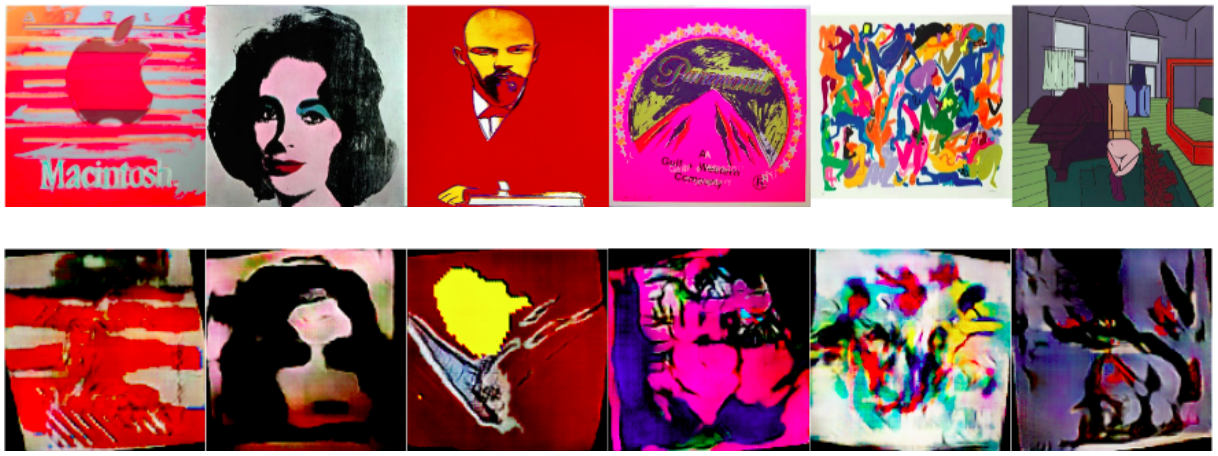


Figure 9 – Pop Art: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

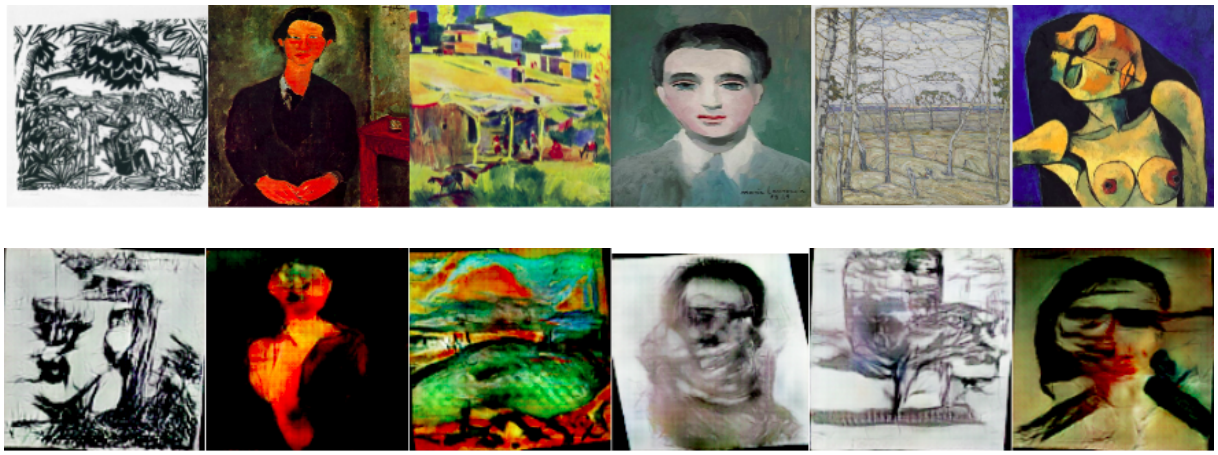


Figure 10 – Expressionism: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

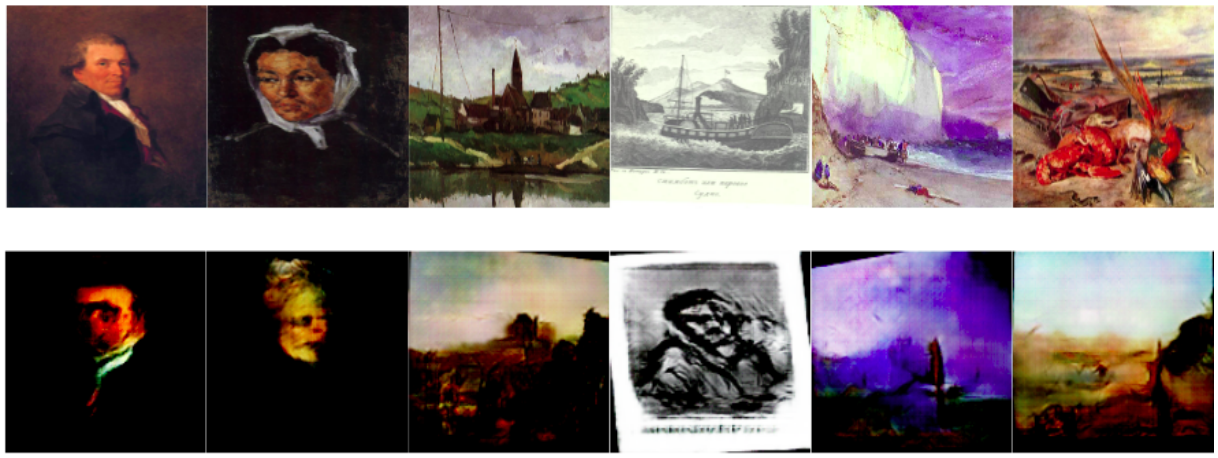


Figure 11 – Romanticism: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

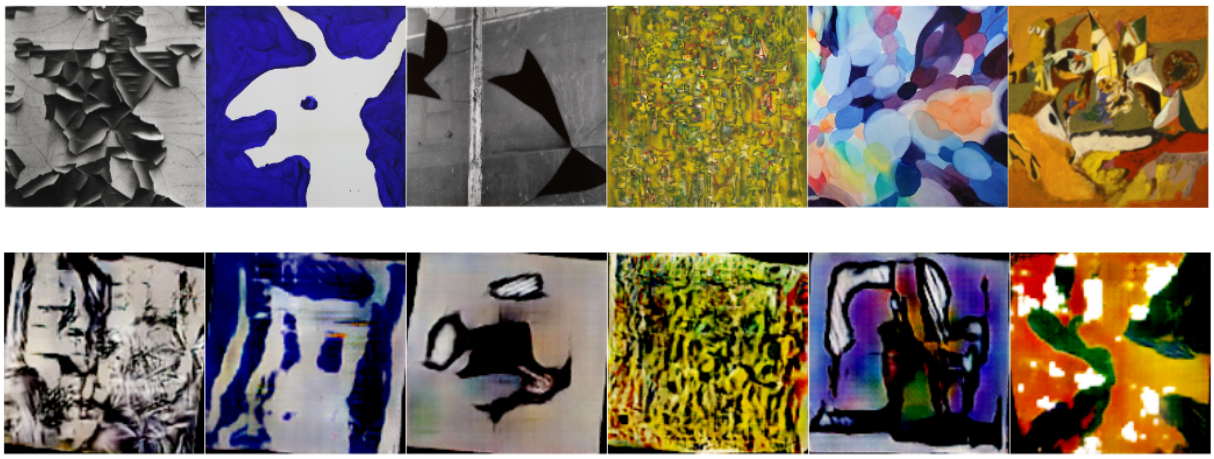


Figure 12 – Abstract Expressionism: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

Figure 13 – Art Nouveau: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

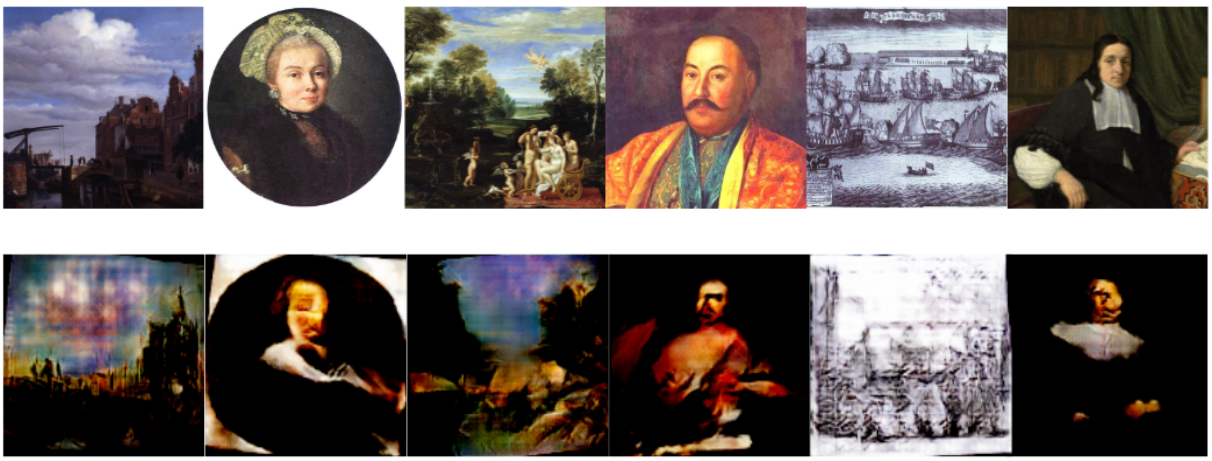


Figure 14 – Baroque: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

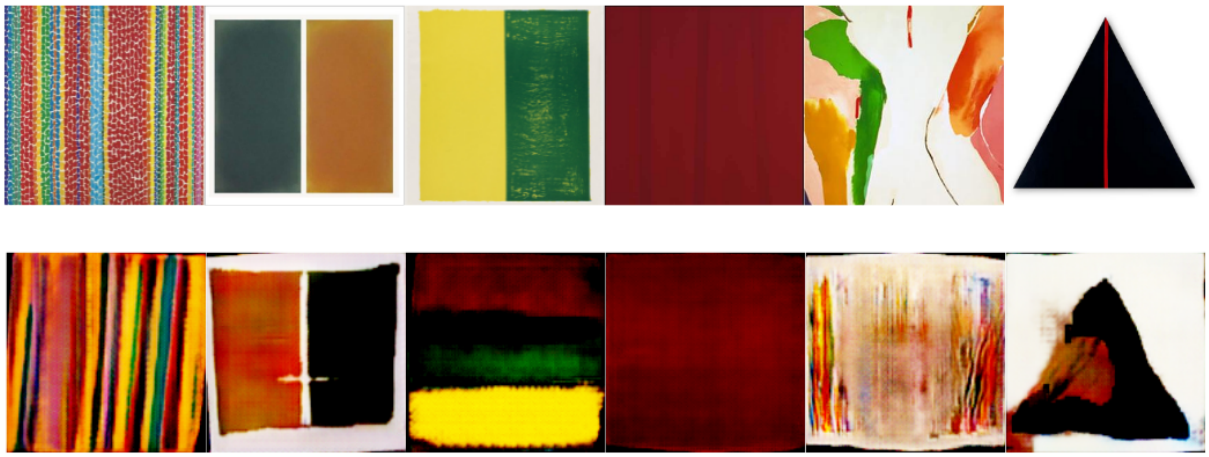


Figure 15 – Color Field Painting: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

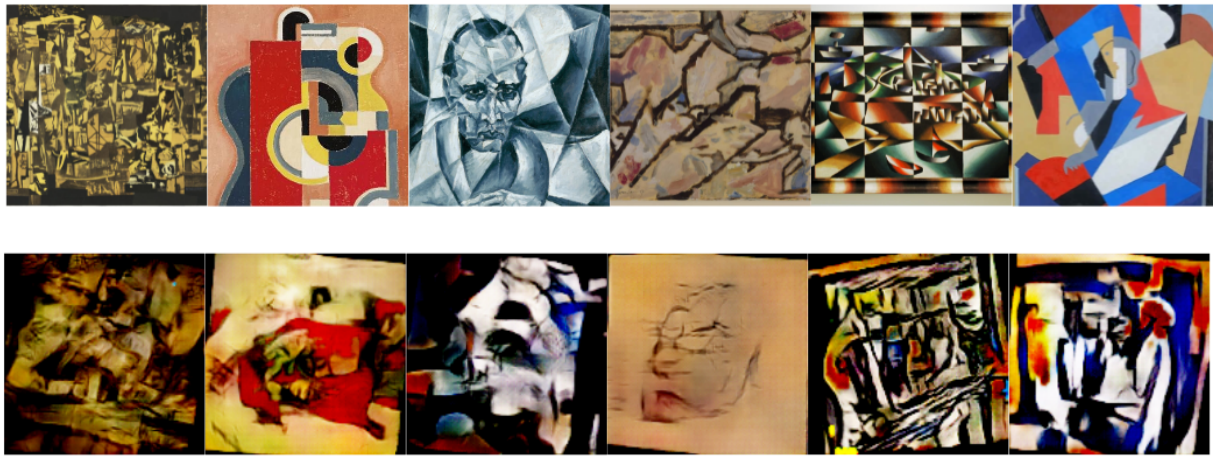


Figure 16 – Cubism: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

Figure 17 – Early Renaissance: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

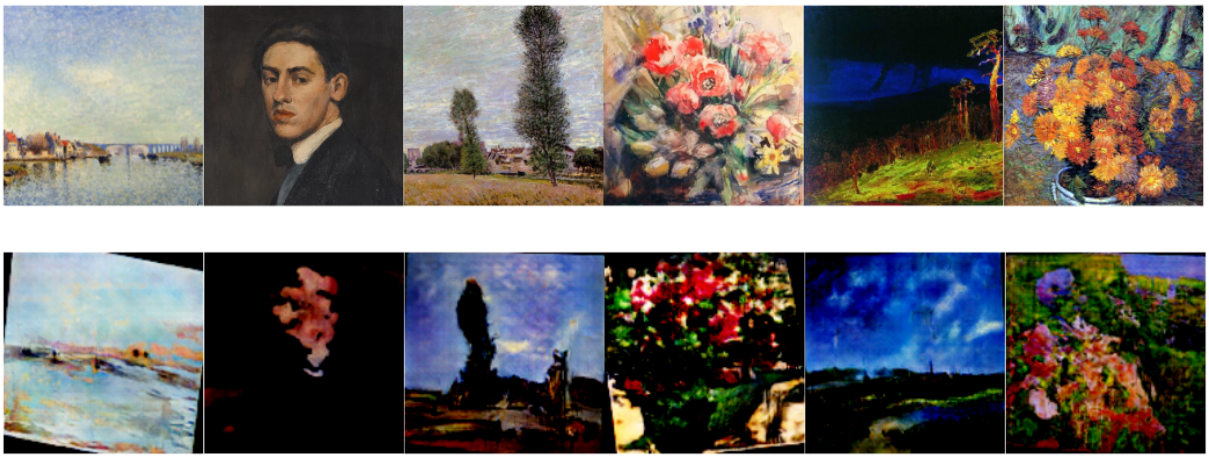


Figure 18 – Impressionism: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

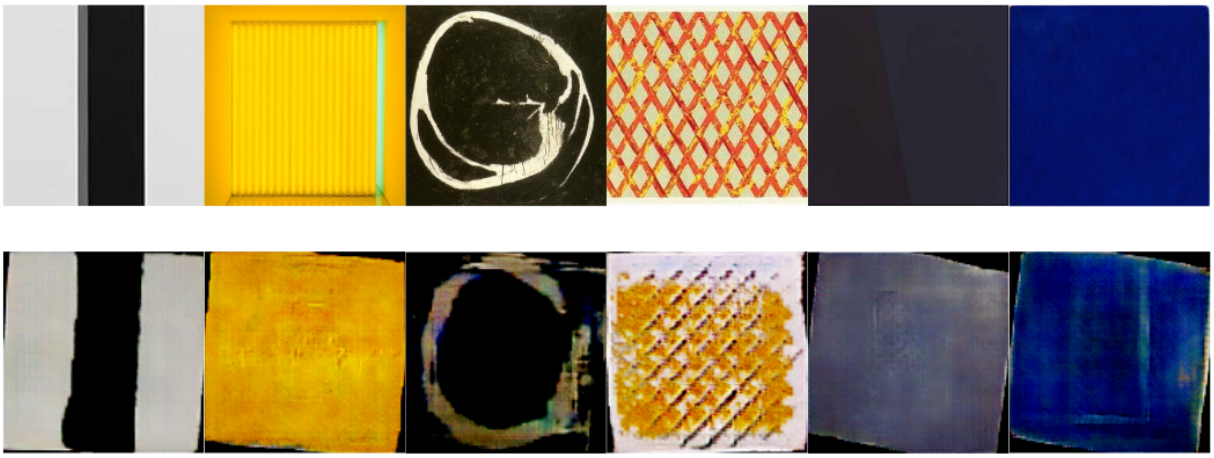


Figure 19 – Minimalism: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

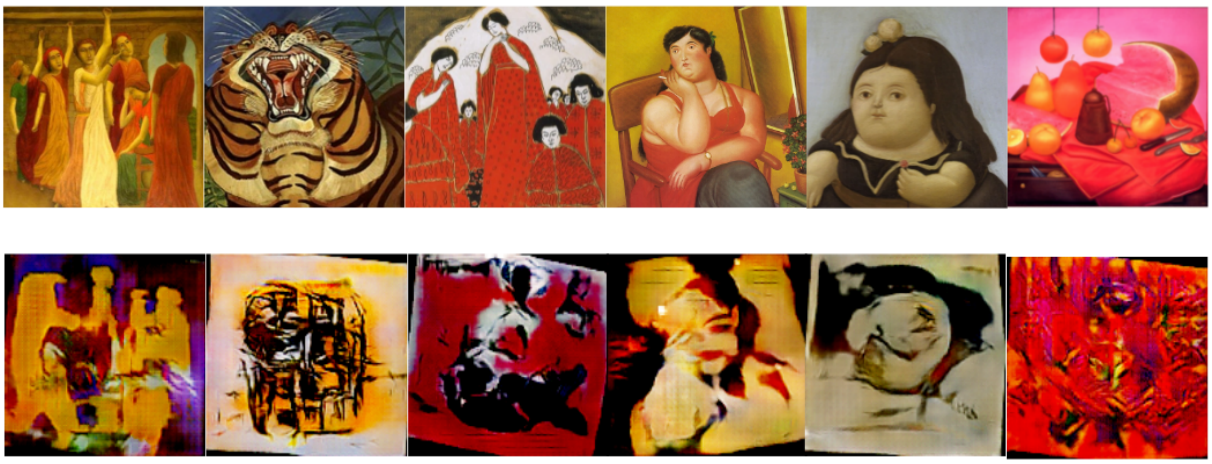


Figure 20 – Naïve Art: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

Figure 21 – Northern Renaissance: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

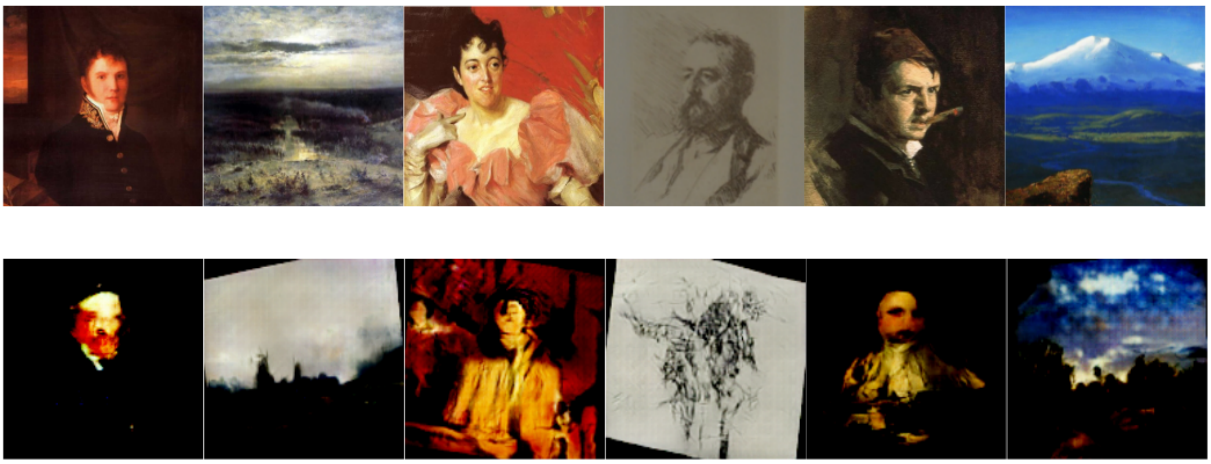


Figure 22 – Realism: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

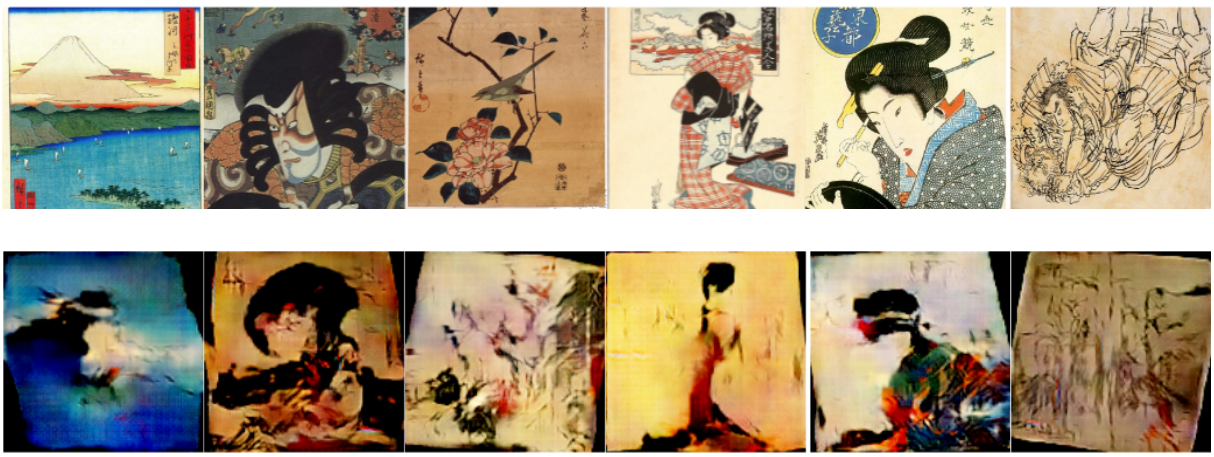


Figure 23 – Ukiyo-e: real images are in the first row and generated images are in the second row. Source: Prepared by the author (2021).

6 CONCLUSION

6.1 Discussion

The difficulty in classifying art styles is directly linked to the characteristics of the artwork image domain: the imbalance of classes and the high diversity within classes and the similarities between artwork that make the boundaries between styles to be rather flexible.

The use of GANs for image augmentation was the main tool we employed to enhance information in this scenario; we had to explore several training architectures and configurations that resulted in the use of a GAN with self-attention layer and the Wasserstein-GP loss function. Also, in order to verify the best sampling strategy, we had to develop a working method called Class-by-Class Performance Analysis. These technical contributions go beyond the context of art style classification; they should be valuable in other contexts where the dataset also has complex classes, especially when augmenting images where color transformation can lead to non-label-preserving transformation.

We verified that our approach and method can actually create better classifiers. In the case of this research, we found that using only geometric transformations, generating images with a GAN with self-attention layer and training it with the Wasserstein-GP loss function allowed us to avoid collapse mode and the vanishing gradient and, consequently, to increase model accuracy by more than 2%.

It is important to emphasize that creating art is still a task restricted to humans, as its creation goes beyond learning patterns and colors. Art communicates a temporal moment, a taste of a specific reality or even the desire not to follow standards – or patterns. What we want to achieve in this work is an improvement in classification performance by maximizing the information provided to classifiers. When generating images with GANs, we are creating new images without the model ever having actually seen real images of artwork, it only knows how close the probability distribution of generated image is to the probability distribution of the real images, a very mathematically-minded scheme.[1]

[1] "A work of art which did not begin in emotion is not art", Paul Cézanne

6.2 Future Work

Some ideas were not explored here, but we believe they would yield future work:

- training classifiers with synthetic images of others classes;
- using style transfer to generate a richer artificial dataset;
- exploring the GAN model's training settings further – for starters, a more robust setup that allows for a larger batch size or training for a longer period of time;
- and mixing artificial images from more than one artistic movements.

The last idea was something we tried only with three art styles - Pop Art, Expressionism and Romanticism. We did not get good results, but maybe this would be a matter of better exploring the combinations of quantities and artistic movements. Could it be that mixing other artistic movements that are more distinct from each other would produce better results?

REFERENCES

ADLER, J.; LUNZ, S. Banach Wasserstein GAN. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2018. (NIPS'18), p. 6755–6764. Cited in page 39.

ARJOVSKY, M.; CHINTALA, S.; BOTTOU, L. Wasserstein Generative Adversarial Networks. In: *Proceedings of the 34th International Conference on Machine Learning - Volume 70*. [S.l.]: JMLR.org, 2017. (ICML'17), p. 214–223. Cited 2 times in pages 38 and 39.

ARORA, R. S.; ELGAMMAL, A. Towards automated classification of fine-art painting style: A comparative study. *Proceedings - International Conference on Pattern Recognition*, p. 3541–3544, 2012. ISSN 10514651. Cited 2 times in pages 21 and 41.

BAHDANAU, D.; CHO, K. H.; BENGIO, Y. Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, p. 1–15, 2015. Cited in page 37.

BAR, Y.; LEVY, N.; WOLF, L. Classification of artistic styles using binarized features derived from a deep neural network. In: . [S.l.: s.n.], 2015. v. 8925, p. 71–84. ISBN 9783319161778. ISSN 16113349. Cited 2 times in pages 41 and 45.

BIANCO, S. et al. Multitask painting categorization by deep multibranch neural network. *Expert Systems with Applications*, Elsevier Ltd, v. 135, p. 90–101, 2019. ISSN 09574174. Cited in page 42.

BROCK, A.; DONAHUE, J.; SIMONYAN, K. Large scale GAN training for high fidelity natural image synthesis. In: *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. [S.l.]: OpenReview.net, 2019. Cited in page 36.

CETINIC, E.; LIPIC, T.; GRGIC, S. Fine-tuning Convolutional Neural Networks for fine art classification. *Expert Systems with Applications*, Elsevier Ltd, v. 114, p. 107–118, 2018. ISSN 09574174. Cited in page 41.

CETINIC, E.; LIPIC, T.; GRGIC, S. How Convolutional Neural Networks Remember Art. *International Conference on Systems, Signals, and Image Processing*, IEEE, 2018. ISSN 21578702. Cited in page 42.

CHE, T. et al. Mode regularized generative adversarial networks. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*, p. 1–13, 2017. Cited in page 36.

CHEN, L.; YANG, J. Recognizing the style of visual arts via adaptive cross-layer correlation. *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, p. 2459–2467, 2019. Cited in page 42.

CHEN, X. et al. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. *Advances in Neural Information Processing Systems*, p. 2180–2188, 2016. ISSN 10495258. Cited in page 36.

CHENG, J.; DONG, L.; LAPATA, M. Long Short-Term Memory-Networks for Machine Reading. In: *EMNLP*. [S.l.: s.n.], 2016. p. 551–561. Cited in page 37.

CHU, W. T.; WU, Y. L. Image style classification based on learnt deep correlation features. *IEEE Transactions on Multimedia*, IEEE, v. 20, p. 2491–2502, 2018. ISSN 15209210. Cited in page 42.

CONDOROVICI, R. G.; FLOREA, C.; VERTAN, C. Automatically classifying paintings with perceptual inspired descriptors. *Journal of Visual Communication and Image Representation*, v. 26, p. 222–230, 2015. ISSN 10959076. Cited 2 times in pages 41 and 45.

DARAS, G. et al. Your Local GAN: Designing Two Dimensional Local Attention Mechanisms for Generative Models. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 14519–14527, 2020. ISSN 10636919. Cited in page 36.

DEVRIES, T.; TAYLOR, G. Dataset augmentation in feature space. *ICLR 2017 workshop*, 2017. Cited in page 33.

ELDEEN, N.; KHALIFA, M. *Detection of Coronavirus (COVID-19) Associated Pneumonia based on Generative Adversarial Networks and a Fine-Tuned Deep Transfer Learning Model using Chest X-ray Dataset*. 2020. Cited in page 34.

ELGAMMAL, A. et al. CAN: Creative Adversarial Networks, Generating "Art"by Learning About Styles and Deviating from Style Norms. *arXiv*, p. 1–22, 2017. ISSN 23318422. Cited in page 43.

ELGAMMAL, A. et al. The shape of art history in the eyes of the machine. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, p. 2183–2191, 2018. Cited 2 times in pages 42 and 49.

FLOREA, C.; TOCA, C.; GIESEKE, F. Artistic movement recognition by boosted fusion of color structure and topographic description. *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017*, IEEE, p. 569–577, 2017. Cited 2 times in pages 41 and 45.

FRID-ADAR, M. et al. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, Elsevier B.V., v. 321, p. 321–331, 2018. ISSN 18728286. Cited in page 33.

GAO, X.; TIAN, Y.; QI, Z. Rpd-gan: Learning to draw realistic paintings with generative adversarial network. *IEEE Transactions on Image Processing*, v. 29, p. 8706–8720, 2020. ISSN 19410042. Cited in page 43.

GOODFELLOW, I. *NIPS 2016 Tutorial: Generative Adversarial Networks*. 2016. Cited in page 38.

GOODFELLOW, I. J.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. Cited in page 34.

GREEN, M. a et al. *Gardner's Art thought the Ages: A Global History*. [S.l.: s.n.], 2011. 585-602 p. ISBN 9781111035181. Cited 2 times in pages 21 and 25.

GULRAJANI, I. et al. Improved Training of Wasserstein GANs. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 5769–5779. ISBN 9781510860964. Cited 2 times in pages 39 and 45.

HAN, C. et al. Synthesizing Diverse Lung Nodules Wherever Massively: 3D Multi-Conditional GAN-Based CT Image Augmentation for Object Detection. *Proceedings - 2019 International Conference on 3D Vision, 3DV 2019*, IEEE, p. 729–737, 2019. Cited in page 33.

HE, K. et al. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778. Cited 2 times in pages 29 and 41.

HEUSEL, M. et al. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 6629–6640. ISBN 9781510860964. Cited in page 36.

HIDAKA, A.; KURITA, T. Consecutive Dimensionality Reduction by Canonical Correlation Analysis for Visualization of Convolutional Neural Networks. *Proceedings of the ISCIE International Symposium on Stochastic Systems Theory and its Applications*, The Institute of Systems, Control and Information Engineers, v. 2017, p. 160–167, 2017. ISSN 2188-4730. Cited 2 times in pages 13 and 29.

HU, J.; SHEN, L.; SUN, G. Squeeze-and-Excitation Networks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 7132–7141, 2018. ISSN 10636919. Cited in page 30.

KARAYEV, S. et al. Recognizing image style. *BMVC 2014 - Proceedings of the British Machine Vision Conference 2014*, p. 1–20, 2014. Cited 2 times in pages 41 and 45.

KASTAN, D. S.; FARTHING, S. *On Color*. [S.l.]: Yale University Press, 2018. Cited in page 32.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, v. 60, p. 84–90, 2012. ISSN 15577317. Cited 3 times in pages 21, 29, and 41.

LECOUTRE, A.; NEGREVERGNE, B.; YGER, F. Recognizing Art Style Automatically in painting with deep learning. *Journal of Machine Learning Research*, v. 77, p. 327–342, 2017. ISSN 15337928. Cited in page 41.

MIRZA, M.; OSINDERO, S. Conditional Generative Adversarial Nets. p. 1–7, 2014. Cited in page 36.

MIYATO, T. et al. *Spectral normalization for generative adversarial networks*. [S.l.]: arXiv, 2018. Cited in page 36.

QIN, Z. et al. A GAN-based image synthesis method for skin lesion classification. *Computer Methods and Programs in Biomedicine*, Elsevier Ireland Ltd, v. 195, 10 2020. ISSN 18727565. Cited in page 33.

RODRIGUEZ, C. S.; LECH, M.; PIROGOVA, E. Classification of Style in Fine-Art Paintings Using Transfer Learning and Weighted Image Patches. *2018, 12th International Conference on Signal Processing and Communication Systems, ICSPCS 2018 - Proceedings*, IEEE, p. 1–7, 2019. Cited in page 42.

SANDLER, M. et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 4510–4520, 2018. ISSN 10636919. Cited in page 30.

SANDOVAL, C.; PIROGOVA, E.; LECH, M. Two-Stage Deep Learning Approach to the Classification of Fine-Art Paintings. *IEEE Access*, IEEE, v. 7, p. 41770–41781, 2019. ISSN 21693536. Cited in page 42.

SHAMIR, L. et al. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art. *ACM Transactions on Applied Perception*, v. 7, 2010. ISSN 15443558. Cited 2 times in pages 21 and 41.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, Springer International Publishing, v. 6, 2019. ISSN 21961115. Cited 2 times in pages 31 and 32.

SUH, S. et al. CEGAN: Classification Enhancement Generative Adversarial Networks for unraveling data imbalance problems. *Neural Networks*, Elsevier Ltd, v. 133, p. 69–86, 2021. ISSN 18792782. Cited in page 34.

SZEGEDY, C. et al. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 1–9. Cited in page 29.

SZEGEDY, C. et al. Rethinking the Inception Architecture for Computer Vision. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 2818–2826. Cited in page 29.

TAN, M. et al. Mnasnet: Platform-aware neural architecture search for mobile. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, v. 2019-June, p. 2815–2823, 2019. ISSN 10636919. Cited in page 29.

TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. *36th International Conference on Machine Learning, ICML 2019*, v. 2019-June, p. 10691–10700, 2019. Cited 3 times in pages 13, 29, and 30.

TAN, W. R. et al. Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. *Proceedings - International Conference on Image Processing, ICIP*, v. 2016-Augus, p. 3703–3707, 2016. ISSN 15224880. Cited 2 times in pages 41 and 45.

TAN, W. R. et al. ArtGAN: Artwork synthesis with conditional categorical GANs. In: *2017 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2017. p. 3760–3764. Cited in page 43.

TOMEI, M. et al. Art2real: Unfolding the reality of artworks via semantically-aware image-to-image translation. In: . [S.l.: s.n.], 2019. v. 2019-June, p. 5842–5852. ISBN 9781728132938. ISSN 10636919. Cited in page 43.

VASWANI, A. et al. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2017. (NIPS'17), p. 6000–6010. ISBN 9781510860964. Cited in page 37.

WAHEED, A. et al. CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection. *IEEE Access*, v. 8, p. 91916–91923, 2020. ISSN 21693536. Cited in page 34.

WANG, Z.; SHE, Q.; WARD, T. E. Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy. *ACM Comput. Surv.*, Association for Computing Machinery, New York, NY, USA, v. 54, n. 2, fev. 2021. ISSN 0360-0300. Cited 3 times in pages 13, 35, and 38.

WU, J. et al. Wasserstein Divergence for GANs. In: . [S.l.: s.n.], 2018. v. 11209 LNCS, p. 673–688. ISBN 9783030012274. ISSN 16113349. Cited in page 39.

WU, Y. et al. End-to-End Chromosome Karyotyping with Data Augmentation using GAN. *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, p. 2456–2460, 2018. Cited in page 33.

XUE, A. End-to-End Chinese Landscape Painting Creation Using Generative Adversarial Networks. *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, p. 3862–3870, 2021. Cited in page 43.

YILDIRIM Özal et al. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. 2018. Cited in page 33.

ZHANG, H. et al. Self-Attention Generative Adversarial Networks. In: CHAUDHURI, K.; SALAKHUTDINOV, R. (Ed.). *Proceedings of the 36th International Conference on Machine Learning*. [S.l.]: PMLR, 2019. (Proceedings of Machine Learning Research, v. 97), p. 7354–7363. Cited 4 times in pages 13, 36, 37, and 45.

ZHONG, S. hua; HUANG, X.; XIAO, Z. Fine-art painting classification via two-channel dual path networks. *International Journal of Machine Learning and Cybernetics*, Springer Berlin Heidelberg, v. 11, p. 137–152, 2020. ISSN 1868808X. Cited in page 42.

ZHU, J. Y. et al. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: . [S.l.: s.n.], 2017. v. 2017-Octob, p. 2242–2251. ISBN 9781538610329. ISSN 15505499. Cited in page 43.

ZHU, Y. et al. Machine: The New Art Connoisseur. In: . [S.l.: s.n.], 2019. abs/1911.10091. Cited in page 42.