Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science and Statistics

Exploring the Impact of AI and Deep Learning in Digital Scenography

Vasiliki Karathanasi

A research paper submitted to the University of Dublin,

in partial fulfilment of the requirements for the degree of

Master of Science Interactive Digital Media

2023

Declaration

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at: http://www.tcd.ie/calendar

I have also completed the Online Tutorial on avoiding plagiarism 'Ready, Steady, Write', located at: https://libguides.tcd.ie/plagiarism/ready-steady-write

I declare that the work described in this research paper is, except where otherwise stated, entirely my own work and has not been submitted as an exercise for a degree at this or any other university.

Signed: ___________________

Vasiliki Karathanasi

22/05/2023

Permission to lend and/or copy

I agree that Trinity College Library may lend or copy this research Paper upon request.

Signed: __________________

Vasiliki Karathanasi

22/05/2023

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Néill O'Dwyer, for his invaluable guidance, encouragement and support throughout the course of this research.

I would also like to extend my appreciation to all the IDM lecturers for their expertise and the knowledge they shared on a wide range of engaging topics.

Finally, I would like to thank my coursemates for our insightful conversations and for contributing to a creative and supportive academic environment.

Abstract

This paper explores the applications of AI and deep learning systems in digital scenography, with a specific focus on their role in scenic design processes. Digital scenography refers to the use of digital tools in creating stage settings and visual elements for live performances, such as theatre and opera productions. This interdisciplinary field encompasses both performance practices and design processes. For the purposes of this paper, the scope is narrowed to the utilisation of AI tools in the scenic design process. Within the design process, the emphasis is placed on the pivotal stage of content generation, where the interplay between digital tools and digital scenography unfolds. In recent times, the emergence of AI-enhanced tools has revolutionised the generation of concept art, enabling designers to amplify their imagination and creativity and produce innovative ideas with unprecedented speed and quality. These cutting-edge tools, categorized as image-to-image and text-to-image systems, harness the latest advancements in AI and deep learning. By seamlessly blending artistic styles and arbitrary images or by interpreting simple textual prompts provided by designers, these systems generate appealing concept art that revitalises the artistic process. The paper's findings revolve around the efficiency of these AI-accelerated systems in generating concept art. Not only do they provide contemporary artists with invigorating ideas, but they also expedite the overall workflow, allowing designers to actualise their vision with greater efficiency.

Table of Contents

Table of Figures

1. Introduction

This paper aims to discuss the recent developments in digital scenography afforded by the fast-paced technological advancements taking place in the digital age. The advent of digital technologies has reshaped traditional performance and design practices in theatres and operas contributing to the evolution of digital scenography. Digital Scenography is a broad term that refers to the use of digital technologies for the creation of a variety of different elements in contemporary theatre, aiming to provide more engaging and immersive experiences for audiences. Developments in performance practices that are only possible using modern digital tools are discussed, however, the paper focuses mostly on the innovations in design practices and content generation in scenography, particularly facilitated by the advent of artificial intelligence and deep learning.

1.1. Digital Scenography and Definitions

Scenography

Scenography is a modern but also an ancient term (Aronson, 2005). The word "scenography" comes from the Greek words "skene" (σκηνή) meaning "stage", and "graphia" (γραφία) meaning "writing" or "drawing". Aronson describes "scenography" as a term that stems from "scene building", as well as Aristotle's references to "skenografia", or "scenic writing". Scenography refers to the creation of the visual and spatial aspects of a performance, including a physical scenery, props, costumes, lighting, and other visual components. Traditionally, scenography involves producing these elements using materials such as wood, fabric, metal, and paint. It also involves working with lighting and sound to create an overall visual and sensory experience for the audience. However, scenography is not restricted to the creation of scenery, costumes and lights. As Lotker and Gough suggest, scenography is not just a design or illustration process, it is a whole discipline with its own rules (Lotker and Gough, 2013). Many scenographers argue that the term "scenography" encompasses the spatial construct, as well as the theatrical text, the performers and the audience

(Aronson, 2017). Others view scenography as a dynamic contribution to the experience of performance correlating scenography with other theatrical terms like mise-en-scène, theatre design, and visual dramaturgy (McKinney and Butterworth, 2009). Howard defines scenography as a combination of the factors: space, text, research, art, actors, directors and spectators. More specifically, Howard argues that scenography requires an active relationship between the performer and the spectator and is incomplete without the performer on stage actively engaging with the audience (Howard, 2001). The definitions are varied and often contradictory and for that reason, they are discussed thoroughly in the context of digital scenography.

Digital Scenography

In the context of digital scenography, the term "digital" refers to the technology utilised. Thus, digital scenography builds and expands on the definition of scenography by incorporating digital tools into the design and production process. Digital scenography describes the use of digital tools for the creation and design of stage settings and visual elements in live performances, such as theatre and opera productions, dance performances and concerts. These digital tools encompass technologies such as projection mapping, virtual and augmented reality, interactive installations, and other digital media to create immersive and interactive stage designs. The above digital technologies are aimed to enhance all aspects of performance, create unique and immersive experiences for audiences, and expand the creative possibilities for designers and artists. Digital scenography has opened up new possibilities for storytelling and audience engagement, allowing designers to create more complex and visually powerful productions than ever before. It also expands the creative alternatives for designers and artists, by facilitating dynamic and flexible stage environments that can be easily customised. However, it has also introduced new challenges, such as integrating digital elements with live performers seamlessly and ensuring that technical glitches do not disrupt the flow of the performance.

In the past, many scholars have proposed a grammar to define the use of digital media in performance for example "filmstage" by Michael Kirby (1966). Auslander (1999) refers to the ‘mediatisation’ of live performance, while Chapple and Kattenbelt (2007) propose ‘intermediality’ to highlight the interplay of live and digital aspects in theatrical space (Vincent, 2021). Other terms that describe this type of technologically engaged performance are multimedia, cyborg/cybernetic (Parker-Starbuck, 2011), new/digital media, virtual, etc. Essentially, they all refer to the performances that creatively utilise digital media technologies for innovative results.

Many scenographers emphasise the use of digital tools as simple means to the creation of the desired result. Vincent uses the term “digital” to refer to all digital media used on stage that affect the scenographic design (Vincent, 2021). This relates to digital performance, described by (Dixon, 2007) as a performance in which “computer technologies play a key role in content, techniques, aesthetics, or delivery forms”. The term digital contains computer technologies, including video, projection, animation, motion capture, and real-time interactivity. According to (Palmer, 2017) digital scenography can be understood as the technology being used to create projections on stage, rather than the projected effects themselves.

Other contemporary definitions of scenography describe it as a whole environment, a world to be explored with all senses which is inseparable from the performance (O’Dwyer, 2021). As such, digital scenography is seen as a theatrical experience that involves active engagement between the performer, a scenic setting in which digital technologies play a major role, and the spectator (Vincent, 2021). Digital scenography can be considered the epitome of techno-cultural innovation, which is afforded by the evolution of the scientific areas: bandwidth, nanotechnologies, processing power, artificial intelligence, and biomechanical engineering (O’Dwyer, 2021). Digital scenography is also widely considered an interdisciplinary field that champions audio-visual, participatory, immersive, site-specific and design-led approaches to performance practice (Thornett, 2017). In this interdisciplinary field, scientific-artistic-hybrid inventions flourish by re-purposing technologies for cultural applications (O’Dwyer, 2021). Another interesting aspect concerns the use of technology. Christopher Baugh supports that the technologies employed for digital scenography are not simply a means to an end; they are frequently used as ends in themselves (Baugh, 2013).

Overall, digital scenography is interpreted in various ways, however, most of the definitions and explanations above share the notion that scenography includes all aspects of the theatrical production, including the performers and audience, and is an essential component of the overall experience. While both traditional scenography and digital scenography are targeted towards creating a representative visual world for a performance, digital scenography offers innovative ways of engaging with audiences and can enhance the overall impact of a production through the use of cutting-edge technologies. Digital scenography is not a replacement for traditional scenography, but rather an extension of it, offering new possibilities for designers and artists to explore and experiment with.

1.2. A brief history of Digital Scenography

Although digital scenography is often portrayed as an innovation, a study of its history reveals that the pursuit of a synthesis between the performer, the stage design, and the audience has been fundamental to the genre from its very beginning. Baugh (2013) claims that "stage technology, machinery and special effects have always been a part of the experience of theatre and performance" and cites examples dating from Ancient Greece.

As early as the 1900s, practitioners across Europe started incorporating elements of film into their scenographic designs. Early multimedia opera productions made use of film footage in familiar ways, such as for background scenery, narrative sequences, and special effects. Despite being a technological innovation, the initial use of film in opera served a similar aesthetic purpose as the perspective scenery, stage machinery, and magic lantern projections of previous centuries. In the early 1920s, the emergence of modernist and avant-garde movements triggered a broader change in perspective regarding the integration of film and other innovative technologies with live performance. Essentially, digital performances emerged from the legacy of the avant-garde (O'Dwyer, 2021).

During the early to mid-1990s a group of artists emerging from the worlds of electronic music, video art, performance art, and theatre began to integrate new digital technologies into live performance (Saltz, 2013). This group of artists sought to pave a new path in performance arts by employing the latest innovations in digital technology. In 1998, the opera Monsters of Grace by composer Philip Glass and stage director Robert Wilson premiered at Wolf Trap Opera in Virginia. The opera gained much attention for its innovative use of digital technology, which included 3D computer-animated films projected on a screen above the performers and viewed by the audience through polarised glasses. This was complemented by live music that was synchronised with the films, creating a theatrical experience that immersed the audience in the world of technology. Described as a "digital opera", this production was widely regarded as a significant milestone in the evolution of digital performances. However, it lacked integration between the live performers and the 3D digital projections and suggested that until artists have a better understanding of the capabilities and limitations of technology, there is still a lot to be explored and developed in digital scenography (Vincent, 2021). It is also important to reflect on the difference between postmodern performance and digital media performance: simply displaying videos or imagery using a digital projector, or screen, does not straightforwardly qualify a performance, set design, prop or costume as digital scenography (O'Dwyer, 2021).

The digital age has encouraged the bloom of new hybrid visual arts. Artists have embraced the fast-paced changes in culture and have developed new methodologies for scenography and performances. These artists have reformed the direction of culture and technological-scientific development, by creatively inventing hybrid artworks. Despite the fact that modern techniques and technologies have become more sophisticated, the fundamental goal of scenography remains the same: to blend the performer and the scenery to provide the audience with a more seamless theatrical experience.

1.3. Digital Scenography in modern theatre and opera

As theatre is inclusive and collaborative by its nature, it functions as an experimental intersection of Arts, sciences and technology. The increased use of technology on the live stage has transformed the theatre into a space for experimentation, investigating not just human-to-human interaction but also human-machine interaction (O'Dwyer, 2021). The traditional methods of mise-en-scène are being steadily enhanced by innovative approaches using new materials and technologies. This stems from the spectators' need to gain more immersive experiences in the digital age. Therefore, there is a significant increase in theatrical performances that experiment with new ways of interaction between the actors and the spectators in order to attract diverse audiences and ensure participation in a fast-consuming digital age. Under these circumstances, digital scenography provides the chance to reshape the traditional theatrical framework and define both genre and repertoire. Essentially, scenography is the domain that allows the explorative combination of theatre with subjects from science, technology and engineering.

The digital age has redefined opera in a similar way. The main difference between opera and theatre lies in the fact that opera is a musical art form, so it focuses on the music, whereas theatre typically focuses on the text. In opera, the composer's work takes precedence and is then interpreted by the director. In theatre, the emphasis is usually put on the written script. In recent years, the opera industry has been described as monotonous and repetitive resulting in many operas struggling to remain financially viable. At the same time, it is particularly challenging to create something visually stimulating in the digital age, especially for the youth. Digital scenography is seen as an ideal way to revitalise classic opera works to suit a more modern perspective as well as a way for opera productions to save money. Digital technology has thus been framed as the solution to this kind of issue (Vincent, 2021).

In an attempt to describe the degree to which digital technologies impact performances and theatrical productions, Caitlin Vincent proposed a classification system named "Modes of Synthesis".

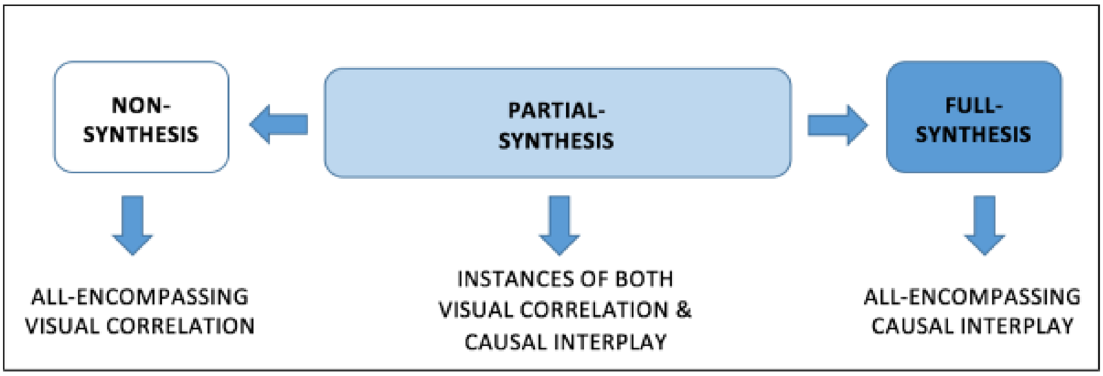

Figure 1: "Modes of Synthesis".

- Non-synthesis: The classification of non-synthesis essentially refers to a superficial use of visuals. Digital elements, such as digital projections, integrated into the opera or theatrical play are necessary for providing a background setting or other effects as they serve as a backdrop, but the performers do not interact with the digital imagery. The digital imagery functions merely as an added layer or texture and their function is strictly aesthetic.
- Partial-synthesis: In partial-synthesis production, both the physical and digital components are essential. At times, digital elements serve as a backdrop exactly like in non-synthesis, but the main difference is that most of the time the projected imagery also facilitates a form of "causal interplay" between the visuals and the live performers. This is done in various ways, for example, when the physical movements of the actors appear to shift the projected imagery in a synchronised way.
- Full-synthesis: Full-synthesis concerns digitally-enhanced productions demonstrating a full integration of live and digital elements. These productions combine digital and live aspects, in terms of both the scenographic design as well as the interaction between the actors and the stage space.

This relationship can be identified as a strategic choice made by the director and creative designers, which leverages available technologies in order to realise the desired narrative on stage.

1.4 Technologies used in Digital Scenography

Nowadays, various digital technologies play a significant role in the enhancement of theatrical design and contribute to cutting-edge interpretations of operatic works (Vincent, 2021). These technologies, which are more commonly used in other scientific disciplines, are now being creatively repurposed leading to unique performances with immense cultural impact. Some examples of digital scenography technologies in practice include video mapping, digital projections, holographic projections, motion capture, virtual and augmented reality and game engines. While these technologies provide a different perspective to digital scenography, there are significant problems relating to the synchronisation of the actors' performance and visual effects.

To address this problem (Roça et al., 2022) propose a creative scenario combining simple software with projection techniques, that can be dynamically adapted during the opera and interact with the artists in real-time. The system proposed by Roça et al. gathers data from the stage using computer vision techniques and then creates real-time digital effects with respect to the opera's author preferences. The results of this work are applied to the electronic opera "TMIE, Standing on the Threshold of the Outside World".

A set design can be partially or fully created with digital artefacts. Creative use of digital projection offers a new layer of story interpretation. A good example is "The Jew of Malta, 2002" where the actors' white costumes were utilised as a canvas surface for the projections which were used to visualise their thoughts and emotions. The main technology employed for that is an image recognition system that can spot the actors' silhouettes in real-time. These silhouettes are used for the creation of virtual masks that are later projected onto the actors.

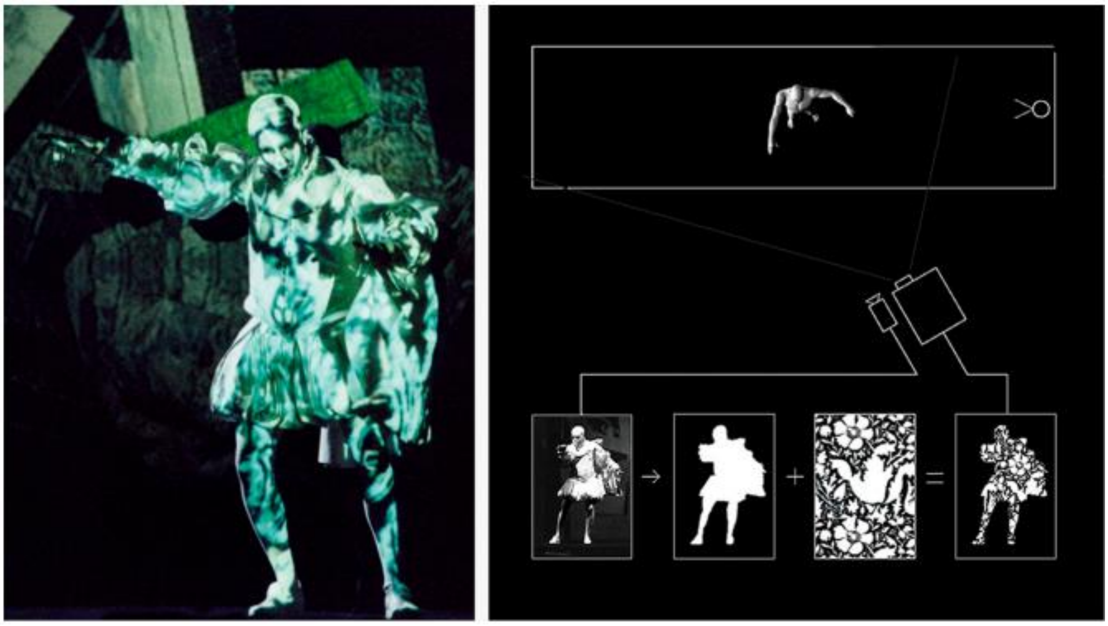

Figure 2: Image recognition and computer vision techniques applied to the opera "The Jew of Malta" (Roça et al., 2022).

Opera productions use digital projections as the digital equivalent of background sceneries. The book "Digital Scenography in Opera in the twenty-first century" references numerous productions for their use of digital technology to create background settings and environments that dynamically change during the narrative of the opera. Digital technologies are also often used for challenging staging requirements such as simulating supernatural effects. A good example is Mozart's fantastical opera The Magic Flute, frequently staged with digital scenography to dramatise the trials of fire and water. Another use is the creation of autonomous virtual entities that interact with or even replace the actors on stage as well as animated video sequences, 3D stereoscopic effects, and virtual performers as presented in English National Opera's The Sunken Garden (2013) (Vincent, 2021).

Projection design or video design is a very common practice in scenic design, however, it presents challenges as far as the positive and negative space is concerned. From a physical sculptural point of view, putting up a screen as part of the set is equal to putting up a flat plane on stage. Even though the scenery or video projected provides a window into content that appears to be rich and fits the story of the play, from that standpoint of the negative and positive space the creative possibilities are essentially constrained by a "wall" which does not provide interesting alternatives for lighting and shadows. In that sense, projection design is efficient only when interesting content is available throughout the duration of the play (Jamerson, J. personal communication 17/04/2023).

Figure 3: "Faux interactivity" as described by (Vincent et al., 2016) is a digital projection of pre-rendered imagery combined with pre-determined choreography that creates the illusion of interactivity. Moby-Dick (2010).

Figure 4: The Magic Flute, the actors are standing on an elevated platform while "digital water" surrounds them.

1.5 Hypothesis

One could claim that digital scenography is comprised of two components: technological tools and cultural artefacts. The tools refer to all the technological digital advancements that are in some way utilised for creative purposes. The cultural artefacts stem from the artists' creative exploitation of the aforementioned tools. The technological tools utilised in digital scenography are not just employed as means to serve the artistic result. They function as individual components, crucial both for the implementation of set design and to inspire artists for new rules, methodologies, and work processes for modern performances.

For the purposes of this paper, the focus is shifted towards the technological components, the newly introduced digital tools and their potential use in enhancing the content generation and design processes. In this context, the research question of this paper is summarised as follows: How the current technological advancements in Artificial Intelligence and Deep Learning can be utilised for scene design and enhance digital scenography? Specifically:

- How Neural Style Transfer as an image-to-image deep learning technique can be utilised not only for the creation of 2D images but also for stage design and,
- How text-to-image techniques can be applied for novel scenographic approaches?

Digital Scenography is concerned with theatrical experiences that involve interaction between the performer or actor, the spectator, and scenic environments enhanced with digital technologies. The methods utilised for set design are so far constrained to specific technologies, such as computer vision, motion tracking, projection mapping, biometric sensing, real-time data analysis, signal processing, and 3D modelling and graphics. However, artificial intelligence applications are not being implemented today in digital scenography to the extent that they are in traditional scientific disciplines as they are still in an experimental stage with more and more artists exploring their potential.

Numerous Neural Style Transfer methodologies are effective in faithfully representing a painting's style, and thus the combination of an arbitrary image and a painting can lead to the creation of various 2D digital artworks, which function as a source of inspiration for artists and designers. Neural Style Transfer for 2D images and videos is a thoroughly researched topic, with scientists proposing variations of the same technique and producing interesting results. More recent research on this topic extended the traditional methodologies to 3D Neural Style Transfer, with various applications to 3D modelling and stereoscopic images.

Similarly, numerous efficient text-to-scene systems have been proposed in the past years for the creation of 2D images and 3D environments based on descriptions extracted from text. With the advent of AI and deep learning, more powerful systems are built.

Thus, the aim of this research paper is to discuss the possibility of expanding the above scientific domains to include artificial intelligence and deep learning applications. AI and deep learning have brought profound developments in various scientific fields and they are promising in contributing to the field of digital scenography. This research paper will identify the interplay between digital scenography and AI-enhanced systems and suggest a potential use of Neural Style Transfer and Natural Language Processing techniques in the scenic design process.

2. Artificial Intelligence and Deep Learning in Digital Scenography

2.1. Definition of Artificial Intelligence and Deep Learning

This section provides fundamental definitions and background knowledge to understand how AI models are used in digital scenography.

Artificial Intelligence

Artificial Intelligence (AI) mimics functions of human behaviour and performs complex tasks. These tasks include problem-solving, object recognition and processing of large data. It is used in a variety of areas like advertising, image processing, medical diagnosis and so on (IBM Cloud Education).

Machine Learning

Machine Learning (ML) is a branch of Artificial Intelligence which focuses on the specific use of data and algorithms so that the system mimics the way humans learn. Such a system is trained, that is, automatically learns to improve, without previously being programmed for it. After being fed with many examples regarding a specific task, the system gradually recognizes a statistical structure among the examples and eventually defines rules that automate a decision or prediction process (Chollet, 2017).

Deep Learning

Deep Learning (DL) is a subset of Machine Learning. It simulates the functions of the human brain, allowing deep learning systems to gather data and make predictions with extreme precision (IBM Cloud Education). The key idea of DL, which refers to the model learning useful representations from input data, is similar to Machine Learning. The key difference, however, lies in how each algorithm performs the learning process. Specifically, deep learning emphasises learning through successive layers that contain increasingly useful representations of the input data. The term "deep" refers precisely to the idea of multiple successive levels of learning. Thus, the model's depth essentially represents the number of levels used for learning (Chollet, 2017). Layer representations are learned through models called Neural Networks (NNs).

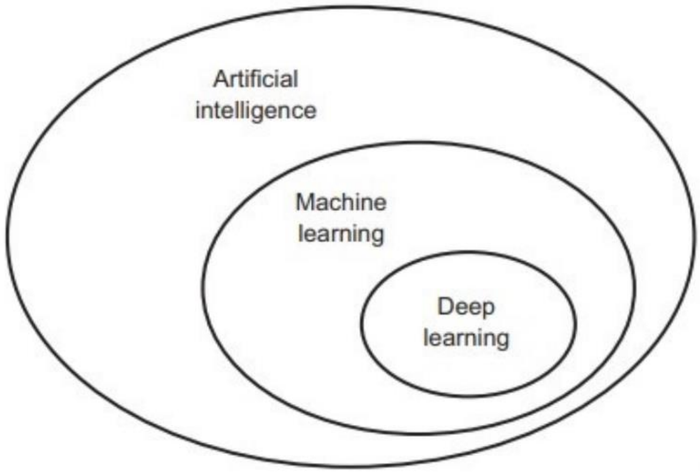

Figure 5: The relationship between artificial intelligence, machine learning and deep learning

Neural Networks

Neural Networks, also known as Artificial Neural Networks, are similarly a subset of deep learning. A neural network consists of layers of neurons: an input layer, one or more hidden layers and an output layer. Each neuron, or node, is connected with the neurons of the previous and next layer through the so-called synapses. For the neural network to be utilised, the training process must take place. The most fundamental function of neural networks is learning, which is achieved through training. Training is an iterative process of adapting the neural network's initial parameters. It is performed using large datasets usually consisting of millions of images and their corresponding labels.

Convolutional Neural Networks

The subclass of Neural Networks that are most useful in image processing is Convolutional Neural Networks (CNNs). CNNs are very effective in pattern recognition and detection, as well as image classification. The layers of a convolutional network consist of nodes, that is computing units that process information hierarchically. Every layer functions as a collection of different image filters. During the training process, each filter extracts a different and specific feature of the input image. Therefore, every input image is encoded and different representations of it are produced. These representations are then processed so that the model generates the desired output. In the context of content generation, the outputs are artistically modified images.

Generative Adversarial Networks

Generative Adversarial Networks (GANs) constitute another type of Neural Networks or else models, which are powerful in generating realistic, high-resolution images. The structure of GANs is different in that they are implemented using two neural networks, the generator and the discriminator. The training of a GAN is done through the interplay between these two neural networks. More specifically, the system is fed with a large dataset of images. The generator is then tasked with generating new plausible examples using that dataset. The discriminator, on the other hand, tries to classify examples as real (existing images of the dataset) or fake (generated by the generator). After each iteration, the discriminator is updated to improve at distinguishing fake examples from real ones, and similarly, the generator is updated based on how well the generated samples fooled the discriminator. This iterative process of learning makes GANs powerful at generating highly representative images (Goodfellow et al., 2020).

Diffusion Models

Diffusion models provide a powerful alternative to traditional generative models, surpassing the efficiency of GANs in artificial synthesis. They can be described as machine learning systems that are trained to denoise random Gaussian noise step by step, to get to a sample of interest, such as an image (Rombach et al., 2022). Diffusion models iteratively apply diffusion during a series of steps and update the Gaussian distribution. As part of the system, a neural network is trained to predict a way to slightly denoise the image in each of these steps. After a certain number of steps, a sample is obtained. Diffusion models play a significant role in text-to-image systems like DALLE-2.

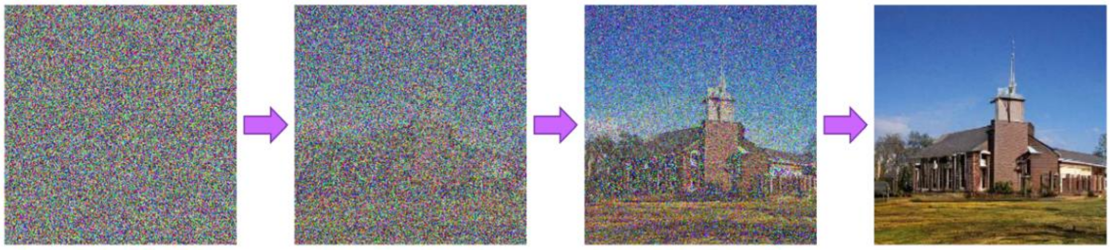

Figure 6: Illustration of the process taking place in Diffusion models ("Diffusers.ipynb,").

Variational Autoencoders

Variational Autoencoders (VAEs) are used, among others, as part of the Diffusion models. Variational Autoencoders are efficient in synthesising high-resolution images (Rombach et al., 2022). As a variation of traditional autoencoders, they function as neural networks used for unsupervised learning. A VAE consists of an encoder network and a decoder network: the encoder takes the input data and creates a lower-dimensional (latent) representation of the input data, while the decoder takes a sample of the latent output and recreates the original input data. In the context of image synthesis, the encoder converts an input image into a low-dimensional representation and the decoder reconstructs the latent representation back to the original image.

2.2 Overview of how Artificial Intelligence and Deep Learning can be used in Digital Scenography

The overarching argument of this paper is that AI systems can reshape traditional design practices in digital scenography by fast generating high-quality concept art. In this section, we discuss the nature and usability of AI-enhanced digital tools that are already employed or could be employed in digital scenography. Most of the AI tools that are analysed below are very recently developed and are still being optimised or fine-tuned to generate more appealing results. As artificial intelligence systems evolve, the creative possibilities are rapidly expanding providing space for experimentation and innovative ideas to be realised. The interplay between AI and digital scenography takes form through the content generation process.

2.2.1 Content Generation

Content generation refers to the use of AI algorithms that are powerful in generating content automatically. For example, various specialised algorithms can create 3D models, textures, and artificial images from scratch. This can save time and effort for designers and allow for the creation of more complex and detailed virtual environments. In the context of digital scenography, generating concept images is an essential and time-consuming part of the design process.

2.2.2 Design Process

The engineering design process, as presented below, is an iterative process consisting of a series of steps applicable to many domains, including scenic design.

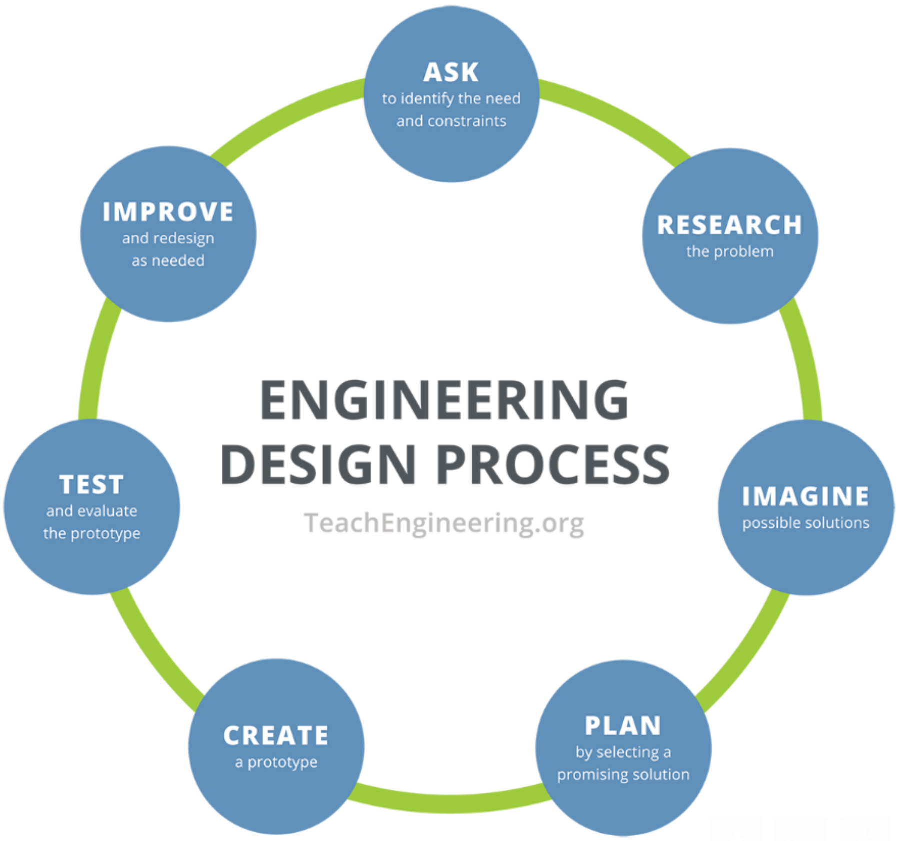

Figure 7: Design process (https://www.teachengineering.org/populartopics/designprocess).

The steps required for each type of design vary significantly, however, most of the different design methodologies focus primarily on four design stages (Mayda and Choi, 2017):

1. Product planning
2. Conceptual design
3. Embodiment design
4. Detail design

The conceptual design or concept generation stage in many disciplines is most commonly realised with sketching and drawing. That is justified by the relation of thinking and drawing, that is, drawing is a fast and effective way to visualise a designer's idea (Won, 2001). With the advent of computer-aided design in the past years, traditional design practices are being reshaped. Even though computer-aided methods are largely utilised during the embodiment and detail design stages, they now have an impact on the conceptual design as well. AI tools, in particular, prove to be extremely useful in generating fast conceptual designs. In addition, the conceptual designs generated by AI tools present new ideas that have not been imagined before, amplifying human imagination and aesthetic intelligence.

Assuming that conceptual design is one of the most crucial stages of any kind of design, including digital scenography, we explore the usability of AI-generated art in traditional scenic designers' working methodology (Dills, 2018). Even though each designer's creative workflow varies significantly, one could argue that a basic scenic design for theatre consists of the following stages:

1. Reading and reflecting on the theatrical script
2. Researching the theatrical play
3. Visualising a design
4. Developing the design

The third stage, visualising a possible design, is a crucial part of the design process. It is a very unique process for each designer and involves a lot of experimentation stages as well as a number of rough designs. The tools utilised for that are numerous and include digital imaging tools like Photoshop, Procreate and Illustrator, and 3D modelling software like Blender, 3DS Max, and Cinema 4D. In this particular stage, content generation with AI-accelerated tools can be utilised as an alternative approach to help scenic designers develop innovative ideas and realise their desired sceneries.

2.3 Discussion of AI techniques for Content Generation: Image-to-Image and Text-to-Image

The two AI-enhanced methods for content generation discussed in this paper are text-to-image and image-to-image, afforded by the rapid development of neural networks, specifically CNNs and GANs. Text-to-image systems use Natural Language Processing and Artificial Intelligence techniques that generate images based on input texts. Designers leverage this technique by using extracts from theatrical texts as an input description of a particular set or environment they wish to visualise. The system can then generate a visual representation of that description and provide designers with useful inspiration in the early stages of design. On the other hand, image-to-image techniques refer to models that take as input arbitrary images and iteratively produce an interesting blend of the input images. This is particularly useful in concept art generation when the desired style of the output is very specific and matches the artistic style of a famous painter.

2.3.1 Image-to-Image Techniques

For thousands of years, painting has been one of the primary ways of expressing human nature and emotions. Painting and drawing enable people to substantiate their thoughts and their perception of the world around them. Significant painters have captured their ideas with different techniques-styles, which are difficult to copy or reprint. Nowadays, paintings and artworks attract the attention of various artists as well as scientists.

Specifically, from the 1990s onwards, various research fields were developed with the aim of processing and transforming images into digital artworks, in search of innovation at the intersection of computer science and art. Thus, the scientific field called Style Transfer eventually emerged. The central idea of Style Transfer is to automatically transform a given image into a new image of artistic interest by co-opting an acclaimed or established artistic style. Style transfer has had impressive results in reproducing the artistic styles of great painters. Over the years, there have been various approaches to artistic image modification. A significant example is Non-Photorealistic Rendering (NPR) (Strothotte and Schlechtweg, 2002), in the field of Computer Graphics and Texture Synthesis (Efros and Freeman, 2001) under the overarching knowledge area of Computer Vision. However, all these methods proved unsuitable for depicting arbitrary styles (Kyprianidis et al., 2013) as they were designed for a specific artistic style at a time. To tackle this flexibility issue, Neural Style transfer was proposed.

Neural Style Transfer

Neural Style Transfer, as a research field, was first presented in the work of Gatys et al. "A Neural Algorithm of Artistic Style" (L. Gatys et al., 2016) and is based on the recent advancement of Neural Networks, specifically Convolutional Neural Networks. The key point of this work is that Convolutional Neural Networks are efficient in extracting semantic information from an image. Therefore, it is possible to separate the information related to the content of an image, from the information related to the style of a famous painting. Style transfer of a selected style to a content image is successful when both the structural elements of this image and the artistic style of the painting are preserved.

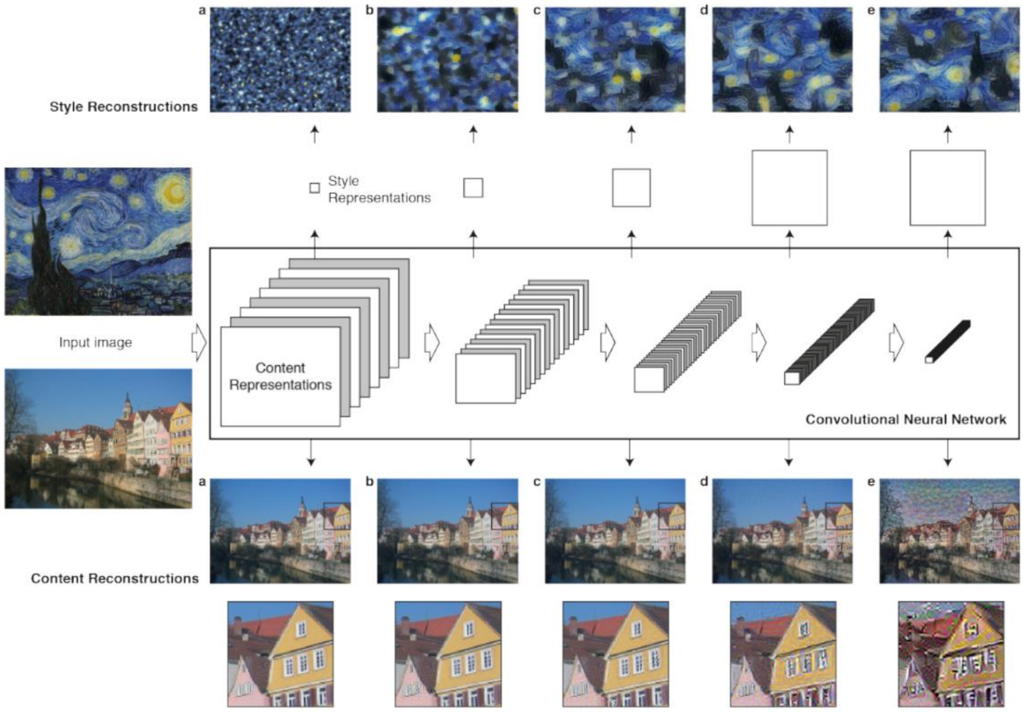

Figure 8: Disentanglement and extraction of the semantic information and style of the input images (L. A. Gatys et al., 2016).

- Content is the abstract structure of the image. It refers to all the objects depicted in the photograph, which can be perceived and described by humans.

- Style is related to image attributes such as colours and textures, which are difficult to be defined and described universally. For example, different visual patterns, combinations of different colour shades, and strokes of different sizes (Gatys et al., 2017).

Neural Style Transfer methods are divided into two subcategories: Image-Optimisation-Based Online Neural Methods and Offline Neural Methods based on Model Optimization. The first category is concerned with all algorithms that perform style transfer and iteratively optimise an initial image. On the contrary, the second category is comprised of the algorithms that first optimise a model and then produce the artistic image in a single pass.

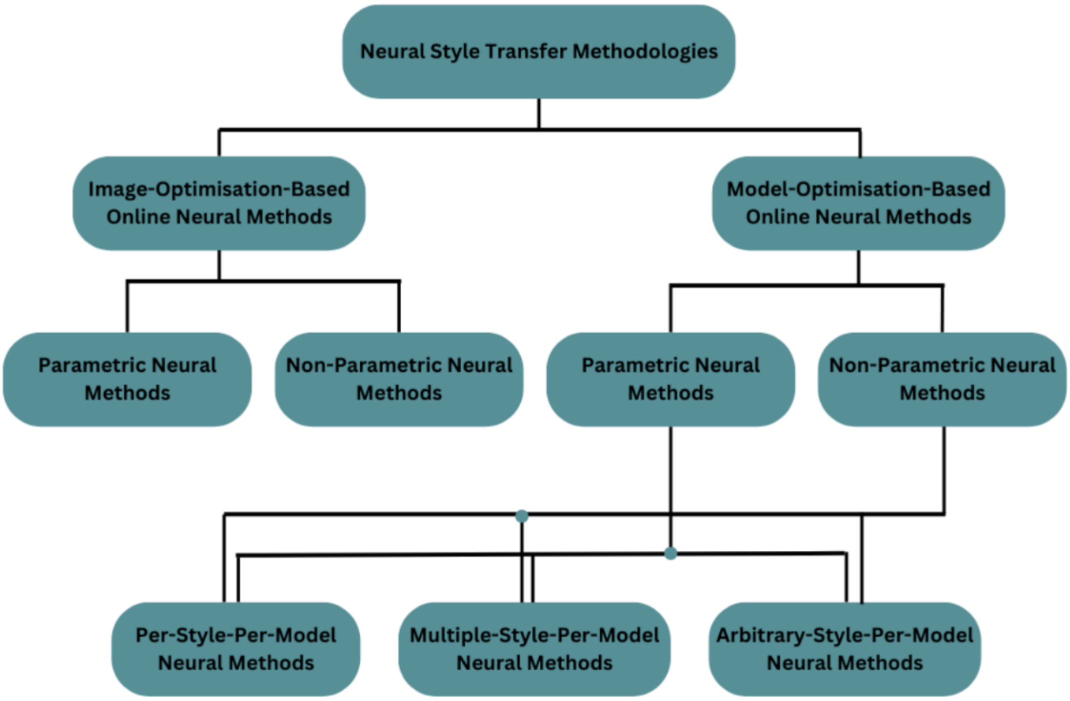

Figure 9: Classification of Neural Style Transfer Methodologies (Karathanasi, 2022).

Online Neural Methods

Online Neural Methods iteratively optimise the initial image. Algorithms using Convolutional Neural Networks first aim to extract and define content and style and then perform the combination. The combination is obtained by initialising a target image with random noise, which is iteratively reconstructed until satisfactory results are obtained. Most algorithms follow the same image reconstruction process, based on gradient descent (Jing et al., 2020). Depending on how style is defined and therefore extracted from the reference image (artistic image), NST algorithms are further divided into parametric and non-parametric methods. The traditional parametric method is the first Neural Algorithm for Artistic Style by Gatys et al. (L. A. Gatys et al., 2016). In a later paper, Gatys et al. proposed an improved parametric method to control image-specific information related to space, colour, and scale. Separating and controlling these three factors widens the variety of results and facilitates the production of detailed, high-resolution images (Gatys et al., 2017). The key disadvantage of the online methods is that they are computationally expensive and time-consuming due to the iterative optimization process, especially when the size of the reconstructed image is large.

Offline Neural Methods

As far as Offline methods are concerned, the optimization is performed a priori, during the model's training, thus the result is produced in a single forward pass. More specifically, to perform the reconstruction of the final artistic image (the target image) based on a given artistic style, optimization of the neural network is required. That is, the network has previously been trained on a large dataset of images, for one or more styles. The specific algorithms pre-train a feedforward network and produce the final stylized result with one forward pass at the testing stage. Algorithms(Johnson et al., 2016), (Ulyanov et al., 2016) follow the original algorithm by Gatys et al (Gatys et al., 2016). In contrast with Online Methods, all fast algorithms belong to this category. The computational cost as well as the execution time are reduced due to prior optimization of the model. Thus, the final artistic image can be generated in just a few minutes.

Overall, offline methods surpass the efficiency of online methods by introducing Fast Style Transfer. Nowadays, Fast Style Transfer and the so-called "Neural Filters" are used for fast stylisation of images in various web or mobile applications and image editors, and are also available online as open-source code. Using one of the widely available software online, the designer can blend an initial conceptual sketch or design of a theatrical stage with a painting of their choice and generate artistically enhanced results. Transferring the style of a desired painting to a concept set design enhances the visualisation in the design process and provides designers with inspiration for a more artistic perspective in their work.

Stereoscopic Neural Style Transfer

Neural Style Transfer enables the fast integration of an artistic style in any two-dimensional concept image. However, integrating an artistic style in scenography which includes a lot of three-dimensional elements is a much more complex task. To mitigate this issue, traditional neural style transfer methods for 2D images can be expanded to apply to a stereoscopic view. The intuitive step is the application of the existing style transfer methods to left and right views of stereoscopic images separately. However, the initial results are not balanced when transferred to the 3D perspective, causing fatigue to viewers. For this reason, a novel neural method proposed by (Chen et al., 2018) effectively extends NST to stereoscopic images and videos, providing new visual content for 3D movies and augmented and virtual reality. An important disadvantage is that the artistic rendering is not particularly successful when applied across two views, as the geometry is not consistent. The perception of depth, when the stylisation is performed, leads to eye discomfort. Therefore, the main goal is to adopt a fast solution to produce consistent stylisations across the two views. The architecture of the proposed model consists of two different networks. The first sub-network StyleNet (Gan et al., 2017) has been used extensively in similar works. Its architecture consists of several convolutional layers following the structure of an image auto-encoder. The second sub-network DispOccNet calculates the disparity maps for an input stereo image pair. The two networks are first trained on each task separately and then trained again as a whole. The model eventually produces symmetric left and right views. The emphasis is given to the overlapping regions in the centre of the stereoscopic images, visible from both views, which are stylised separately.

Summarising, Neural Style Transfer is a method that can efficiently learn the artistic style of a specific painting or mimic the style of a famous painter and transfer it to a selected image. The applications of NST and stereoscopic 3D are various in digital media: movies, games, and AR/VR displays, especially in the domain of content creation. As NST is a trending topic in content creation, it can similarly address the emerging need for 2D and 3D content in digital scenography. This process is a fast solution to concept image generation. It is also an easy tool for artists, designers, directors and producers to provide an extra layer of information in scenic design, by enriching simple set designs with widely recognised artistic styles.

2.3.2 Text-to-Image Techniques

Text-to-image techniques refer to the generation of visual representations based on textual descriptions or scripts. These techniques leverage advancements in Natural Language Processing to create visual elements for numerous applications. The early development of text-to-image systems primarily focused on visually representing nouns and a limited number of spatial prepositions, such as maps and charts (Zakraoui et al., 2019). Throughout the last decade, a number of text-to-image methods have been developed, however, the most efficient systems were only introduced in the last couple of years. Especially in the domain of theatre, a limited number of text-to-image systems have been utilised.

The work "Theatrical Text-to-3D Virtual Scenography" (Velonoromanalintantely et al., 2013) is a good example of a text-to-scene visualization system particularly tailored to the domain of the theatre. Text-to-scene systems are utilised for the creation of 3D environments based on descriptions extracted from text. Such a tool enables non-expert users to experiment with new methods in staging, designing, and production in theatre. The user feeds the system with descriptive elements, and a 3D view is automatically generated based on scenographic elements identified in the text. For the visualisation to take place, the theatrical text is first preprocessed with the use of annotations or tags, to identify semantic information. A knowledge base is then utilised to create constraints on the extracted information, which eventually leads to the generation of the virtual scene. Essentially, this system affords a user-friendly tool for the automatic generation of pictures or 3D scenes from simple text.

In more detail, this declarative system is organized into three main stages: Annotation, Interpretation, and Generation. During the first stage, the annotations or tags provide additional semantic information to a video or text. The descriptions are comprised of a set of properties such as spatial relations or lighting. These descriptions are later translated into geometric and photometric constraints in order to generate the 3D view of the theatrical scene. During the annotation process, the original text is annotated with two types of tags and converted to an XML file. The first set of tags (<TI>, <CH>, <DID>, <DIA>, <EM>) is used to identify the structure of the text: title, act, scene, names of the characters, and stage directions. The second set of tags is concerned with the descriptive elements, such as the orientation and position of the characters and the accessories, the posture, the expressions, and so on, (e.g.: "The chair is against the table" tagged: <REL target ="chair" relation="against" landmark="table"/>). The second stage is the Interpretation of the above properties. All the necessary information is stored in a Knowledge Base.

A Knowledge Base is a structure that stores information about semantic properties. In the specific system, the knowledge base contains all the data necessary for the interpretation stage: description of emotions, style of the play, the mood of the character, spatial relations for objects, etc. A querying process extracts the necessary information from the knowledge base and produces geometric data according to the user's needs. The interpretation stage generates constraints that lead to the definition of information about the objects and characters of the 3D model. Finally, the generation step is essentially the display of the resulting virtual scene using 3D software like Blender. This system proved to be helpful in creating different virtual environments, as the knowledge base can be extended for different contexts such as video games, simulators, urban development, etc.

After years of limited contributions of text-to-image techniques, the advancements in AI and Deep Learning in Natural Language Processing afforded a dramatic surge in the automatic creation of images based on text. More specifically, the inception of Generative Adversarial Networks in 2014 introduced text-guided synthesis of images using deep learning. In 2015 Google introduced Deep Dream ("Deep Dream Generator,"), a computer vision programme that leverages CNNs to modify images using an algorithm that creates dream-like, psychedelic effects. However, the popularity of image synthesis sparked just in 2021 with the release of CLIP by OpenAI ("CLIP"). CLIP is a model capable of classifying images when provided with their visual categories. It is trained on a very large dataset of images from the web and can associate visual concepts with their names. More specifically, after being fed with images and their labels, CLIP can predict the label that is most likely to be paired with an image. The ability to associate natural language with images is the revolution that made possible the development of generative systems (Oppenlaender, 2022).

Text-to-image generation systems are facilitated by the popularity of deep generative models. Deep generative models are able to synthesise digital images given an input prompt in natural language. Simply writing prompts in natural language is a process called prompt engineering, prompt programming, prompt design or just prompting (Oppenlaender, 2022).

Text-to-image techniques belong to the broader field of image synthesis. Image synthesis is one of the computer vision fields with the most dramatic development in recent years (Rombach et al., 2022). Image synthesis executed by practitioners is commonly known as "AI Art" or "generative art". The field of image synthesis leverages the power of neural networks and especially variations of Generative Adversarial Networks (GAN) (Goodfellow et al., 2020) and Variational Autoencoders (VAE) (Kingma and Welling, 2019).

Three of the most popular systems, Stable Diffusion, Midjourney and DALL-E are capable of generating artificial photorealistic images of high quality and are analysed below. The common attribute between those AI-enhanced models is scraping billions of pictures from the Internet and then creating concept images pulled from this immense repository of ideas. As neural networks, each of these tools learns from the available pictures and each time creates something totally new.

Midjourney

Midjourney ("Midjourney") is a currently open beta system, created by Midjourney lab. Midjourney was released in February 2022, yet the script became popular only in June 2022 through the Discord-based community. It is described as a generative software system that synthesises digital images from text descriptions. The generated images tend to have a more surreal character (Borji, 2022) leading Midjourney to be more popular among artists and designers.

The creation process is done in Discord as follows: after entering the prompts (/imagine command), Midjourney generates four results based on the input text. The user can then make the decision to select one of the produced images and ask Midjourney to upscale it or produce further variations of the selected image. It is pertinent to mention, that if no results are satisfactory, the user can repeat the generation task by providing the same keywords as the input text, as Midjourney will never generate the same results. This is due to Midjourney randomly selecting matching images from the input labels and providing different proportions and positions in the final image (Jaruga-Rozdolska, 2022). The images generated using Midjourney can be of variable size.

A good example that describes the workflow using Midjourney is McCann's work for a production of The Phantom of The Opera: The text prompt provided was "/imagine: a huge, malevolent pipe organ, gothic baroque style." McCann then selected one of the Midjourney-generated images, modified the aspect ratio and then developed it further into a matte painting style adding more scenic details (Barbour, 2023).

Figure 10: Concept art generation for The Phantom of the Opera, created by Garry McCann.

DALL-E 2

DALLE-E 2 is the most optimised version of DALL-E (2021), another successful model created by OpenAI ("DALL·E 2"). DALL-E takes simple text descriptions and turns them into photorealistic images that have never existed before. It involves various capabilities such as editing and retouching photos based on NLP descriptions or filling, replacing (inpainting) and extending (outpainting) specific parts of an image. DALLE-2 generates highly representative and complex images at higher resolutions with a greater comprehension of the input text. After generating initial images, it can also create variations of a selected image with different angles and styles. DALL-E 2 is trained on 650 million image-text pairs scraped from the internet. After learning a variety of labelled images, it can apply new attributes to each generation and amplify its creative potential. Images generated by DALL-E are always 512x512 pixels.

Stable Diffusion

Stable Diffusion, released by StabilityAI in 2022 (Rombach et al., 2022), is a generative model used to create highly detailed images based on text descriptions. While it is primarily used for the generation of images, it can also be modified to run other tasks such as inpainting, outpainting and image translation. It is trained on the LAION-5B database which includes images of size 512 x 512 and uses a text encoder called CLIP ViT-L/14, to condition the model on text prompts. Compared to other models, it is relatively light and can run on a GPU. Stable Diffusion is open source and provides more options for customisation, with basic knowledge of programming. The images generated using Stable Diffusion can be of variable size.

All of the above models are user-friendly and generate a variety of results. A discussion of the generated results is provided in the next chapter.

The creative process for concept art generation involves a variety of different methodologies, tools and approaches depending on the designer's preferences or desired outcomes. For example, a designer might first generate a rough sketch of the stage using Stable Diffusion. Then this sketch could be fed as input to a generative model trained to improve specific attributes of the input image. Then the inpainting and outpainting techniques can be performed, before editing the final output in a graphics tool. It is worth mentioning that this process might also have an iterative character. The user can provide a variety of textual prompts before concluding on the combination of words which produces satisfactory results. The interaction between the user and the AI responses in this iterative process can also steer the designer towards other unanticipated artistic directions.

A quantitative comparison between Midjourney, DALL-E 2 and Stable Diffusion concerning portrait generation, showcased that Stable Diffusion scores higher in creating photorealistic faces. The poor performance of Midjourney is due to the surreal and anime character of generated pictures. On the other hand, DALL-E 2 is trained using deepfake safeguards so to prevent it from memorising images available online. That means that the model is more efficient in generating portraits of imaginary people than portraits in more complex scenes (Borji, 2022). Iteration is an inherent attribute in image synthesis with generative models. For example, using VQGAN-CLIP, the user enters the prompt "ufo landing, fantastic, on Vellum, trending on /r/art". The image is initialised with random noise and gradually takes shape with each iteration. After a number of steps (iterations), the user has a good idea of the final output and has the option to abort the synthesis process and redefine the parameters that will give more specialised results (Oppenlaender, 2022).

Figure 11: Image synthesis process (Oppenlaender, 2022).

In summary, the recent text-to-image generative models are the most effective in generating concept images for theatre. Midjourney, DALL-E and Stable Diffusion are considered cutting-edge techniques with promising applications, especially in generating concept art for theatre productions. By leveraging the power of pre-trained models and image synthesis techniques, they generate complex and imaginative images based on textual prompts. This process provides theatre designers with different opportunities during their work as they can input simple text and quickly obtain concept art that captures the essence of their vision. These AI tools enrich the designers' toolkits and pave new paths in the design process.

Both text-to-image and image-to-image systems analysed in this chapter are used in various image processing applications. Digital Scenography is another field that these image synthesis systems can enhance. Artists can make use of text-to-image techniques to generate innovative ideas using simple textual descriptions, especially when faced with lack of inspiration. Subsequently, they can leverage image-to-image techniques to enrich an existing design with a selected artistic style. The whole conceptual design process facilitated by these tools is completed in just a few minutes.

3. Experiments

With the growing availability of text-to-image generation systems, everyone can experiment with generating digital images and "AI art". More specifically, a variety of text-to-image systems have emerged online and are mainly available as web applications, open-source notebooks (e.g., Google Colab) or on GitHub. Based on discussions with practitioners of the field, it is a very common practice to utilise a range of text-to-image models to create initial images, as well as other AI-powered tools to make edits to those images.

In this section, we present the designer's working methodology as discussed by contemporary practitioners, and more specifically the results obtained from text-to-image models utilised during the visualisation stage of the design process. Moreover, we experiment with the most commonly used tools for image generation, namely Midjourney, DALL-E and Stable Diffusion.

3.1 Artists' Working Methodology

The main question asked to artists working in theatre productions is "Are there tools to use AI content in the design process?" And the answer is yes. Digital media artists working with theatrical productions agree that AI-powered tools are now applicable and useful in their work, by providing solutions in terms of time and cost (Jamerson, J. personal communication 17/04/2023). Especially text-to-image models are extensively used in the visualisation process of the design replacing other slower digital media tools.

The initial part of a scenic design work methodology includes *reflection* on the theatrical play, as well as extensive *research* of Google Images, Pinterest, and other visual collections for the designers to get initial ideas regarding the theatrical play. AI tools can accelerate this process by quickly generating a number of images that fit the designers' description based on the theatrical text.

Following is the *conceptual design* stage. A good scenic design takes into account positive and negative space. Positive space is a term referring to the scenic components present on stage, for example, levels, props, furniture, ramps, stairs, entrances, exits, essentially anything that occupies space on the stage. The negative space is occupied by the directors and consequently the actors. The positive space sets the mood for the play but most importantly functions as a complementary space that provides the necessary tools to define and achieve the desired performance of the actors on stage (Jamerson, J. personal communication 17/04/2023). Until recently, defining a working relationship between the positive and negative space in the design process was mainly achieved using Photoshop or 3D modelling software. Traditionally, Photoshop would be used to create collages of a desired set design, including a combination of images of props, furniture and actors to visualise the three-dimensional space of the stage. This process is indeed helpful in determining the spatial characteristics of the positive space and also in providing a general idea about the aesthetics, the positions and the dimensions of the theatrical components, the colours, etc.

Figure 12: Scenic design rendering with Photoshop [Studio Work – Jason Jamerson Design]

However, in the past months, this design stage has been replaced by multi-steps of different AI tools: First, image generation takes place using text-to-image tools and then a variety of highly specialised AI tools are used for different types of customisation.

In more detail, the first step is achieved using text-to-image tools, as the initial and main content is generated by a simple text prompt by the designer. The text-to-image models facilitate the visualisation stage through fast content generation: with each text prompt, the models generate a variety of digital images that provide designers with ideas that fit their stylistic needs. Mastering prompt engineering can even lead to text-to-image models generating a variety of almost fully realised images, however, the human factor is absolutely necessary in order to be able to customise the final output. The images generated by Midjourney, DALL-E, and Stable Diffusion provide a general concept idea, yet the control of the generated results is limited. And so, supplementary AI tools can be used together with the aforementioned tools to gain control over the spatial characteristics and dynamically determine the content. Therefore, in order to control what the content is and how it is presented, the second step is to make use of other more specialised AI tools to affect and modify the generated content towards a desired artistic direction.

Figure 13: Concept art generated with Midjourney "The mysterious Fairyland, whose moon glimmers and dewdrops rest on the forested grasses." (Forsee, 2022).

A good example of a supplementary specialized AI tool commonly used among artists in the second stage is ControlNet. ("ControlNet," 2023) is a neural network model for controlling Stable Diffusion models. Some of the functions of ControlNet include turning a scribble into a photorealistic image, detecting and extracting human poses, and most importantly controlling the depth of the image so this looks like a real 3D space. By providing ControlNet with the information that the input picture is a photograph of a three-dimensional room, the AI system performs the so-called depth estimation. Depth estimation is the process where the AI tries to imagine the depth of the input image using depth maps and provides variations of the spatial position of components on stage. Black is used to define the features in the back, while the whiter the features, the closer they are positioned to the front. This way, the designer is provided with a tool to create different variations of the same design but with different customisation options such as colours, furniture, fabric and so on.

Figure 14: Depth map of a Stable Diffusion generated image ("ControlNet," 2023).

In summary, the designer's methodology consists of the stages: reflection on the play, research, conceptual design and implementation. Artists have enriched their design toolkits with AI-enhanced tools. Text-to-image models, in particular, are the ones most often used by designers and theatremakers in the conceptual design and visualisation stage. Text-to-image models (e.g., Midjourney, DALL-E, Stable Diffusion) take simple, natural language descriptions of the designer's idea, analyse and recognise the language, and finally create an image based on the processed text. Then, a variety of extremely finely tuned programs (e.g., ControlNet) can be utilised for customisation.

The outcome of this process is a highly representational scenic design that functions as a valuable guide in theatrical productions.

To validate this, we run experiments with the three most commonly used text-to-image models, namely, Midjourney, DALL-E, and Stable Diffusion. The pictures below represent prospective designs for “A Midsummer Night’s Dream” by William Shakespeare.

Midjourney was asked to generate the following:
“A long shot, studio photograph of a theatrical stage set for A Midsummer Night's Dream. On stage is the mysterious Fairyland whose moon glimmers and dewdrops rest on the forested grasses.” (Forsee, 2022)

Figure 15: “A long shot, studio photograph of a theatrical stage set for A Midsummer Night's Dream. On stage is the mysterious Fairyland whose moon glimmers and dewdrops rest on the forested grasses.” Generated with Midjourney (Forsee, 2022).

3.2 Midjourney

Midjourney is very powerful in generating initial concept art and has surpassed other digital media tools for concept design. Midjourney is currently in open beta with free initial generations and a paid subscription system. Midjourney, which seems to be preferred among artists was asked to generate the following: “A long shot, studio photograph of a theatrical stage set for A Midsummer Night's Dream." Midjourney first generates four different versions of the input text.

Figure 16: Initial generated versions.

By selecting one of the four generated images, Midjourney is asked to create variations:

Figure 17: Midjourney creates variations of the first generated image.

Leveraging the NST approach to stylise an input image using as a reference a famous painting, we ask Midjourney to generate the following: "A long shot, studio photograph of a theatrical stage set for A Midsummer Night's Dream inspired by van Gogh."

Figure 18: Stylised generated versions and variations and upscaling of the third generated image.

3.3 DALL-E

We repeat the above procedure using DALL-E. We ran the experiments through their website: https://labs.openai.com/

Figure 19: "A long shot, studio photograph of a theatrical stage set for A Midsummer Night's Dream", generated with DALL-E.

Figure 20: "A long shot, studio photograph of a theatrical stage set for A Midsummer Night's Dream inspired by van Gogh", generated with DALL-E.

Figure 21: Variations of the first generated image.

3.4 Stable Diffusion

We use the Colab notebook provided by Hugging Face to run Stable Diffusion: ("Stable Diffusion," 2022).

Figure 22: "A long shot, studio photograph of a theatrical stage set for A Midsummer Night's Dream", generated with Stable Diffusion.

Figure 23: "A long shot, studio photograph of a theatrical stage set for A Midsummer Night's Dream, inspired by van Gogh", generated with Stable Diffusion.

4. Discussion

4.1 The Advantages of using AI and Deep Learning in Digital Scenography

The most important argument of this paper is that AI tools, categorised into image-to-image and text-to-image techniques, enhance digital scenography as far as the design process is concerned.

After researching the different design methodologies, we concluded that scenic design consists of the following stages: reflection on the theatrical text, research around the theatrical text, generation of concept designs and development of the final design. AI tools prove to be very useful in the design process, especially in the concept generation stage. AI tools generate high-quality images that amplify the designer's creativity and imagination and accelerate the overall working methodology. The way AI and especially text-to-image models visualise ideas is more powerful than other traditional design tools in terms of time and quality. In this section, we elaborate on why AI tools are successful and useful in the design process.

First, AI tools function as a limitless source of information during the research stage of the design process, exactly like classic search engines. Designers are enabled to look at different historical periods and art through the visualisation of simple prompt texts, for example, "gothic baroque style" or "19th-century interior", instead of researching the same information on web browsers. This accelerates the overall process, as the designer can easily have access to and visualise historical and artistic information, without having to explore many different sources across the Internet.

A great advantage of AI tools is that they are very useful in the visualisation of complex ideas. Brainstorming is one of the first steps an artist takes during the design process and very often, it is hard to imagine and visualise innovative and engaging ideas. Moreover, it is time-consuming to efficiently render a concrete idea, integrating various aspects of the envisioned design using traditional means like sketching or digital drawing. AI tools, on the other hand, generate detailed high-quality visuals with a variety of structural components including materials, textures, props, lighting effects, or even actors on stage, all in just an initial design.

In addition, AI generates images or visual worlds that do not exist. A powerful attribute of AI is its capacity to combine different and unrelated concepts, that the human brain cannot easily conceive, in a meaningful way. This is particularly useful when the desired design for a production is intended to be dreamlike or surrealistic.

AI tools save time and effort for scenic designers. Initial concept art can be generated in just a few minutes. Even though many of the designs generated visualise some ideas that the artists would not consider, a meaningful foundation is provided fast with the potential to be extended and reshaped towards a desired artistic direction. This is particularly helpful when the concept art is generated for the purposes of a big production, where different creative teams need to collaborate towards a mutual vision. At the same time, the director of the production can support or reject the scenic designer's proposed conceptual design without delaying the overall design process, as the scenic designer can easily generate something new and devote their efforts to a detailed design only after the director's approval.

As mentioned previously, scenic design approaches slightly differ for theatrical and operatic works. The scenery in theatres is to a great extent dependent on the theatrical text, while in opera scenography can be much more open to interpretation as it heavily relies on the music and the composer. As such, stage design in opera includes concepts that surpass simple visuals. AI is efficient in providing dreamlike, bizarre, abstract or even contrasting aesthetic outcomes that match the feelings provoked by the music.

Finally, it is important to mention that while AI can create a variety of different general conceptual images, it has also a great capability in generating highly detailed and specific outputs when requested in such a way.

4.2 Challenges and limitations of using AI and Deep Learning in Digital Scenography

Using AI tools for design generation is considered controversial and many people even blame AI for taking away jobs. In reality, AI functions as a very powerful but supplementary tool, as the human factor is necessary for the practical realisation of artistic work. A literature review of relative work indicates that there is not adequate research concerning the application of the generative models discussed in the field of digital scenography. This is due to the experimental nature of these tools as well as the fact that they are applied for various artistic purposes. This is also justified by the fact that Midjourney, DALL-E and Stable Diffusion are all very recently developed tools.

The experiments conducted in the previous chapter using Midjourney, DALL-E and Stable Diffusion showcase some of the barriers of these AI models when applied to scenography. The first observation is that all models, when provided with the same text prompt, generate a variety of interesting results, some of them being more artistic while others being more photorealistic. In any case, these images cannot be directly applied to a set design without a number of modifications executed by the designers themselves.

Thus, the most important barrier is that no matter how efficient AI tools are in complex tasks, they cannot surpass humans in the overall design process. As far as conceptual design is concerned, AI wins in terms of speed when confronted with a human, but the completion of scenic design is not possible without the intervention of human artistic nature and intelligence.

Another very important factor to consider is that scenography concerns the three-dimensional rendering of ideas, which means that all the current AI tools which generate 2D images are not by themselves capable of completing scenic design tasks, rather they just provide a mutual ground for a director and designer to first communicate their vision and then implement it in the real three-dimensional space.

Even though image generation systems are powerful in creating concept art, the results depend on the designer's effective use of these systems. In text-to-image models that translates to providing effective prompts. Writing prompts that generate the desired results depends on the prompt engineering skills the user steadily acquires through experimentation with different prompt modifiers. Having an understanding of the model's training data set as well as the configuration parameters is helpful in getting more specialised results. In addition, apart from effective prompt design, producing high-fidelity images requires the user's knowledge of photography composition rules, for example, which aspect ratio to be used for a specific subject. In the above experiments they key words "long shot", "studio photograph" and "theatrical stage" are pivotal for the generation of images that can apply to stage design.

As far as image-to-image techniques are concerned, Neural Style Transfer is a widely researched topic with many published modified versions and applications for images. However faithfully these methods can mimic a painting, they cannot substitute the creative work of an artist, especially when the output is not abstract but strictly defined by the needs of a production. In other words, NST can indeed be used to generate images that match the artistic style of a painter and influence a set design, but it is impossible to create complete artwork without the creative skills of a real artist.

Considering all the above, we conclude that anyone using these AI tools can generate content in just a few minutes, yet the results are by no means finished artworks. AI products can be used as a starting point in the creative process that provides inspiration and visual stimuli to designers. "The skill comes in translating the ideas to three-dimensional spaces in a way that makes sense, dramaturgically", says Gary McCann, an Irish scenic and costume designer (Barbour, 2023).

4.3 Future developments and potential applications

One of the barriers of generative art, as mentioned in the previous section, is the gap between the generation of a concept image in two-dimensional space and its application in three-dimensional space, that is transferring the 2D content into a 3D scene. Since the unveiling of Neural Style Transfer, related work in the field has produced impressive results working with 2D images. One logical extension of this field is to move from 2D space (e.g., images) to 3D space (e.g., 3D models, 3D scenes). After almost a decade of experimentation and similar works in computer vision, 3D style transfer remains a relatively unexplored problem and therefore, we briefly refer to recent works in this direction, to gain a deeper understanding of what style transfer in a 3D setting would look like.

Han et al. proposed the first NST method for the automatic creation of novel 3D shapes using locally extracted styles, "3D Shape Creation by Style Transfer" (Han et al., 2015). The main contribution of this work is the redefinition of "content" and "style" for 3D shapes. The style transfer proposed is a method that creates a variety of different shapes by recombining existing or new styles with existing contents. A similar approach to 3D style transfer is provided in the paper "Analogy-Driven 3D Style Transfer" (Ma et al., 2014). This style transfer technique leverages the concept of analogies. Given a source S and a target T shape, the goal is to generate a 3D output model O with the target's structure while obtaining the style characteristics of an exemplar E shape. For example, the algorithm computes the analogy between two 3D objects, a source armchair and a target sofa, and then applies the specific analogy to a given exemplar armchair to create a new sofa. The 3D shapes created in this work are visually appealing and suitable for different graphic applications.

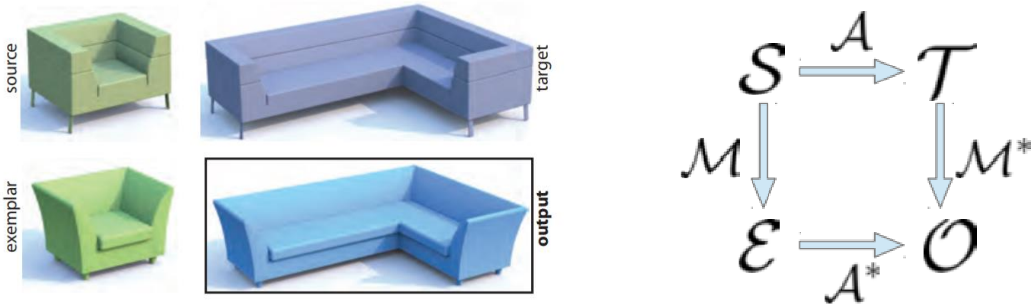

Figure 24: Analogy driven 3D style transfer (Han et al).

The application domains for 3D style transfer are diverse with the most prominent ones being augmented and virtual reality. Any 3D models produced can be easily embedded in virtual worlds, virtual gaming, and virtual collaboration environments. More importantly, scenic designers can leverage this process and generate countless new 3D scenic components by combining the content and style of existing previous 3D models.

In addition, a style transfer method proposed by Höllein et al. is concerned with 3D scene reconstructions. In contrast with the previous approaches that focused on explicit 3D models, this work introduces 3D environment rendering based on famous paintings. As with stereoscopic neural style transfer, the main challenge is the stylisation in the 3D perspective. When a style is transferred to a complex mesh, the artistic patterns become stretched out leading to an inconsistent result which is not appealing to the viewer. The approach followed in this work generates successful room-scale scene stylisations with equally sized and distributed rendering results without visual artefacts (Höllein et al., 2022).

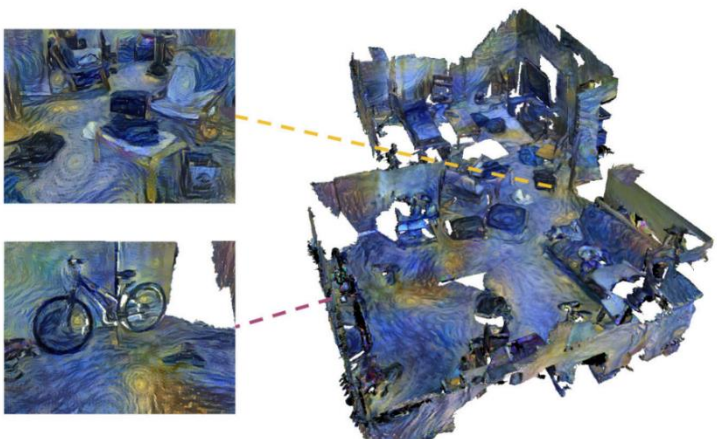

Figure 25: Neural style transfer in 3D-scene reconstruction (Höllein et al., 2022).

Applying an artistic style in a 3D scene provides designers with a more informative visualisation of a potential design as it is rendered in a way that closely matches realistic theatrical stages.

Finally, Text2Mesh (Michel et al., 2022) is effectively a combination of NST in 3D as well as text-to-3D, which is afforded by recent developments in joint embeddings of text and images with CLIP. This method facilitates the editing of the style of 3D objects according to a target text prompt (text-driven stylization). Essentially, the desired style is expressed through a text description, very similar to how an artist or a designer provides a verbal or textual description of the desired aesthetic outcome. It is important to mention that as far as text-to-image systems are concerned, there is always room for improvement when bridging the comprehension gap between the user and AI systems. With the advancement of subfields of natural language processing, generative models will become more efficient in generating images closer to the user's textual descriptions.

All in all, the next steps to enhancing the conceptual designs for scenography include leveraging methods that: increase the model's capability in interpreting the user's textual prompts and extend 2D renderings to the visualisation of 3D scenes.

Since digital scenography and its applications are still in the experimental stage there is not enough data to show a direct correlation between the use of digital technology and increased box office sales. Thus, the impact of digital technology in theatres and operas is unclear and it is important to be investigated in the future.

5. Conclusion

This research paper provided an overview of the creative possibilities of AI systems and their potential applications in scenic design. Considering that the most significant characteristic of digital technologies is interactivity, digital scenography is one of the most effective interdisciplinary fields to practise human-machine interaction. This interaction is present in both the performance and the design process. While this paper touched on different aspects of interactions in performances, it mostly focused on the interaction of designers and computational systems during the design process, something which got very recently afforded by the advancements of deep learning and artificial intelligence.

It is claimed that technological advancements conventionalise other traditional artistic practices. However, they also provide another perspective on traditional practices or unlock different types of creativity. An interesting parallelism is painters being threatened by the invention of photography. While some artists were indeed negatively affected, photography opened up more opportunities for artists willing to embrace its creative possibilities (Oppenlaender, 2022). Similarly, AI has posed the problem of what is considered art and what is not and a lot of professionals express doubt, disbelief, and worry when technology intervenes in artistic aspects. In the field of stage design, artists are enthusiastic and embrace technology by supporting that AI tools do not substitute their work, but rather empower it by providing fast and appealing concept art:

"AI design tools have helped me conceptualize faster, communicate with stronger visuals earlier, and empower collaborative conversations with real-time iterations. AI helps me share my ideas and get to my best work more efficiently." (Jamerson, J. personal communication, 17/05/23)

"Most creatives have a mindset of scarcity, that AI will take what few opportunities are already available for designers. For my part, I've always found the more unique and

diverse our ways of working and toolkits - the more impactful our creations become. I have to believe in abundance, that AI will only expand the possibilities of our craft. We as designers are responsible for finding that path forward." (Forsee, D. personal communication, 17/05/2023)

All in all, it is unquestionable that AI has accelerated scenic design workflows and has empowered designer's work by providing tools that ignite their imagination and creativity. The fusion of AI in other facets of digital scenography will serve as the catalyst for the creation of innovative and captivating experiences for audiences in the days to come.

6. Bibliography

Aronson, A. (Ed.), 2017. The Routledge Companion to Scenography. Routledge, London. https://doi.org/10.4324/9781317422266

Aronson, A., 2005. Looking Into the Abyss: Essays on Scenography. University of Michigan Press.

Barbour, D., 2023. Lighting and Sound America [WWW Document]. URL https://edition.pagesuite-professional.co.uk/html5/reader/production/default.aspx?pubname=&edid=1c6d00ec-4996-42ea-86cd-134c02b6f16a&pnum=2 (accessed 5.12.23).

Baugh, C., 2013. Theatre, Performance and Technology: The Development and Transformation of Scenography. Bloomsbury Academic.

Borji, A., 2022. Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2. https://doi.org/10.48550/arXiv.2210.00586

Chen, D., Yuan, L., Liao, J., Yu, N., Hua, G., 2018. Stereoscopic Neural Style Transfer. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6654–6663.

Chollet, F., 2017. Deep Learning with Python. Manning Publications Company.

CLIP: Connecting text and images [WWW Document], n.d. URL https://openai.com/research/clip (accessed 5.12.23).

ControlNet: A complete guide - Stable Diffusion Art, 2023. URL https://stable-diffusion-art.com/controlnet/ (accessed 5.16.23).

DALL·E 2 [WWW Document], n.d. URL https://openai.com/product/dall-e-2 (accessed 5.12.23).

Deep Dream Generator [WWW Document], n.d. URL https://deepdreamgenerator.com/about (accessed 5.12.23).

Diffusers.ipynb [WWW Document], n.d. . Hugging Face. URL https://colab.research.google.com/github/huggingface/notebooks/blob/main/diffusers/diffusers_intro.ipynb#scrollTo=xkyOEnzuVbsq (accessed 5.15.23).

Dills, C., 2018. Read, render, realize [WWW Document]. Dram. Mag. Online. URL https://dramatics.org/read-render-realize/ (accessed 5.16.23).

Dixon, S., 2007. Digital Performance: A History of New Media in Theater, Dance, Performance Art, and Installation. https://doi.org/10.7551/mitpress/2429.001.0001

Efros, A.A., Freeman, W.T., 2001. Image quilting for texture synthesis and transfer, in: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '01. Association for Computing Machinery, New York, NY, USA, pp. 341–346. https://doi.org/10.1145/383259.383296

Forsee, 2022. Practical Artificial Intelligence for Stage Design [WWW Document]. HowlRound Theatre Commons. URL https://howlround.com/practical-artificial-intelligence-stage-design (accessed 5.16.23).

Forsee, D., 2023. AI tools in scenic design work.

Gan, C., Gan, Z., He, X., Gao, J., Deng, L., 2017. StyleNet: Generating Attractive Visual Captions With Styles. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3137–3146.

Gatys, L., Ecker, A., Bethge, M., 2016. A Neural Algorithm of Artistic Style. J. Vis. 16, 326. https://doi.org/10.1167/16.12.326

Gatys, L.A., Ecker, A.S., Bethge, M., 2016. Image Style Transfer Using Convolutional Neural Networks. Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2414–2423.

Gatys, L.A., Ecker, A.S., Bethge, M., Hertzmann, A., Shechtman, E., 2017. Controlling Perceptual Factors in Neural Style Transfer. https://doi.org/10.48550/arXiv.1611.07865

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2020. Generative adversarial networks. Commun. ACM 63, 139–144. https://doi.org/10.1145/3422622

Han, Z., Liu, Z., Han, J., Bu, S., 2015. 3D shape creation by style transfer. Vis. Comput. 31, 1147–1161. https://doi.org/10.1007/s00371-014-0999-1

Höllein, L., Johnson, J., Nießner, M., 2022. StyleMesh: Style Transfer for Indoor 3D Scene Reconstructions. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6198–6208.

Howard, P., 2001. What is Scenography? Routledge, London. https://doi.org/10.4324/9780203424520

Jamerson, J., 2023a. AI tools in scenic design.

Jamerson, J., 2023b. AI tools in scenic design work.

Jaruga-Rozdolska, A., 2022. Artificial intelligence as part of future practices in the architect’s work: MidJourney generative tool as part of a process of creating an architectural form 95–104. https://doi.org/10.37190/arc220310

Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M., 2020. Neural Style Transfer: A Review. IEEE Trans. Vis. Comput. Graph. 26, 3365–3385. https://doi.org/10.1109/TVCG.2019.2921336

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual Losses for Real-Time Style Transfer and Super-Resolution, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 694–711. https://doi.org/10.1007/978-3-319-46475-6_43

Karathanasi, V., 2022. Creation of artificial artworks using specific styles with convolutional neural networks. University of Patras, Patras, Greece.

Kingma, D.P., Welling, M., 2019. An Introduction to Variational Autoencoders. Found. Trends® Mach. Learn. 12, 307–392. https://doi.org/10.1561/2200000056

Kyprianidis, J.E., Collomosse, J., Wang, T., Isenberg, T., 2013. State of the "Art”: A Taxonomy of Artistic Stylization Techniques for Images and Video. IEEE Trans. Vis. Comput. Graph. 19, 866–885. https://doi.org/10.1109/TVCG.2012.160

Lotker, S., Gough, R., 2013. On Scenography: Editorial. Perform. Res. 18, 3–6. https://doi.org/10.1080/13528165.2013.818306

Ma, C., Huang, H., Sheffer, A., Kalogerakis, E., Wang, R., 2014. Analogy-driven 3D style transfer. Comput. Graph. Forum 33, 175–184. https://doi.org/10.1111/cgf.12307

Mayda, M., Choi, S.-K., 2017. A reliability-based design framework for early stages of design process. J. Braz. Soc. Mech. Sci. Eng. 39, 2105–2120. https://doi.org/10.1007/s40430-017-0731-y

McKinney, J., Butterworth, P., 2009. The Cambridge Introduction to Scenography, Cambridge Introductions to Literature. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511816963

Michel, O., Bar-On, R., Liu, R., Benaim, S., Hanocka, R., 2022. Text2Mesh: Text-Driven Neural Stylization for Meshes. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13492–13502.

Midjourney [WWW Document], n.d. . Midjourney. URL https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F (accessed 5.12.23).

O'Dwyer, N., 2021. Digital Scenography: 30 Years of Experimentation and Innovation in Performance and Interactive Media. https://doi.org/10.5040/9781350107342

Oppenlaender, J., 2022. The Creativity of Text-to-Image Generation, in: Proceedings of the 25th International Academic Mindtrek Conference, Academic Mindtrek '22. Association for Computing Machinery, New York, NY, USA, pp. 192–202. https://doi.org/10.1145/3569219.3569352

Palmer, S., 2017. Light and Projection, in: The Routledge Companion to Scenography. Routledge.

Parker-Starbuck, J., 2011. Cyborg Theatre:: Corporeal/Technological Intersections in Multimedia Performance. Palgrave Macmillan.

Roça, C., Augusto, C.A., Rebelo, S.M., Martins, P., 2022. Real-Time Dynamic Digital Scenography: An Electronic Opera as a Use Case, in: Wölfel, M., Bernhardt, J., Thiel, S. (Eds.), ArtsIT, Interactivity and Game Creation, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer International Publishing, Cham, pp. 155–167. https://doi.org/10.1007/978-3-030-95531-1_11

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-Resolution Image Synthesis With Latent Diffusion Models. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695.

Saltz, D.Z., 2013. Media, Technology, and Performance. Theatre J. 65, 421–432.

Stable Diffusion [WWW Document], 2022. URL https://colab.research.google.com/github/huggingface/notebooks/blob/main/diffusers/stable_diffusion.ipynb (accessed 5.16.23).

Strothotte, T., Schlechtweg, S., 2002. Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation, Illustrated edition. ed. Morgan Kaufmann, San Francisco, CA.

Thornett, L., 2017. Scenography expanded: an introduction to contemporary performance design, edited by Joslin McKinney and Scott Palmer. Theatre Perform. Des. 3, 191–193. https://doi.org/10.1080/23322551.2017.1395248

Ulyanov, D., Lebedev, V., Vedaldi, A., Lempitsky, V., 2016. Texture Networks: Feed-forward Synthesis of Textures and Stylized Images. https://doi.org/10.48550/arXiv.1603.03417

Velonoromanalintantely, R., Andriamarozakaina, T., Sanza, C., Gaildrat, V., Martinez-Thomas, M., Matthieu, P., 2013. Theatrical Text to 3D Virtual Scenography, in: Plemenos, D., Miaoulis, G. (Eds.), Intelligent Computer Graphics 2012, Studies in Computational Intelligence. Springer, Berlin, Heidelberg, pp. 23–40. https://doi.org/10.1007/978-3-642-31745-3_2

Vincent, C., 2021. Digital Scenography in Opera in the Twenty-First Century. Routledge, London. https://doi.org/10.4324/9781003093305

Vincent, J., Vincent, C., Vincs, K., Mccormick, J., 2016. Navigating control and illusion: functional interactivity versus 'faux-interactivity' in transmedia dance performance. Int. J. Perform. Arts Digit. Media 12, 44–60. https://doi.org/10.1080/14794713.2016.1161955

What is Artificial Intelligence (AI) ? | IBM [WWW Document], n.d. URL https://www.ibm.com/topics/artificial-intelligence (accessed 5.11.23).

Won, P.-H., 2001. The comparison between visual thinking using computer and conventional media in the concept generation stages of design. Autom. Constr., CAADRIA 10, 319–325. https://doi.org/10.1016/S0926-5805(00)00048-0

Zakraoui, J., Saleh, M., Al Ja'am, J., 2019. Text-to-picture tools, systems, and approaches: a survey. Multimed. Tools Appl. 78, 22833–22859. https://doi.org/10.1007/s11042-019-7541-4