Style Transfer
Bridging Art, Culture and Technology

Leah Kawka leah.kawka@campus.lmu.de
Report for Seminar Creating Art(efacts) WiSe 24/25
Department of CVML @ LMU Munich

Research Question

In the ever-evolving dialogue between code and artistic creation, what role does the technology Style Transfer hold in contemporary Culture and Art?

Contents

Abstract

Style transfer is a transformative technique in computer vision that merges artistic styles with digital image content, bridging the gap between art, culture, and technology. This report provides a comprehensive examination of style transfer, starting with its conceptual foundations and historical intersections with art. It defines key style features and explores the evolution of techniques, from early neural style transfer to more advanced GAN-based and diffusion-based approaches. Further, the study investigates applications of style transfer across various domains, including open-source model platforms, browser-based applications, and commercial implementations. It also examines the societal impact and ethical challenges associated with this technology. A particular focus is given to visual artifacts as stylistic elements in contemporary art, analyzing their role in cultural reinterpretation and their influence on identifying a new style in artworks. Finally, the report discusses future directions, highlighting emerging trends in style transfer research, and the broader implications of computer vision-driven artistic generation. The conclusions underline the increasing relevance of style transfer as a tool for creative expression and cultural preservation.

1 Introduction

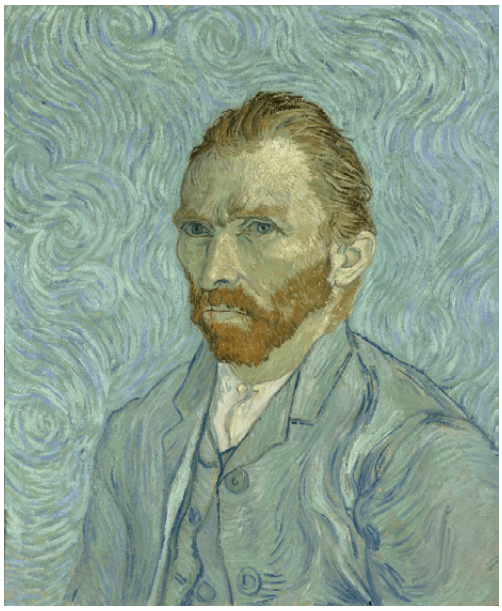
Style Transfer is a well-established technique in computer vision and pattern recognition that enables the generation of artistic images by transferring the visual style of one image onto the content structure of another. This approach has been widely explored in academia and industry for decades [1]–[3], leading to significant advancements in image synthesis, artistic rendering, and creative applications. By blending the content features of one image source with the stylistic attributes of another, style transfer allows for the transformation of images into visually unique, artistically inspired compositions. As research in this field progresses, style transfer continues to evolve, expanding its applications beyond 2D images into 3D graphics, video processing, and real-time rendering technologies.

1.1 Intersection of Style Transfer and Art History

The stylization of artworks has always been a fundamental practice in artistic creation [4]. With the advent of digital images, image editing has gained significance as an important artistic tool. By applying filters and adjusting parametric image properties, digital representations have been modified to shape novel artistic styles or reference well-known artworks. [5] With the increasing interest in machine learning research for the automated generation of stylistically adapted images, the development of trained models and the growing demand for large datasets have followed. These technologies have simultaneously become an integral part of artistic creation, making a comprehensive conceptual clarification of the term *style transfer* particularly relevant from both an art historical and computer science perspective. The definition of art style has been a recurring topic in Western art history and philosophy for centuries [6], [7]. Art epochs and art movements reflect on societal and cultural values, religion, philosophy, and social changes. As a concept of scientific research, it allows conclusions to be drawn about an artwork's author, time, place, and various interrelated influences. In contrast, the term *style* in the context of style transfer as a computer vision method is more specifically defined.

1.2 Defining Style Features

Due to the technical demands of computer science, early style transfer research primarily focused on the recognizable characteristics of artistic styles. That included textural features such as the distinctive brushstrokes of Van Gogh (a) or the woodblock print aesthetics of Ukiyo-e (b). [Fig. 1]



(a) *Self Portrait* by Vincent van Gogh, 1889 [8]

(b) *Under the Wave off Kanagawa* by Katsushika Hokusai, ca. 1826-1836 [9]

Fig. 1

The following examples highlight the challenge of categorization from the perspective of different scientific disciplines. Since generative systems are trained on taxonomical or ontological training data, a taxonomical classification of portraits by two extensively researched artists, Frida Kahlo and Pablo Picasso, should help clarify the ambiguities.

Pablo Picasso's globally renowned self-portraits, such as those from the Blue Period (e.g. in 1901) and another in the Cubist style (e.g. in 1907), can be easily categorized. This is mainly because the Blue Period is uniquely associated with Picasso, and he is widely recognized as one of the founders of Cubism.

However, towards the end of his life, Picasso adopted complete artistic freedom in his style. In 1972, he produced two very different artworks in style, *Self-Portrait Facing Death* and *The Young Painter*. The latter bears a striking resemblance to a photographic portrait of Picasso from 1904 [Fig. 2], raising the question of whether *The Young Painter* is indeed a self-portrait or a depiction of another individual.

Fig. 2: Pablo Picasso photographed by Ricard Canals i Llambí, 1904 [10]

Frida Kahlo's self-portrait paintings often contained strong autobiographical elements, blending realism with fantasy. In addition to being associated with the post-revolutionary Mexicayotl movement, which sought to define a distinctly Mexican identity, Kahlo has been described as both a surrealist and a magical realist. To this day, blog authors and online encyclopedia contributors sporadically refer to her as a surrealist artist. [Fig. 3] Time magazine quoted her in 1953, declaring:

> *They thought I was a Surrealist, but I wasn't. I never painted dreams. I painted my own reality.* [11]

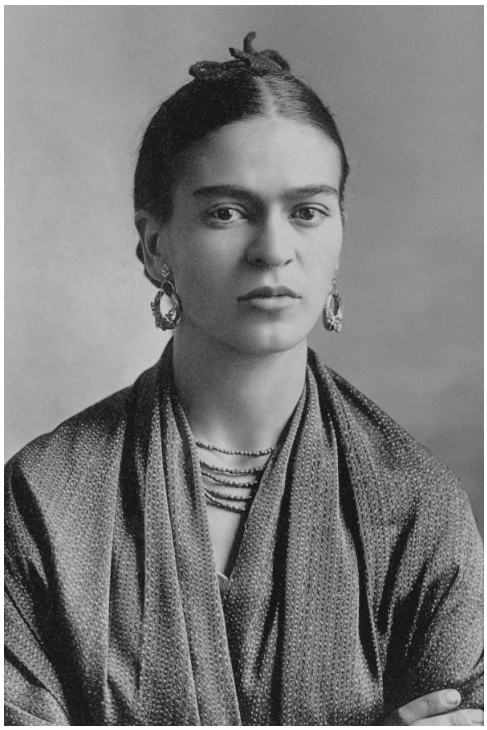


Fig. 3: Frida Kahlo photographed by Guillermo Kahlo, 1932 [12]

These ambiguities highlight key challenges in designing data systems. Beyond technique, time of creation, or artistic style, even the interpretation of an image's subject can vary significantly. Such distinctions are particularly relevant for the development of ontologies and taxonomies, the clarification of authorship in digital humanities, and the categorization of training data.

Gatys et al. [13] identified the problem as follows:

> *[. . .] style of an image [. . .] might be the brush strokes in a painting, the colour map, certain dominant forms and shapes, but also the composition of a scene and the choice of the subject of the image – and probably it is a mixture of all of them and many more.*

Defining Style is inherently challenging, making it a non-trivial problem.

2 Foundations of Style Transfer

Style transfer in computer vision can be understood as a texture transfer problem, where the goal is to generate a texture from a source image while preserving the semantic content of the target image. Various nonparametric algorithms achieve realistic texture synthesis by resampling pixels from a source texture [1], [14]–[16]. Innovative texture transfer methods utilized nonparametric approaches, employing different strategies to maintain the structural integrity of the target image. Efros and Freeman introduced a correspondence map incorporating image intensity features to guide the texture synthesis [1]. Similarly, Hertzmann et al. used image analogies to transfer texture from an already stylized image onto a target image [3]. Ashikhmin's method focuses on transferring high-frequency texture details while preserving the overall coarse structure of the target image [17]. Lee et al. improved this technique by further integrating edge orientation information to refine the texture transfer process [18]. Despite their effectiveness, all these algorithms share a fundamental limitation. They rely exclusively on low-level image

features, such as edges, colors, and textures of the target image. This limits their ability to capture intricate artistic styles. Style Transfer has advanced significantly with a key idea [13].

> *[. . .] a style transfer algorithm should be able to extract the semantic image content from the target image (e.g. the objects and the general scenery) and then inform a texture transfer procedure to render the semantic content of the target image in the style of the source image.*

3 Evolution of Style Transfer Techniques

This chapter outlines key milestones in Style Transfer methods in chronological sequence.

3.1 Neural Style Transfer

Neural Style Transfer (NST) is a computational technique that blends one image's content with another's artistic style using deep neural networks. It extracts structural elements from the content image while capturing stylistic features such as colors, textures, and patterns from the style image. NST methods encode style features using a single Gram matrix to capture statistical correlations and create a new image that retains the original content but appears rendered in a different artistic style. The method was first introduced in the 2015 paper ***A Neural Algorithm of Artistic Style*** by Leon Gatys et al., which was later accepted at CVPR in 2016 under the name:

Image Style Transfer Using Convolutional Neural Networks [13]

This fundamental breakthrough in style transfer technology presents a neural algorithm for artistic style transfer that allows for the separation and recombination of content and style from different images. This is achieved using Convolutional Neural Networks (CNNs) [19], which extract high-level image representations by leveraging features optimized for object recognition. The algorithm effectively renders the semantic content of an arbitrary photograph in the visual style of various well-known artworks, producing high-quality perceptual images by optimizing the alignment of content and style representations within the network.

Fig. 4: Illustration of accuracy

This illustration demonstrates the method's accuracy. These images blend the content of a photograph with the artistic styles of several renowned paintings. They were generated by optimizing an image to simultaneously align with the photograph's content representation and the artwork's style representation. The original photograph depicting the Neckarfront in Tübingen, Germany, is shown in A (Photo: Andreas Praefcke) [Fig. 4].

The painting serving as the style reference for each generated image is displayed in the bottom left corner of its respective panel. B *The Shipwreck of the Minotaur* by J.M.W. Turner, 1805. C *The Starry Night* by Vincent van Gogh, 1889. D *Der Schrei* by Edvard Munch, 1893. E *Femme nue assise* by Pablo Picasso, 1910. F *Composition VII* by Wassily Kandinsky, 1913.

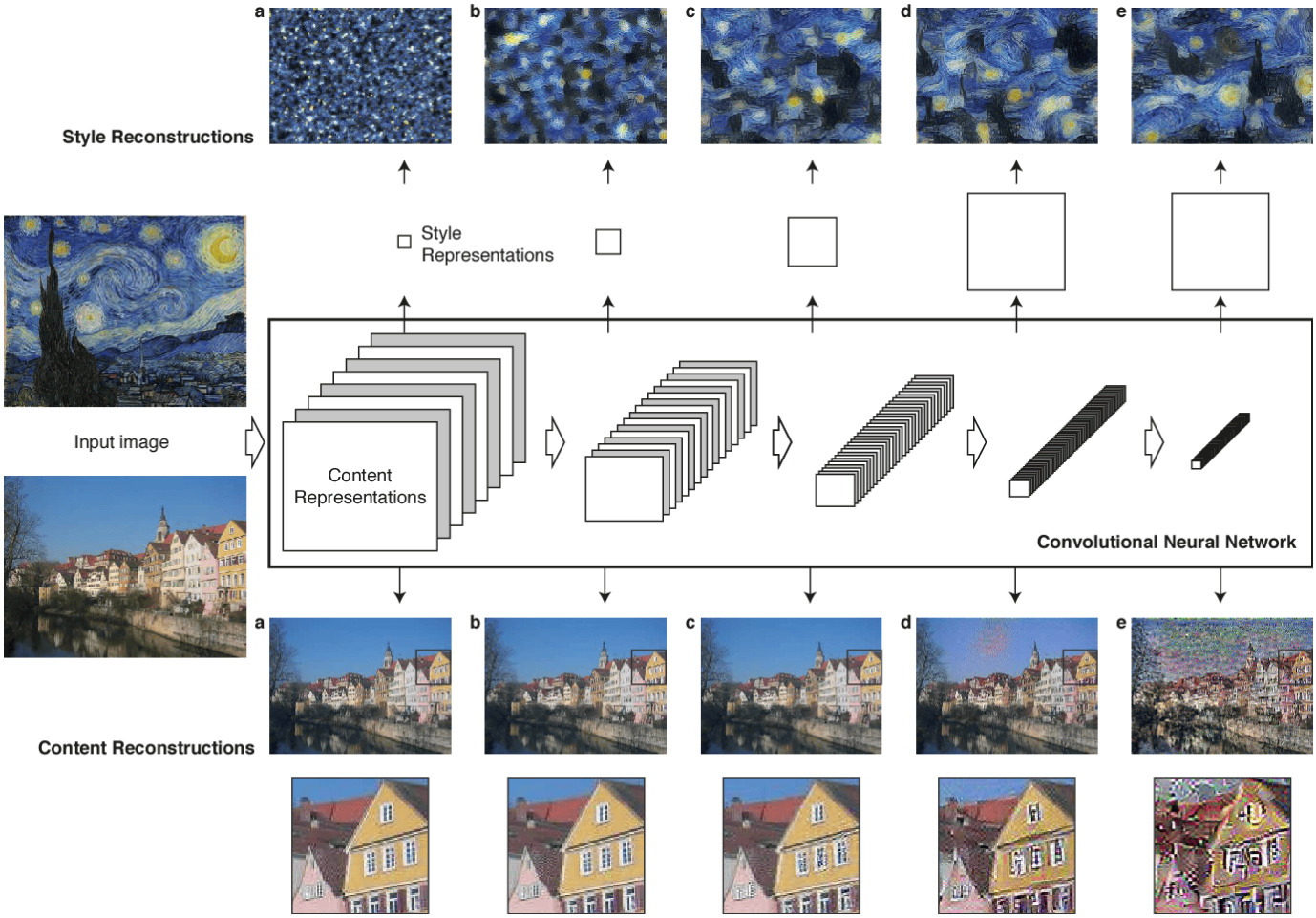


Fig. 5: Illustration of image representations in a CNN

This figure illustrates image representations in a CNN, where an input image undergoes filtering at each stage. The authors use a pre-trained VGG network [20], with normalized weights ensuring each convolutional filter's mean activation remains one across images and positions. As the processing hierarchy deepens, the number of filters increases, while down-sampling (e.g., max-pooling) reduces image size and the total units per layer. To visualize the content representation, the information at different processing stages in the CNN is reconstructed by generating the input image using only the network's responses from a specific layer. Reconstructions from lower layers (a–c) are nearly identical to the input, while higher layer reconstructions (d, e) lose fine pixel details but retain the overall high-level content of the image. Higher layers of the CNN capture the high-level content of the image, such as objects and their arrangement. The content representation is obtained by performing gradient descent on a white noise image to match the feature responses of the original image in a particular layer [Fig. 5].

Building on the original CNN activations, a feature space is utilized to capture the texture characteristics of an input image. The style representation is derived by computing correlations between features across multiple CNN layers. These feature correlations are represented by the Gram matrix, which encodes the inner product between

vectorized feature maps in a layer. The style loss is calculated as the mean-squared distance between the Gram matrices of the original image and those of the generated image.

A progressively expanding subset of CNN layers is used to reconstruct the style of an input image. As more layers are incorporated, the generated images increasingly reflect the texture of the reference style, while gradually discarding information about the scene's global spatial arrangement. The style of one image is transferred onto another by synthesizing a new image that matches the content representation of one image and the style representation of the other. This is achieved by jointly minimizing the distance of the feature representations of a white noise image from the content representation of the content image in one higher layer, and the style representation of the style image defined on every layer of the CNN. The total loss function is a linear combination of content and style loss.

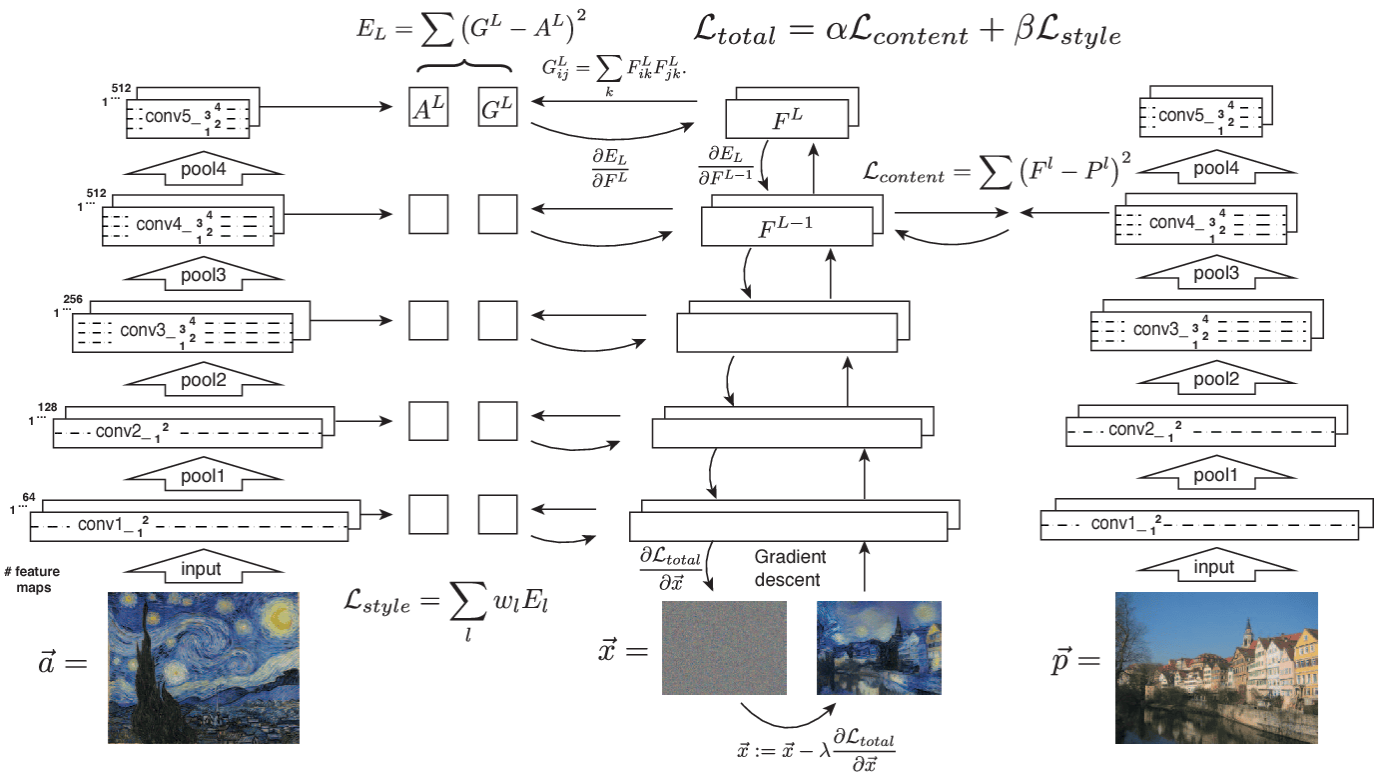


Fig. 6: Style transfer process

The style transfer process begins with extracting and storing content and style features. The style image ($\vec{a}$) is passed through a neural network, where its style representation (A_l) is computed across multiple layers (left). Similarly, the content image ($\vec{p}$) is processed, and its content representation (P_l) is stored in a specific layer (right).

Next, a random white noise image ($\vec{x}$) is introduced into the network, where its style features G_l and content features F_l are computed.

The style loss (L_{style}) is determined by calculating the element-wise mean squared difference between G_l (style features of $\vec{x}$) and A_l (style features of $\vec{a}$) across all layers (left). Similarly, the content loss (L_{content}) is derived from the mean squared difference between F_l (content features of $\vec{x}$) and P_l (content features of $\vec{p}$) in the selected layer (right).

The total loss (L_{total}) is then computed as a weighted sum of the content and style losses. Using error back-propagation, the gradient of this loss with respect to the pixel values is calculated (middle). This gradient is then iteratively applied to update the image $\vec{x}$, optimizing it to match the style features of $\vec{a}$ and the content features of $\vec{p}$ (middle, bottom). By independently manipulating content and style representations, the study highlights the potential for image synthesis and artistic manipulation. Therefore, the authors further examine the impact of content and style matching, as well as the effect of different CNN layers on the final visual outcome [Fig. 6].

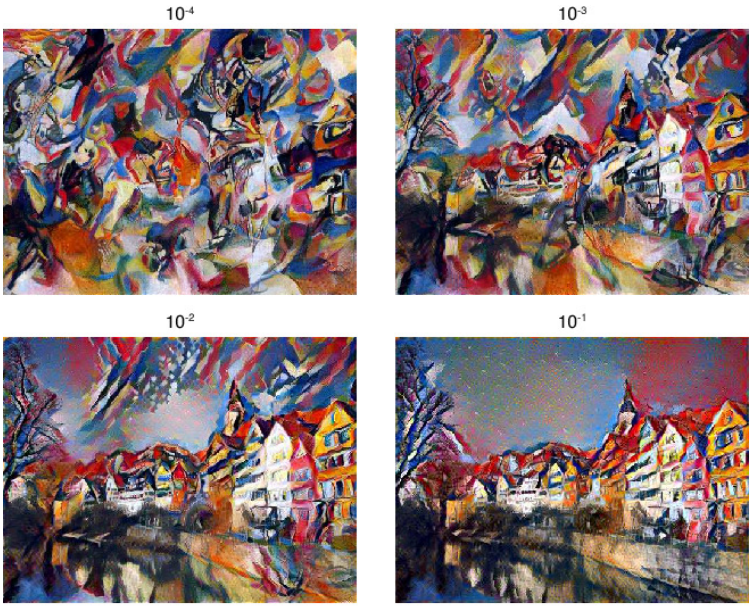


Fig. 7: Content and style matching

The balance between content and style matching in the generated images is determined by the relative weighting of the respective source images. The ratio α/β, which controls the emphasis on content versus style, increases from top left to bottom right. When the style weight is dominant, the result closely resembles a texturized version of the style image (top left). Conversely, prioritizing content results in an image with minimal stylization, preserving the original structure (bottom right). In practice, a smooth interpolation between these two extremes allows for flexible adjustments in the final output [Fig. 7].

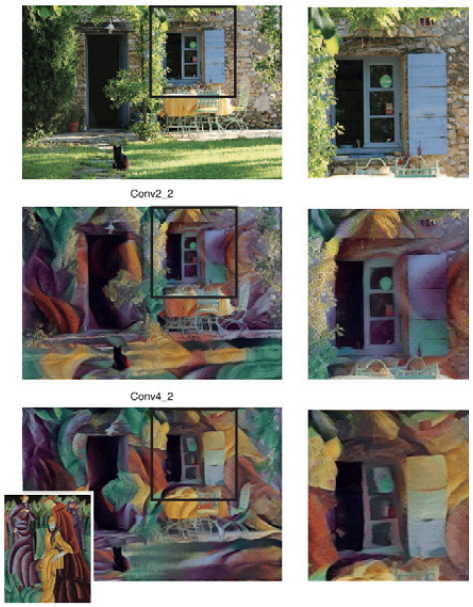


Fig. 8: Matching content representations at different network layers

In Figure 8 the impact of matching content representations at different network layers is illustrated. When content is matched at layer `conv2_2`, the fine structure of the original photograph is primarily preserved, resulting in a synthesized image where the painting's texture appears to be blended over the photograph (middle). In contrast, matching content at layer `conv4_2` leads to a more potent fusion of the painting's texture and content, effectively displaying the photograph in the artistic style of the painting (bottom).

Both images were generated using the same parameter settings $\alpha/\beta = 1 \times 10^3$. The painting used as the style reference, *Jesuiten III* by Lyonel Feininger (1915), is shown in the bottom left corner. [Fig. 8]

One of the key limitations of the style transfer algorithm is the resolution of the generated images. As the number of pixels increases, the dimensionality of the optimization problem and the number of units in the CNN grow linearly. Consequently, the synthesis speed is highly dependent on image resolution. Another challenge is the presence of low-level noise in the generated images. While this is less problematic in artistic style transfer, it becomes more noticeable when the content and style images are photographs, compromising the output's photorealism. The noise often resembles the patterns of CNN's filters, so the authors suggest that post-processing denoising techniques could improve image quality after optimization.

Fig. 9: Challenges of photorealistic style

Additionally, the algorithm has limitations in generative flexibility. It lacks control mechanisms beyond the predefined optimization process, meaning it does not actively guide or influence generation beyond style transfer constraints. Furthermore, the method is restricted to applying a single style per execution, limiting the ability to blend multiple styles seamlessly within a single output [Fig. 9].

3.2 GAN-Based Style Transfer

Generative Adversarial Networks (GANs) [21] combine CNNs, which serve as foundational building blocks in its architecture, to overcome shortcomings of single CNNs. In a GAN-based model, a generator network learns to transform images from domain A to domain B. In contrast, a discriminator network ensures the output looks authentic in the target style. Unlike NST, which often uses a fixed pre-trained network for losses, GANs learn the style mapping through adversarial training. Style transfer via GANs typically requires a training dataset for the desired style. Once trained, though, the model can produce stylized results in one forward pass, like fast NST models.

ArtGAN: Artwork Synthesis with Conditional Categorical GANs, 2017 [22] The ArtGAN article introduces a GAN architecture designed for generating complex images, particularly artworks, to address the challenge of synthesizing images with abstract characteristics [Fig. 10].

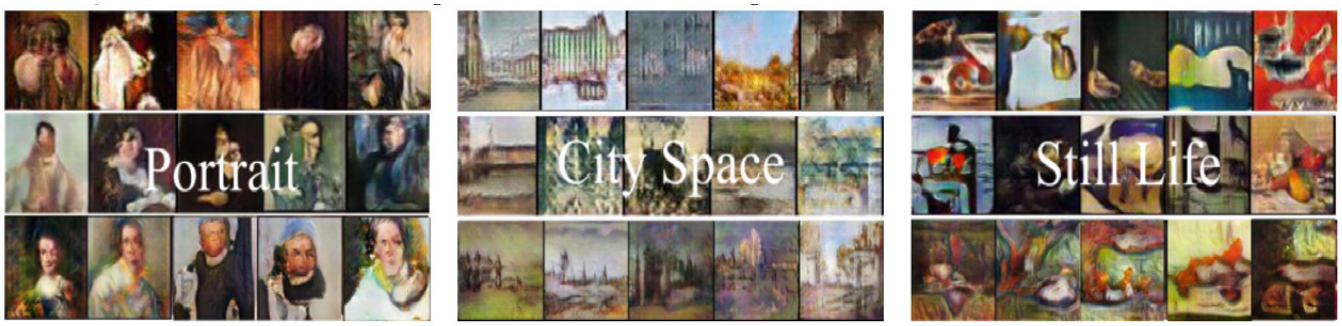


Fig. 10: Artwork Generation: Comparison between DCGAN (top), GAN/VAE (middle), and ArtGAN (bottom)

A key innovation is the back-propagation of the loss function related to additional image labels from the discriminator to the generator, allowing faster learning and improved image quality.

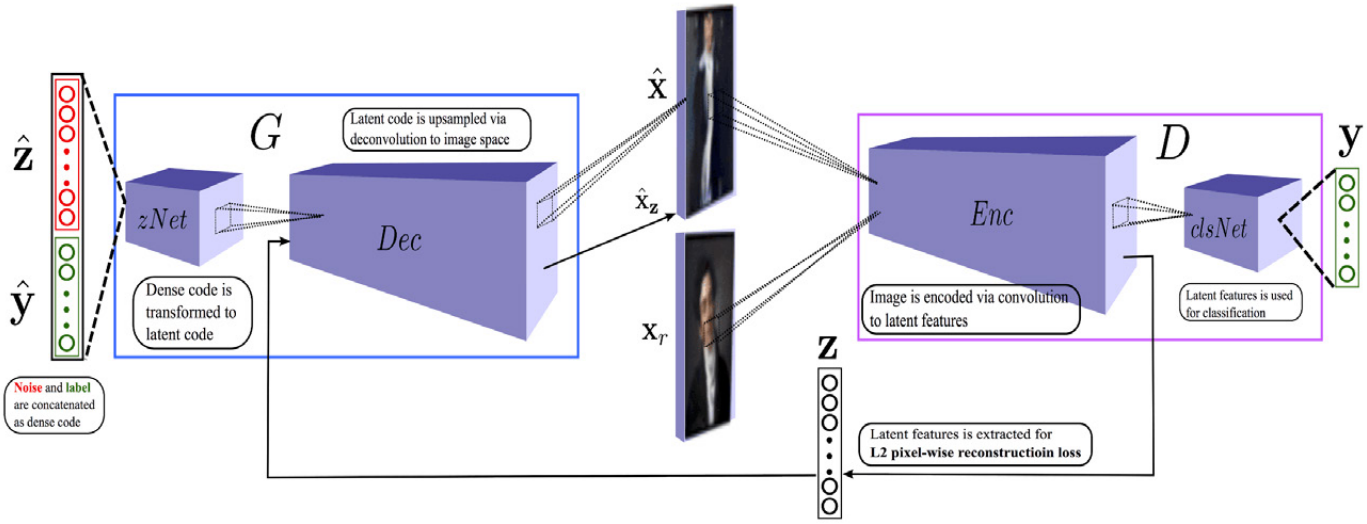
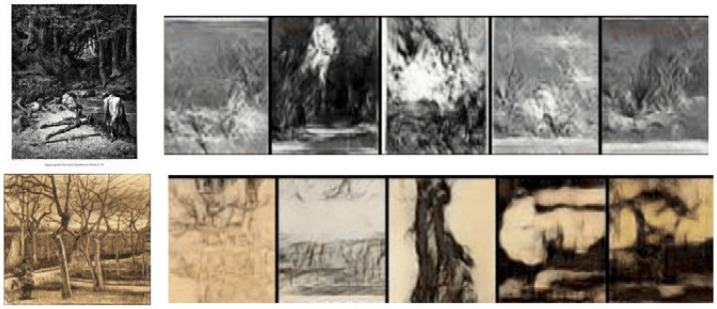


Fig. 11: ArtGANs architecture

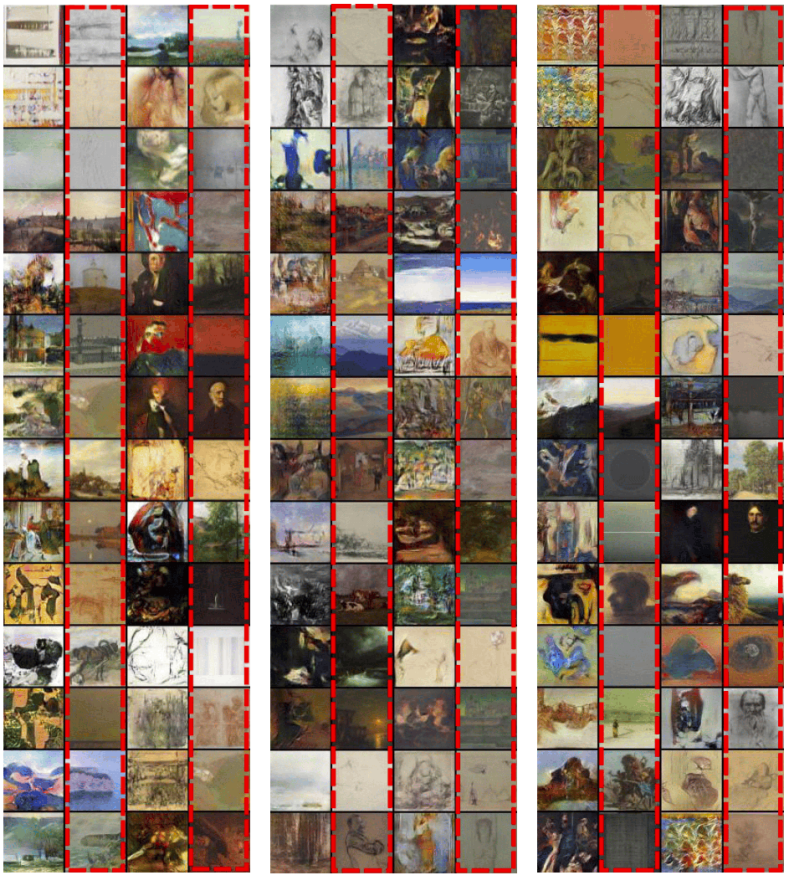
The architecture of ArtGAN closely resembles that of a standard GAN, with a few key modifications. Additional input labels, denoted as $\hat{y}$, are incorporated into the generator (G), while the discriminator (D) outputs a probability distribution over the labels. Furthermore, a connection is established between the encoder (Enc) and decoder (Dec) to facilitate image reconstruction, authorizing the application of an L2 pixel-wise reconstruction loss. A unique feature is the incorporation of label feedback. Randomly assigned labels are given to each generated image, and this label information is integrated into the loss function of the discriminator. By feeding these labels back into the generator, the model enhances its ability to generate images that align with the assigned categories. The discriminator in ArtGAN is responsible for distinguishing real from fake images and functions as a classifier. It outputs a probability distribution over the assigned labels, using cross-entropy loss to back-propagate the error. This classifier-based training allows the discriminator to provide more informative feedback to the generator, accelerating the learning process and improving image quality. The generator and the discriminator are trained simultaneously using the min-max formulation of GANs. And unlike conventional GANs, ArtGAN employs a sigmoid activation function. This choice influences the learning dynamics, impacting how the discriminator processes outputs and contributes to improved stability during training. In addition to the adversarial loss, the L2 pixel reconstruction loss is included to prevent structural coherence in the generated images and mode collapse. ArtGAN's architecture also features an encoder-decoder design, where the decoder in the generator shares the same network with the encoder in the discriminator. This architectural modification strengthens the connection between the generator and discriminator, enabling more effective learning and better reconstruction of images [Fig. 11].



(a) Real (b) Synthesis

Fig. 12: Artwork synthesis

The researchers explore the application in artwork synthesis based on genre, artist, and style [Fig. 12, 13]. Trained on the dataset Wikiart [23] ArtGAN demonstrates by experimental results the capability to synthesize artworks reflecting diverse artistic styles and produce natural-looking images with defined shapes on datasets such as CIFAR-10 [24].



(a) Genres (b) Artist (c) Style

Fig. 13: Example of an inserted image.

The authors outline areas for improvement. Enhancing ArtGAN with a more profound architecture could help capture finer details and more complex concepts, addressing its limitations in generating intricate features. Additionally, the reliability of log-likelihood as a performance metric is questioned, as evaluating abstract visual components is challenging. ArtGAN's output may not fully reflect artistic quality. Furthermore, improving the accuracy of photorealistic image generation remains an area for potential refinement. [25]

StyleGAN: A Style-Based Generator Architecture for Generative Adversarial Networks, 2019 [26] In 2019, NVIDIA introduced the first StyleGAN model, a GAN that revolutionized photo-realistic image generation. This model was groundbreaking as it automatically learned to separate different visual attributes of images without requiring human supervision. After training, various combinations of these attributes could be synthesized to create highly diverse and realistic outputs. The research paper presents a novel generator architecture for GANs, incorporating style transfer concepts to enhance image synthesis capabilities. The architecture enables the automatic and unsupervised decomposition of high-level attributes, such as pose and identity in human faces, and stochastic details, like freckles or hair. This scale-specific control mechanism enhances the interpretability and precision of the outputs by allowing greater control and flexibility. Furthermore, the study introduces two automated techniques for evaluating interpolation quality and feature disentanglement, making them applicable to any GAN-based generator model. The following paragraphs investigate the architecture, style mixing, stochastic variance in detail, and disentanglement.

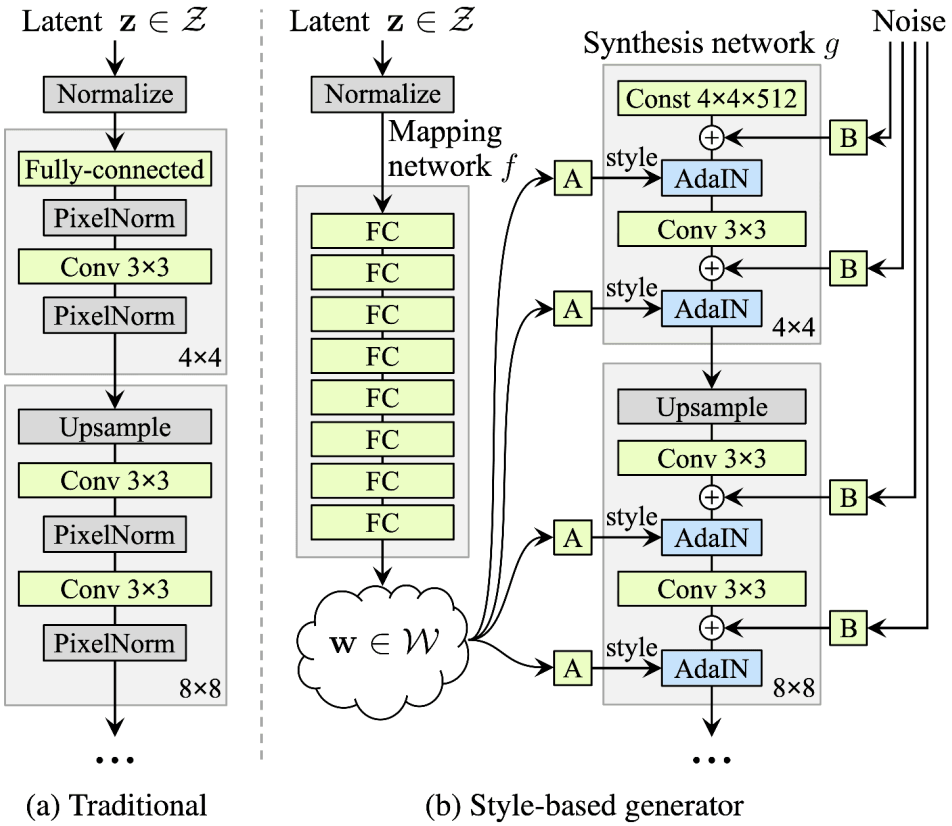


Fig. 14: StyleGAN architecture

The Style-based generator uses an intermediate latent space (W) and adaptive instance normalization (AdaIN) to control image synthesis, along with stochastic variation for finer details. In a traditional generator, the latent code is fed solely through the input layer. However, the StyleGAN architecture** first maps the input to an intermediate latent space (W), which then influences the generator via adaptive instance normalization (AdaIN) at each convolution layer. Additionally, Gaussian noise is introduced after each convolution, before applying nonlinearity. In this framework, A represents a learned affine transformation, while B applies per-channel scaling factors to the noise input. The mapping network f consists of eight layers, whereas the synthesis network g has 18 layers—two per resolution level (from 4^2 to 1024^2). The final layer's output is converted to RGB using a separate 1×1 convolution, following the approach of Karras et al. [27]. Overall, the generator in this architecture has 26.2M trainable parameters, compared to 23.1M in the traditional generator, highlighting its increased complexity and capacity [Fig. 14].

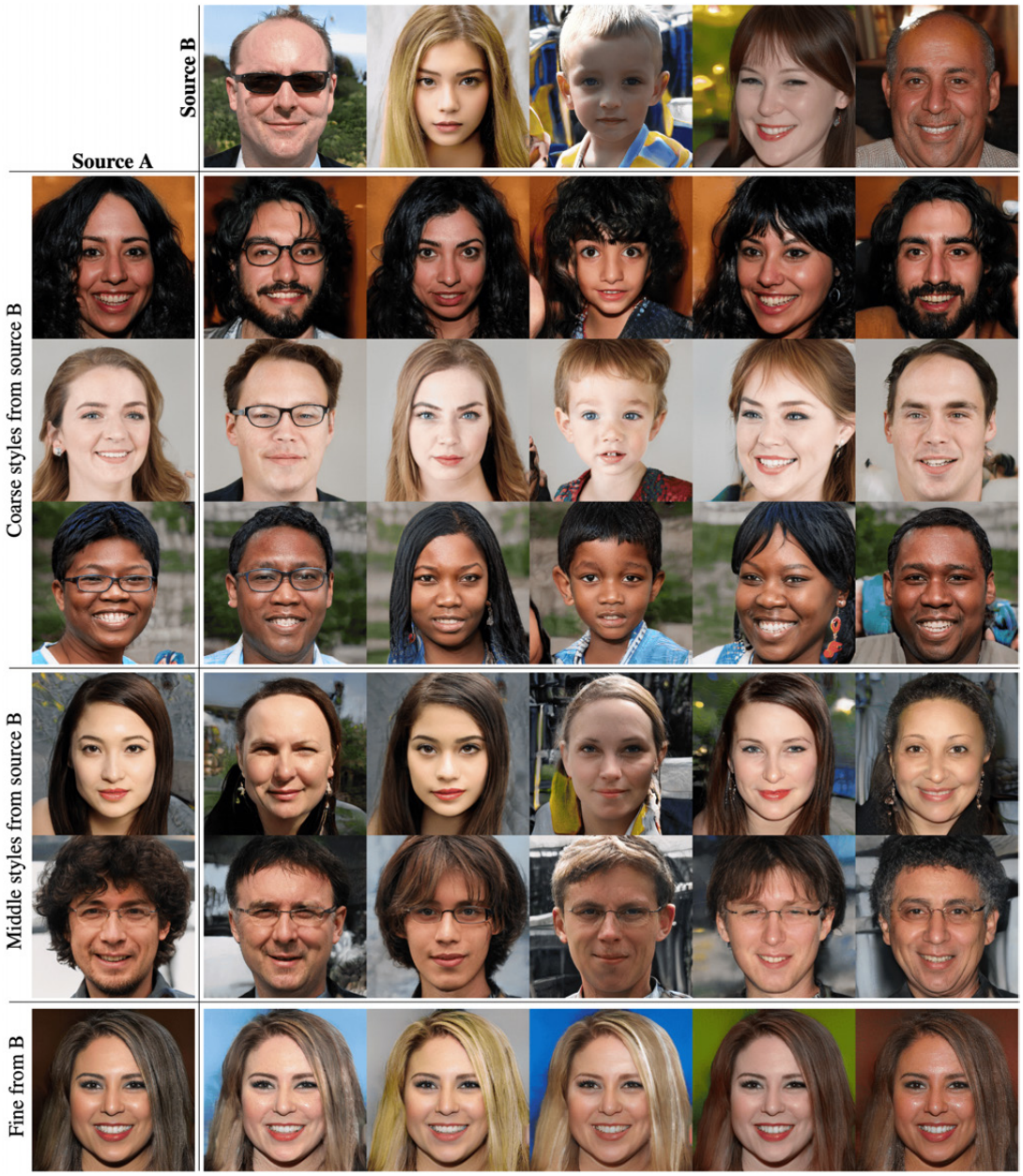


Fig. 15: Style mixing regularization

Style mixing regularization is employed, where images are generated using two random latent codes to encourage styles to localize. This prevents the network from assuming that adjacent styles are correlated. Two sets of images were generated from latent codes of sources A and B, with additional images created by selectively copying styles from B while retaining features from A. Copying coarse-resolution styles (4^2–8^2) from B transfers pose, face shape, and eyeglasses, while colors and finer details remain from A. Middle-resolution styles (16^2–32^2) influence facial features, hairstyle, and eye openness from B, while pose and face shape stay from A. Fine-resolution styles (64^2–1024^2) mainly affect color and microstructure, adopted from B, keeping A's overall shape. This method enables precise control over image attributes at different levels of detail [Fig. 15].

The generator introduces explicit noise inputs as single-channel images of uncorrelated Gaussian noise to each synthesis network layer, enabling direct stochastic detail generation. Examples of stochastic variation show how input noise influences fine details while preserving global structure. Two generated images (a) appear nearly identical, but zooming in (b) reveals differences in individual hair placement. A standard deviation map (c) over 100 realizations highlights noise-affected regions, mainly hair, silhouettes, background, and eye reflections, while identity and pose remain unchanged [Fig. 16].

(a) Generated image (b) Stochastic variation (c) Standard deviation

Fig. 16: Stochastic variation

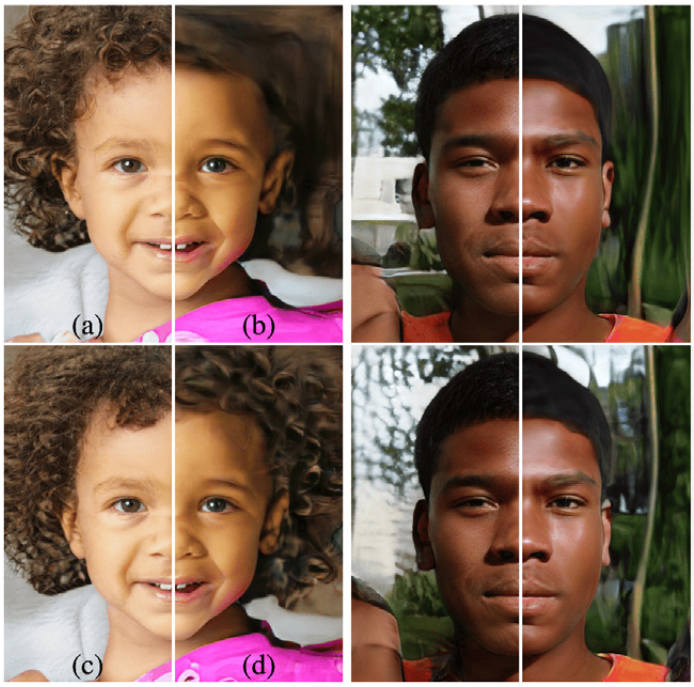
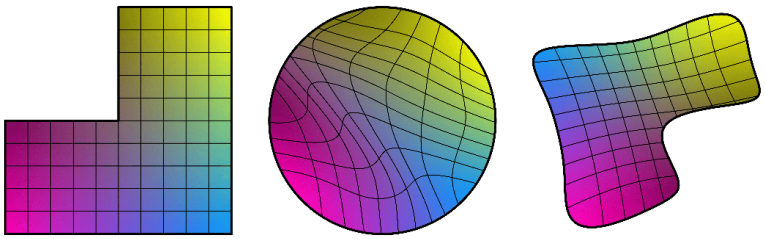


Fig. 17: Example of an inserted image.

Further, the effect of noise inputs varies across generator layers. When applied to all layers (a), images appear natural, while removing noise (b) results in a featureless, painterly look. Noise in delicate layers (c) (64^2–1024^2) enhances hair curls, background details, and skin texture, whereas noise in coarse layers (d) (4^2–32^2) influences large-scale hair curling and background structure [Fig. 17].

The intermediate latent space does not have to support sampling according to any fixed distribution. Its sampling density is induced by a learned mapping, which can be adapted to ”unwarp” the space so that the factors of variation become more linear. Perceptual path length and linear separability are used to quantify disentanglement.



(a) Distribution of features in training set

(b) Mapping from $\mathcal{Z}$ to features

(c) Mapping from $\mathcal{W}$ to features

Fig. 18: Image mapping

An illustrative example with two variation factors, such as masculinity and hair length, shows how missing combinations in training (a) force a curved Z-to-image mapping to prevent invalid samples (b). However, the learned Z-to-W mapping (c) corrects much of this distortion, restoring a more natural representation [Fig. 18].

Besides the StyleGAN architecture, NVIDIA set new standards for datasets. To support further advancements, the research also introduces Flickr-Faces-HQ (FFHQ) [28], a high-quality dataset of human faces with significant diversity [Fig. 19]. This dataset improves the training and evaluation standards for GAN-based facial synthesis, further solidifying StyleGAN's impact on generative computational intelligence.

Fig. 19: Example of an inserted image.

An uncurated set of images generated by the style-based generator (config F) on the FFHQ dataset uses a modified truncation trick ($\psi = 0.7$) for resolutions 4^2–32^2, enhancing image quality while preserving diversity.

Two subsequent versions followed: ***StyleGAN2***, 2020 [29] and ***StyleGAN3***, 2021 [30].

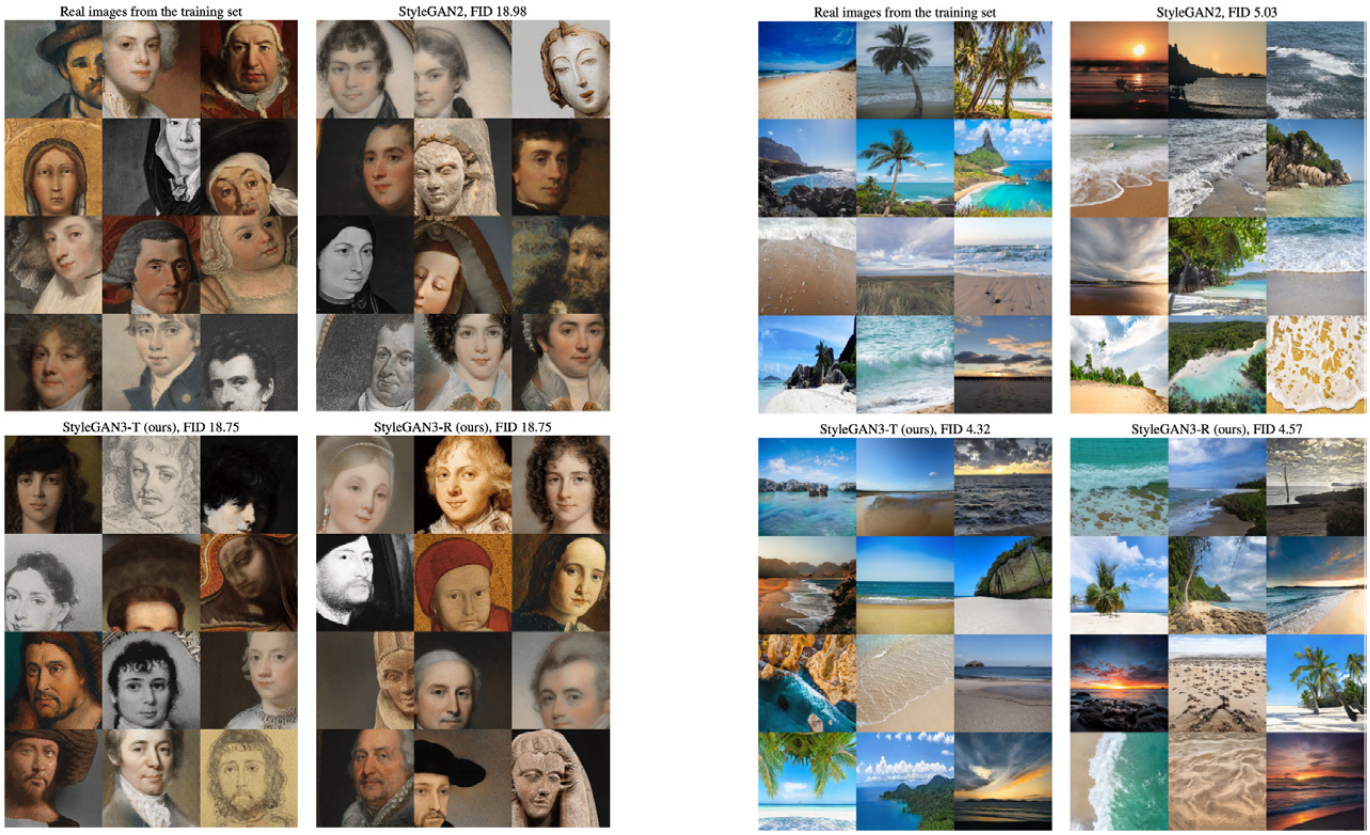


Fig. 20: Comparing feature mapping in StyleGAN2 and StyleGAN3

StyleGAN2 2020 focuses on refining StyleGAN, identifying and eliminating artifacts by redesigning normalization methods and regularising the synthesis network to favour smooth mappings. The revised StyleGAN2 offers improved image quality and enables better attribution of generated images to their source. StyleGAN3 2021 tackles the problem of images appearing glued to specific coordinates by addressing signal processing and aliasing issues within the generator network. By enforcing continuous equivariance to translation and rotation, the resulting alias-free GANs produce images more naturally integrated with their depicted objects, paving the way for better video and animation. Both had notable architectural changes [30]. However, StyleGAN2 addressed artifacts by removing AdaIN, introducing weight demodulation for scale-invariant feature maps, and improving path length regularization for better structure and sharpness. StyleGAN3 further evolved by making the generator fully convolutional, eliminating explicit positional encoding, and using Fourier feature mapping to improve translation and rotation consistency while removing texture sticking artifacts [Fig. 20].

The StyleGAN3 paper highlights several limitations and areas for improvement. Its alias-free generator assumes smooth training data, which can cause issues with jagged edges in pixel art, letterboxed images, or low-quality JPEGs. Measuring equivariance remains ambiguous beyond 40 dB due to inherent resampling limitations, especially for arbitrary rotations. The design also involves trade-offs between antialiasing and high-frequency detail retention, while the generator layers have limited ability to introduce global transformations, making input features crucial for orientation. The paper suggests making the discriminator equivariant for future improvements, as inconsistencies, like misaligned teeth in face rotations, stem from its pixel-based learning. It also proposes reintroducing stochastic noise in a hierarchical manner, improving path length regularization to ensure smooth feature movement, and extending equivariance to scaling and other transformations. Additionally, it suggests performing antialiasing before tone mapping, as current GANs, including StyleGAN3, operate in the sRGB color space after tone mapping.

CycleGAN: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, 2020 [31]
The key challenge addressed by this research is performing image-to-image translation without requiring paired examples of corresponding images in the two domains.

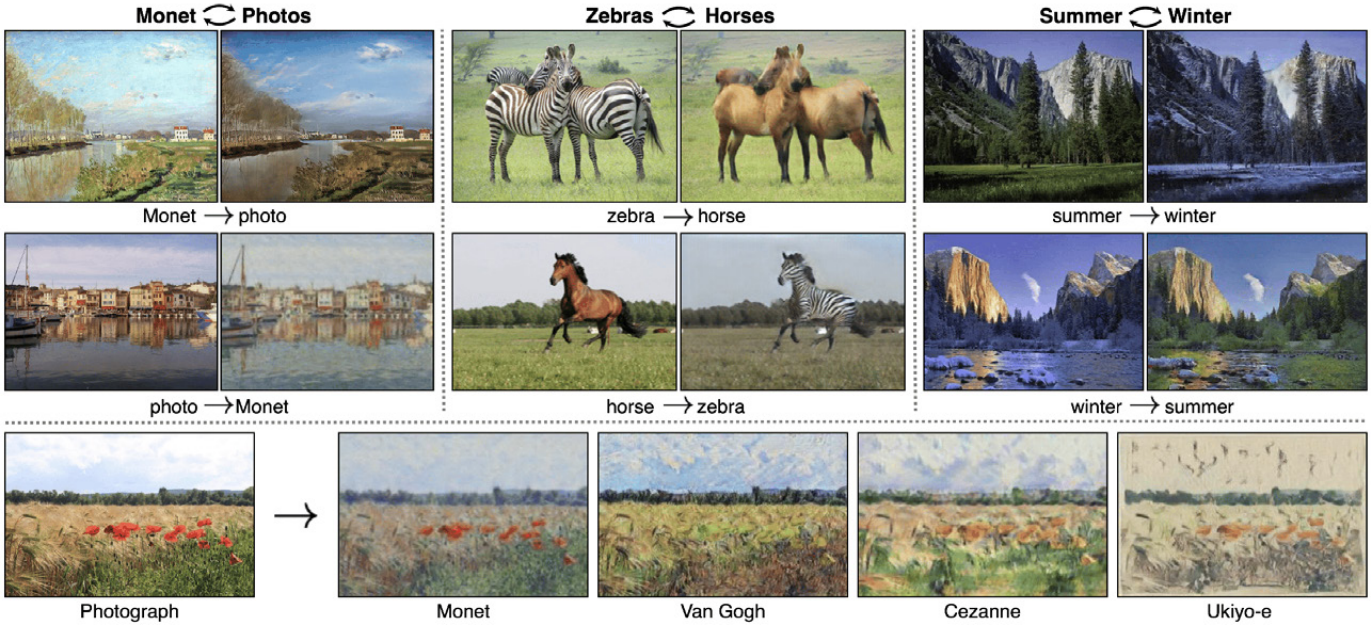


Fig. 21: Unpaired image-to-image translation

To solve this, the study introduces Cycle-Consistent Adversarial Networks (CycleGANs), which use two mappings between the source and target domains along with a cycle consistency loss to ensure that transformations are reversible. This method learns to capture and transfer characteristics from different image collections, enabling tasks such as style transfer, object transfiguration, and seasonal changes. The effectiveness of CycleGAN is demonstrated through qualitative and quantitative results, highlighting its ability to perform unpaired image translation. CycleGAN operates in an unsupervised manner, learning from unpaired images to discover meaningful mappings between domains. This allows transferring artistic styles, such as converting a Claude Monet painting into a realistic photograph or vice versa. Additionally, it can be used for domain adaptation tasks, such as transforming horses into zebras and vice versa, or even changing winter landscapes into summer scenes and vice versa, showcasing its versatility in image transformation tasks [Fig. 21].

The methodology enables unpaired image-to-image translation by learning a mapping $G : X \to Y$, ensuring that $G(x)$ is indistinguishable from real images in domain Y using an adversarial loss, without requiring paired data.

The model consists of two mapping functions, $G : X \to Y$ and $F : Y \to X$, which translate images between domains. Two adversarial discriminators, D_X and D_Y, distinguish between real and translated images, enforcing realism through an adversarial loss that makes $G(x)$ and $F(y)$ indistinguishable from actual samples in their respective domains.

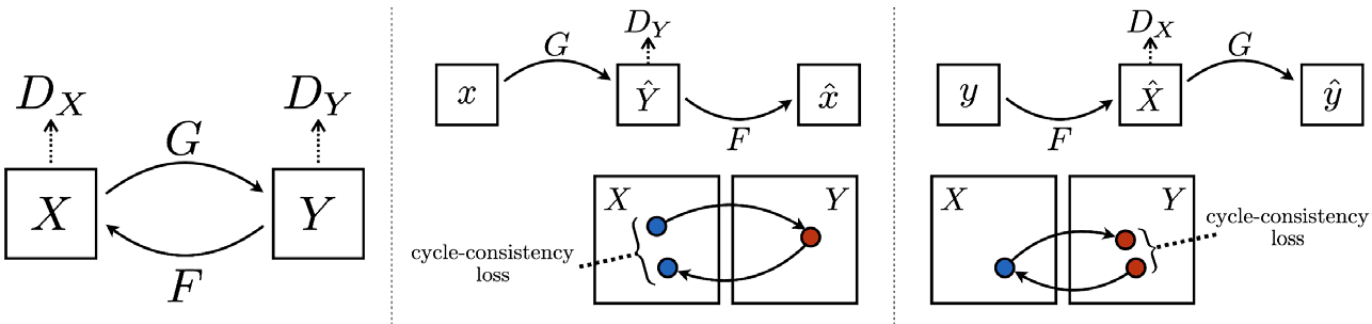


Fig. 22: CycleGANs process

A crucial component is the cycle consistency loss, ensuring that translating an image to another domain and back preserves its original form. Forward consistency is defined as $x \to G(x) \to F(G(x)) \approx x$, and backward consistency as $y \to F(y) \to G(F(y)) \approx y$, minimizing reconstruction errors with $L_{\text{cyc}}(G, F)$.

The objective function combines adversarial losses and cycle consistency loss, with λ controlling their relative importance.

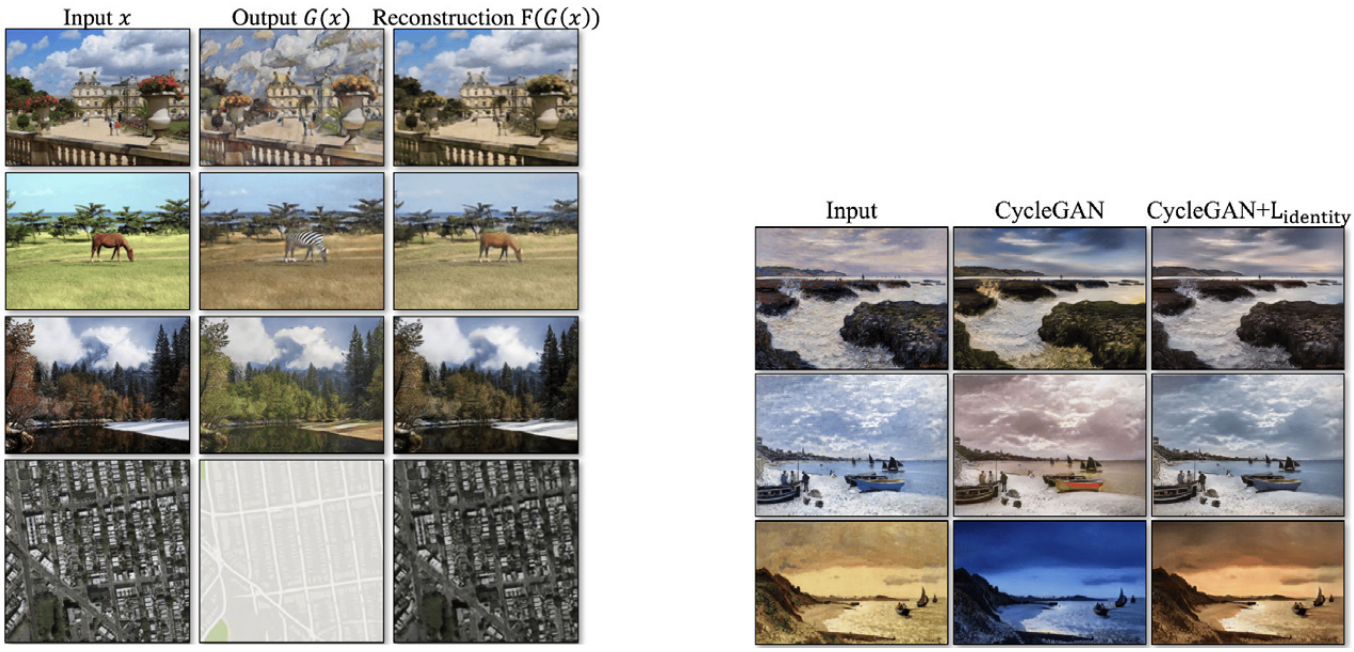


Fig. 23: Visualization of CycleGANs image processing

The generator architecture includes three convolutions, residual blocks, fractionally-strided convolutions, and an output layer mapping features to RGB. PatchGANs classify 70×70 overlapping image patches as real or fake.

To stabilize training, least-squares loss replaces the standard adversarial loss, and a history buffer of generated images reduces model oscillation. The model is trained using Adam optimization (batch size = 1, learning rate = 0.0002). Conceptually, CycleGAN acts as two autoencoders, $F \circ G : X \to X$ and $G \circ F : Y \to Y$, where each input reconstructs itself through an intermediate representation in the opposite domain. For painting $\to$ photo translation, an additional loss function was introduced to ensure that the mapping preserves the color composition between the input and output images [Fig. 23].

CycleGAN is applied to various tasks, including collection style transfer (e.g., Monet to photos), object transfiguration (e.g., horses to zebras), and season transfer (e.g., summer to winter) [Fig. 21].

CycleGAN has achieved impressive results [Fig. 24] in image-to-image translation but also has notable limitations. These challenges stem from the network architecture, dataset characteristics, and the complexity of certain transformations. Typical failure cases of CycleGAN include limited transformations in specific tasks, such as dog-to-cat conversion, where only minimal changes are made to the input. Similarly, in horse-to-zebra translation, the model struggles with scenarios it hasn't encountered in training, such as horseback riding, leading to unrealistic results. A performance gap remains between models trained with paired data and those using unpaired data, like CycleGAN. In cases such as photos-to-labels translation, where tree and building labels are permuted, this gap is difficult to close and may require weak semantic supervision. While CycleGAN effectively handles color and texture changes, it struggles with structural modifications and extreme transformations, highlighting the need for additional refinements in future research [Fig. 25].

Fig. 24: CycleGANs image-to-image translation

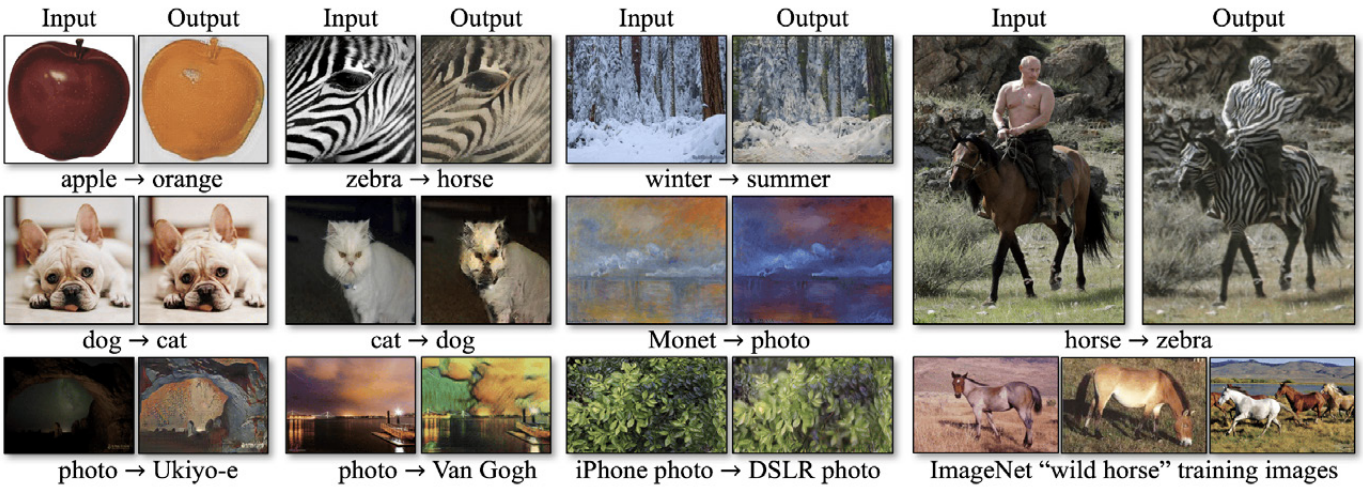
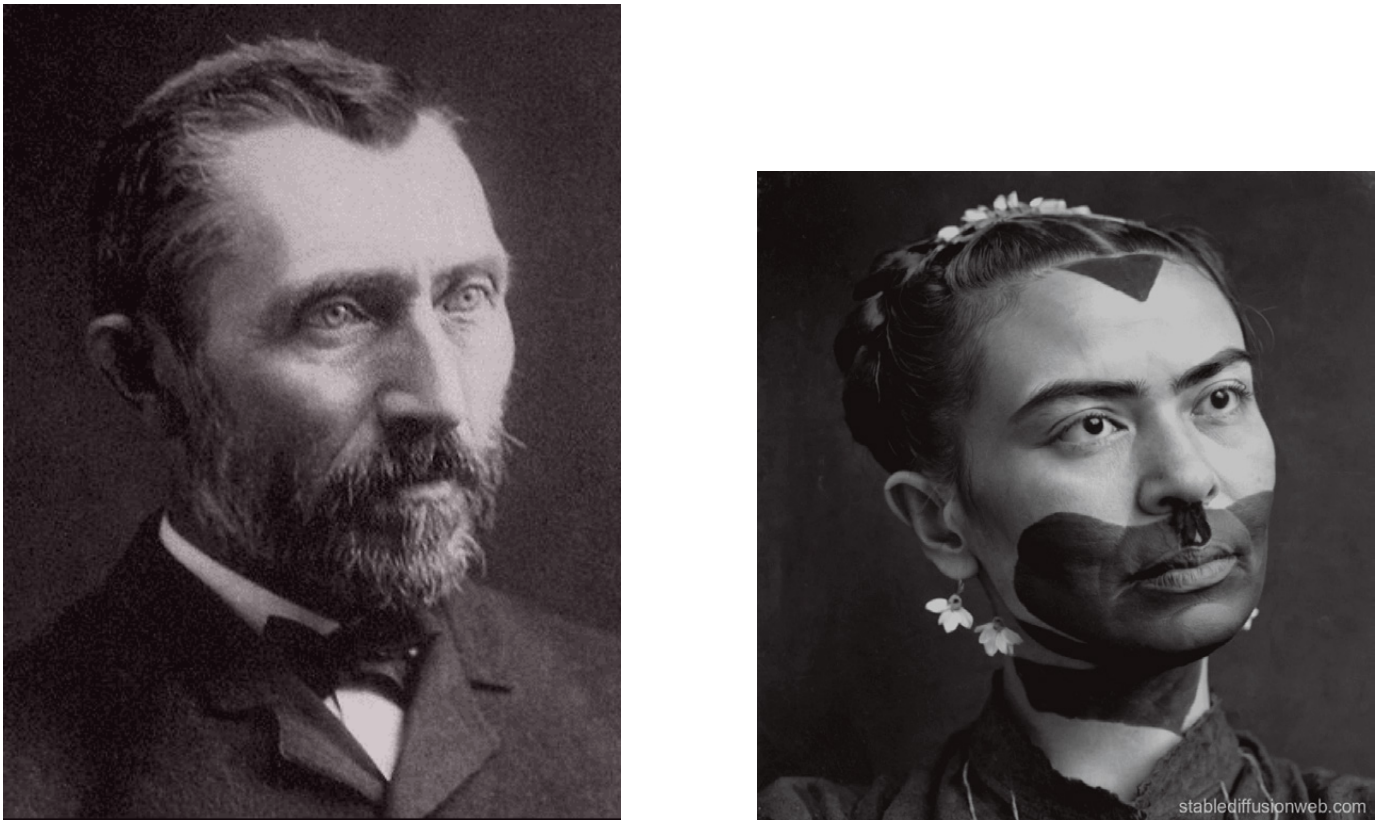


Fig. 25: Limitation characteristics

3.3 Diffusion-based Style Transfer

Latent Diffusion Models (LDMs) [32] represent the latest advancement in generative style transfer, leveraging their ability to generate images from noise. A simple approach involves using text-to-image diffusion models for image-to-image translation, where an input image is provided alongside a text prompt describing the desired style. Models like Stable Diffusion and DALL·E 2 [33] are trained to reverse the noise diffusion process, generating high-quality images conditioned on text, allowing them to replicate a vast range of visual styles. Diffusion-based style transfer takes advantage of these pre-trained models by guiding generation through text prompts or modifying the latent input using a reference image. The key advantage is that all styles are already embedded in the trained model, eliminating the need for additional training-only prompt engineering is required. This allows Stable Diffusion to convert a photograph into a stylized photo (b) or painting, simply by prompting ”paint it as Frida Kahlo” while using a photograph that gained notoriety due to unsubstantiated claims that it depicted Vincent van Gogh (a), as a reference. This text-guided approach has made diffusion models a powerful tool for style transfer and image editing [Fig. 26].



(a) *Artiste-Photo* by Victor Morin, ca. 1886 [34]

(b) Output Image

Fig. 26: Promt: paint it like Frida Kahlo

ControlNet: Adding Conditional Control to Text-to-Image Diffusion Models, 2023 [35] This article introduces ControlNet, a neural network architecture designed to enhance text-to-image diffusion models, such as Stable Diffusion, by incorporating spatial conditioning controls. This extension allows for more precise control over image generation, addressing challenges in accurately representing complex layouts, poses, shapes, and structures, which are often difficult to achieve using text prompts alone [Fig. 27]. ControlNet serves as a plugin for latent diffusion models, significantly improving the flexibility and accuracy of text-to-image synthesis by conditioning the model on additional structural guidance. The plugin architecture locks the production-ready large diffusion models. It reuses their deep and robust encoding layers, which are pre-trained with billions of images, as a strong backbone to learn a diverse set of conditional controls [Fig. 28].

Fig. 27: ControlNet combining multiple input types

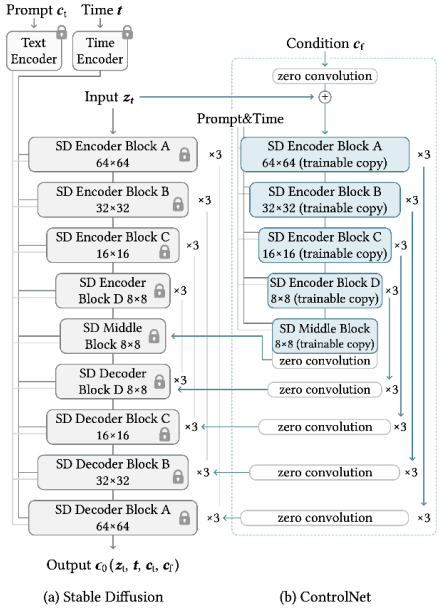


Fig. 28: ControlNet plugin architecture

The trainable copy and the original, locked model are connected with zero convolution layers, with weights initialized to zeros so that they progressively grow during the training. This architecture ensures that harmful noise is not added to the deep features of the large diffusion model at the beginning of training, and protects the large-scale pretrained backbone in the trainable copy from being damaged by such noise.

ControlNet enables using various conditioning inputs such as Canny edges, Hough lines, user scribbles, human key points, segmentation maps, shape normals, and depths to control image generation. The method supports single or multiple conditions, with or without text prompts. ControlNet's extra conditions can be controlled to affect the denoising diffusion process, notably using Classifier-Free Guidance (CFG) and CFG Resolution Weighting [Fig. 29]. ControlNets do not change the network topology of pretrained Stable Diffusion models so that they can be directly applied to various Stable Diffusion models.

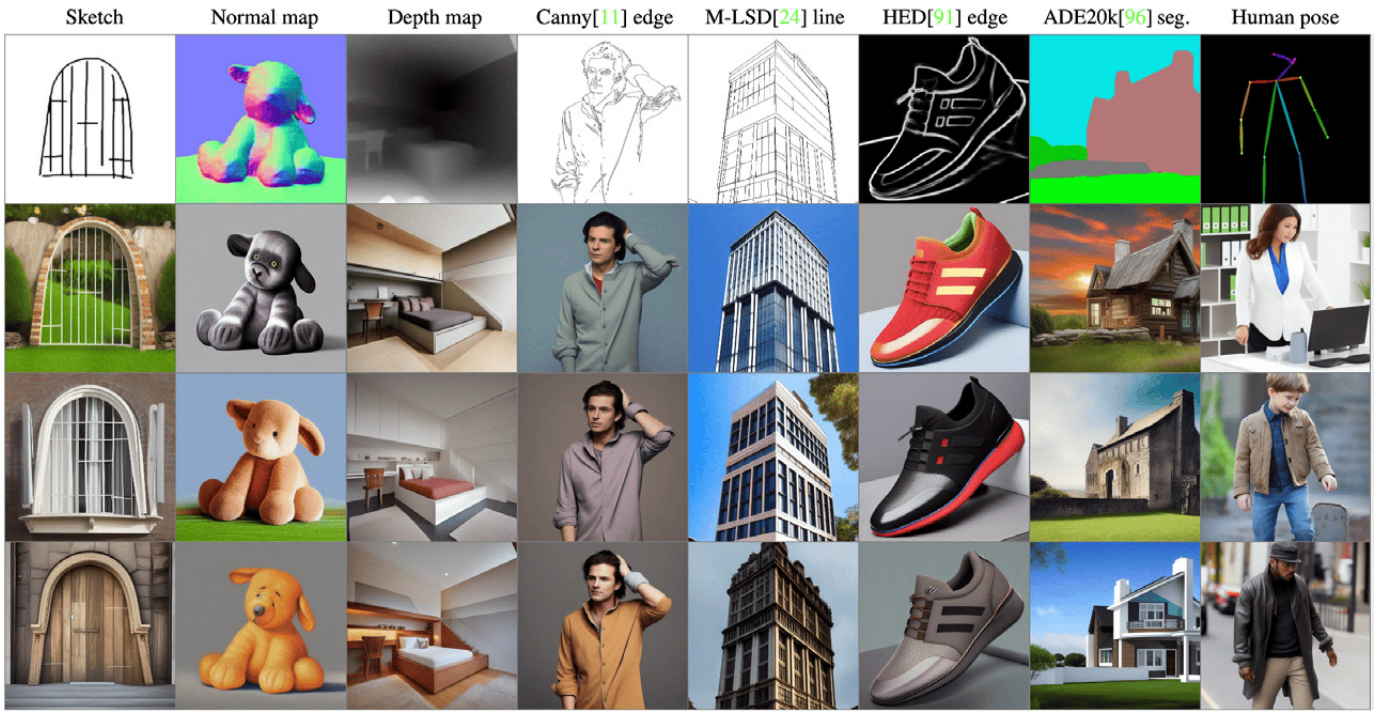


Fig. 29: Conditioning inputs

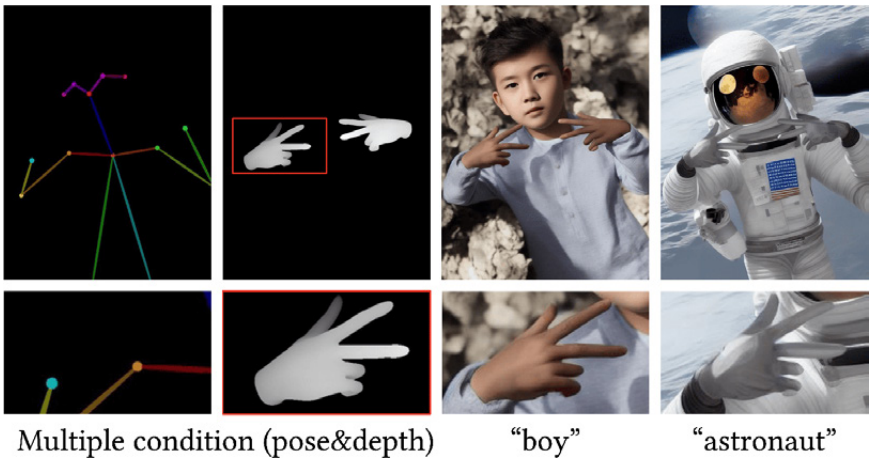


Fig. 30: Composition of multiple ControlNets

Therefore, multiple ControlNets can also be composed to apply multiple conditioning images to a stable diffusion instance [Fig. 30]. Several technical limitations should be considered when evaluating ControlNet. Its performance depends on pretrained models, making it reliant on the underlying diffusion model. While it can be trained on small datasets, the quality and size of training data impact results. Computational demands remain a factor as fine-tuning still requires GPUs, despite being more efficient than training diffusion models from scratch. ControlNet handles various prompt settings but may struggle with conflicting instructions or ambiguous inputs, where the model interprets shapes without clear guidance. The sudden convergence phenomenon poses another challenge, as the model quickly adapts to input conditions, but its underlying mechanisms and control remain unclear. While LDMs reduce computational costs compared to pixel-based methods, their sequential sampling remains slower than GANs. Lastly, training Stable Diffusion models remains complex and resource-intensive, with super-resolution models already facing intrinsic limitations [36].

4 Style Transfer Applications and Cultural Representation

There are various approaches to trying out or utilizing this technology. The most suitable solution depends primarily on the background and skills of the interested person.

4.1 Open Source Model Platform

Hugging Face [37] is a leading machine learning platform specializing in natural language processing, computer vision, and generative computational intelligence. It offers open-source tools, pre-trained models, APIs, and diffusion-based pipelines, making AI development more accessible to researchers and developers. With resources like the Hugging Face Model Hub, Pipelines, and Spaces, users can explore different styles, fine-tune models, and deploy interactive style transfer applications online.

4.2 Free Style Transfer Browser Applications

Free browser applications are available for fully automated style-based image transformation, offering optimized image processing as a chargeable update.

(Go Art) Fotor [38] applies a variety of artistic filters and effects to images reminiscent of sketches, cartoons, or other well-known art genres.

Stable Diffusion Web [32], [39] is built on the open-source Stable Diffusion model. It enables users to generate realistic, artistic, and surreal visuals from text prompts or through image-to-image transformation. With an intuitive interface, users can adjust image resolution, fine-tune prompts, and apply style variations, making the platform accessible to beginners and advanced creators.

DALL·E [33] by OpenAI leverages deep learning techniques, specifically a variant of GPT trained for image synthesis, to generate images from text descriptions [Fig. 31].

(a) Stable Diffusion Web

(b) DALL·E

Fig. 31: Promt: paint a Portrait of Vincent van Gogh

MidJourney [40] was developed as an independent research project and operates through Discord. It gained popularity among artists, designers, and storytellers, who use it for concept art, marketing, and creative design projects. These systems have become fundamental for the NFT and digital art markets, where businesses and artists create immersive digital experiences in gaming, virtual realities, and 3D Modeling. These systems have become fundamental for the NFT and digital art markets, where businesses and artists create immersive digital experiences in gaming, virtual realities, and 3D Modeling.

4.3 Commercial Implementation of Style Transfer

Initially developed for artistic purposes, Style Transfer has found widespread commercial applications across various industries. In advertisement, eye-catching promotional content is created, transforming product photos into artistic renderings that align with brand aesthetics. In the fashion industry, clothing designs are visualized through the use of textures, and stylized product previews are generated. It enhances film production, gaming, and digital content creation by transforming live footage into animated or painterly styles. Neural filters are applied to social media and mobile sales apps to invigorate user-generated content. Brands integrate style transfer into e-commerce platforms, allowing customers to apply different effects to products, rooms, or body parts. *Adobe* offers a variety of customizable automation methods in the mentioned areas, ranging from Stable Diffusion models to neural filters. [41] [Fig. 32]

Fig. 32: Using Photoshop Neural Filters, I changed the photograph of Frida Kahlo [Fig. 3]

While style transfer is commonly associated with creative applications, it also plays a role in critical fields like medicine and defense, where advanced image processing is essential. In medicine, style transfer enhances diagnostic imaging by improving the clarity of MRI, CT, and X-ray scans. It can help radiologists detect anomalies more easily and standardize imaging across different devices, ensuring consistency in medical analysis. Additionally, generated stylized medical images are valuable for training and simulation, allowing healthcare professionals to visualize complex conditions innovatively. In defense and security, style transfer is applied to satellite image enhancement, improving the resolution and readability of surveillance data. This can aid in reconnaissance and intelligence analysis by clarifying low-quality or obscured images. Another critical use is camouflage detection, where AI transforms visual data to help identify hidden or disguised objects. Additionally, style transfer creates diverse environmental conditions in military training simulations, improving situational preparedness.

4.4 Societal Impact and Challenges

The provision of trained open-source models democratizes access to technology and its exploration. However, despite its potential, style transfer requires careful validation, particularly in medical diagnostics and security

intelligence, to ensure accuracy and reliability. The representation of human features should also be diverse and representative of the respective culture and its ethical values. Since neither visual nor content-related biases can be fully corrected by random parameters, thoroughly developed tests and comprehensive statistical analyses are necessary and must be professionally assessed. Watermarks and traceable authorship can help identify and actively combat deliberate image manipulations, such as Deep Fakes, which disproportionately affect minorities and women. A professional review of the automated generation of training data is particularly important. For example, classifying scientific data requires subject-matter expertise, especially in fields like art history, where cultural appropriation is not always easy to assess, or when dealing with historical records documented inconsistently. The Western perspective on art history cannot be universally applied to other cultures. As highlighted in a comment in the introduction to ArtGAN [22]:

> *In the philosophy of art, aesthetic judgement is always applied to artwork based on one's sentiment and taste, which shows one's appreciation of beauty.*

5 Artefacts as a Style in Contemporary Art

5.1 Cultural Reinterpretation of Artifacts through Artworks

Generated renderings are an integral part of contemporary art. Indications of technical prerequisites as visual features suggest the appropriation of innovative technology. Artifacts make the renderings identifiable as such and thus serve as visual style components. The intensive and, at times, visionary engagement with mechanical and technological developments has always been inherent in the history of art and philosophy. Reflecting the spirit of the times, the visual characteristics of artifacts evolve alongside technological advancements. These visual features are technical artifacts that can be deliberately used as stylistic elements or accepted conceptually across various genres.

Blurred color transitions create blind spots or out-of-focus areas (a); blending techniques emerge like a painted photographic style (b); and glitches appear, such as errors in extremities, the number of limbs, or deformations (c). These elements inspire artistic storytelling as well as conceptual reflection on the rapid pace of technological advancement. [Fig. 33]

(a) *Edmond de Belamy* by Obvious (collective), 2018 [42]

(b) Created with DALL·E by Encik Tekateki, 2022 [43]

(c) *Incorrect prompt conversion: Three arms* by Fährtenleser, 2024 [44]

Fig. 33: AI generated art

5.2 Artworks Identifiable by Artifacts as a Style Feature

The following contemporary artworks, ranging from sculpture and conceptual art to prints and virtual reality, show stylistic characteristics of visual artifacts.

Artwork Series ***Adversarially Evolved Hallucination*** by Trevor Paglen, 2017-ongoing

The artist Trevor Paglen curated training datasets on allegorical art and metaphor, drawing from image vocabularies rooted in literature, philosophy, poetry, folklore, and spiritual traditions. He trained GANs to generate images, collectively titled *Hallucinations* [45]. Often perceived as a flaw or glitch in the system, hallucination is intrinsic to image-generation models, forming a fundamental aspect of computational intelligence. Paglen's series *Adversarially Evolved Hallucinations* reveals this spectral and hallucinatory realm by exploring the latent space of computer vision. His work examines how we can engage with these opaque structures from within while challenging the often-exaggerated claims surrounding automated image production systems. These generated artworks are exhibited as dye sublimation prints [46].

3D-Rendering ***The 3D-Google-Earth-Model #7*** by Achim Mohné, 2020

The generated video [47] features Peter Weibel's land art piece *The Globe as a Suitcase*, 2004, located in the Austrian Sculpture Park in Premstätten.The artist simulated a drone flight around the sculpture, which was created using a three-dimensional model based on Google Earth data and photogrammetric processing. Weibel's sculpture symbolizes a once mobile society that has now come to a standstill. The quirks of the algorithm and the evolving ways we perceive art in public space highlight the longstanding relationship between photography and sculpture, dating back to the early 1930s. The project demonstrates how digital algorithms introduce imperfections and unexpected distortions, using satellite imagery from virtual globes like Apple Maps and Google Earth, while opening new perspectives on public art. Traditional land art, which is inherently connected to the relationship between space and time, is reinterpreted in virtual space, fundamentally transforming concepts of distance, proximity, and time. *The 3D-Google-Earth-Model #7* captures these digital anomalies, surreal distortions, floating trees, and missing paths, reflecting the simplified language of the algorithm.

Exhibition ***Forced Amnesia*** by Mary-Audrey Ramirez, 2023

The artist Mary-Audrey Ramirez stages in the exhibition a transition into another world, inspired by video games and digital technology [48], [49]. The designed scenery imitates a virtual gaming environment, establishing a connection to its own iconography and terminology. It is a world of creature-like beings, which the artist developed using the text-to-image generator Stable Diffusion. The morphological artifacts of the generated creatures shape their aesthetic appearance. Ramirez connects these motifs to the ongoing discourse on artificial intelligence, an established field of research since 1956, as well as to considerations of posthumanism. The exhibition is defined by visual tactility and complex materialities. At its center is the Boss Pit, a white vinyl surface staged as a battle arena. The sculpture *Somebody's Basilisk* stands within it as a gatekeeper, symbolizing progress and guarding the transition to a new stage of development. Other works incorporate religious visual concepts, including a sculptural triptych with the central panel titled *The Lovers (in Love)*, flanked by two versions of *The Haters (in Goo)*, digital images as satin prints mounted on panne velvet. Advanced 3D printing techniques are also showcased in sculptures like *Happy Face (Parasites)*, made from sand and binder. These pieces are exhibited alongside multimedia installations that include animated films created with the Unreal Engine, a computer graphics game engine. The experience is accompanied by an electronic ambient soundtrack composed for the exhibition by Simon Goff.

3D Aluminum Print Sculpture ***Low Poly Tree*** by Achim Mohné, 2023

The awarded sculpture *Low Poly Tree* [50] takes form of a virtual tree as it appears in Google Earth's 3D mode, bringing it back into reality at a 1:1 scale, $207.5 \times 80.3 \times 73.8$ cm. Positioned opposite its real-life counterpart on Grafenwerth Island, Germany, the polished steel tree reflects its surroundings, remaining unchanged, unlike the living tree that grows and evolves with the seasons. This work examines the representation of nature in art, ranging from painting to digital imagery, and highlights how platforms like Google Earth influence our perception of the natural world. The mirrored surface invites self-reflection and social media interaction. Produced using digital techniques, sustainable materials, and climate-neutral methods, the sculpture's carbon footprint is measured and offset.

6 Future Directions

In addition to the well-researched 2D style transfer methodology, which ranges from abstract to photorealistic images, research efforts are increasingly focusing on extending 2D results into 3D dimensions. As outlined by

Kotovenko et al. [51], some researchers in this field explore solutions that leverage access to the style target geometry. However, the challenge remains in handling complex and diverse 3D scenes with full 360-degree camera rotation, as the target depth is typically defined for a single fixed viewpoint [52]–[54]. At the same time, another line of research focuses on modifying only the geometry of shapes without incorporating style textures [55], [56]. These transformations are attempted using 2D-based information and common 3D representations. [57]–[61]. Another approach to providing style information involves integrating text prompts that specify desired changes. The method [62] employs text-based inputs, where the user specifies an artist's name to generate an image in that particular artistic style. Still, methods that rely on image or text prompts rarely alter the geometry of the scenes. Despite these advancements, all of these approaches effectively transfer colors and textures but still struggle to replicate the geometric structure of the scenes realistically. The recent study introduces a method for stylizing 3D scenes using another 3D style scene as a reference. To address the challenges of scene-to-scene translation, the approach first partitions the content scene into simpler components, then identifies the most suitable style segment for each content cluster, and finally refines the stylization by optimizing the Wasserstein-2 distance between each pair of clusters. By leveraging this distribution-matching approach along with a robust representation of style scenes, the method achieves an accurate reproduction of the style scene's geometry. Unlike traditional techniques that render scenes to align with feature distributions in an image encoder space, this approach offers a more structured and geometry-aware alternative to conventional 3D scene processing. To evaluate the method's effectiveness, ablation studies are conducted on various components, assessing the influence of different parameters on the final performance [51]. The computational demands and scalability of the computationally intensive generation of style transfer in film sequences remain challenging.

7 Conclusion

This report comprehensively examines the computer vision methodology of style transfer. A closer analysis highlights the crucial role this technology plays across cultures. The rapid technological advancements in this field have been simultaneously documented and evaluated within Western contemporary art. Continuing advancements in machine learning, optimization strategies, and real-time processing will further enhance the potential of style transfer. Improved accessibility, adaptability, and integration across industries, including art, security, and education, will drive its continued development. Style Transfer has significant potential in academic research for preserving, restoring, and analyzing cultural heritage. As the technology evolves, striking a balance between creativity, efficiency, and ethical responsibility will be essential to maximizing its benefits.

Acknowledgments

I want to express my deepest gratitude to the seminar advisors, Olga Grebenkova and Timy Phan, for their comprehensive mentorship. I am genuinely grateful for their unwavering support and invaluable guidance throughout the semester. I would also like to sincerely thank Prof. Dr. Björn Ommer for offering such an insightful and engaging seminar.

References

[1] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 2023, pp. 571–576.
[2] A. Hertzmann, "Paint by relaxation," in *Proceedings. Computer Graphics International 2001*, 2001, pp. 47–54.
[3] C. Jacobs, D. Salesin, N. Oliver, A. Hertzmann, and A. Curless, "Image analogies," in *Proceedings of SIGGRAPH*, 2001, pp. 327–340.
[4] J.-P. Goude, "So far so goude - exhibition," 2025, accessed: 6 March 2025. [Online]. Available: https://www.jeanpaulgoude.com/en/work/exhibition/so-far-so-goude
[5] A. Gursky, "Official website of andreas gursky," 2025, accessed: 6 March 2025. [Online]. Available: https://www.andreasgursky.com/en
[6] J. M. Robinson, "Artistic style," *Routledge Encyclopedia of Philosophy*, 1998, accessed: 6 March 2025. [Online]. Available: https://www.rep.routledge.com/articles/thematic/artistic-style/v-1
[7] H. Wölfflin, "Kunstgeschichtliche grundbegriffe," München, 1915, accessed Jul. 5, 2025. [Online]. Available: https://digi.ub.uni-heidelberg.de/diglit/woelfflin1915
[8] V. van Gogh, "Self portrait," Oil on canvas; 540x650 mm; Musée d'Orsay, public domain, 1889, gift of Paul and Marguerite Gachet 1949. [Online]. Available: https://artsandculture.google.com/asset/self-portrait/9gFw_1Vou2CkwQ?hl=de
[9] K. Hokusai, "Under the wave off kanagawa jap. kanagawa oki nami ura," Colour woodblock print; 246x365 mm; part of the series Thirty Six Views of Mount Fuji, 1826-1836, held by Art Institute of Chicago © Clarence Buckingham Collection, public domain, accessed Jul. 5, 2025. [Online]. Available: https://www.artic.edu/artworks/24645/

[10] R. C. i Llambí, "Pablo picasso, photographed in paris, 1904," Gelatin silver print; Wikimedia Commons, 1904, public domain, accessed Jul. 5, 2025. [Online]. Available: https://commons.wikimedia.org/wiki/File:Pablo_Picasso,_1904,_Paris,_photograph_by_Ricard_Canals_i_Llamb%C3%AD_cut.jpg

[11] Time Magazine, "Art: Mexican Autobiography," *Time*, vol. 61, no. 17, p. 92, apr 1953, accessed Jul. 5, 2025. [Online]. Available: https://time.com/archive/6620658/art-mexican-autobiography/

[12] G. Kahlo, "Frida kahlo photographed 16 october 1932," Gelatin silver print; Wikimedia Commons,, oct 1932, public domain, accessed Jul. 5, 2025. [Online]. Available: https://commons.wikimedia.org/wiki/File:Frida_Kahlo,_by_Guillermo_Kahlo.jpg

[13] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2414–2423.

[14] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1033–1038 vol.2.

[15] L.-Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000. [Online]. Available: https://api.semanticscholar.org/CorpusID:3131710

[16] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick, "Graphcut textures: image and video synthesis using graph cuts," *ACM Trans. Graph.*, vol. 22, no. 3, p. 277–286, Jul. 2003. [Online]. Available: https://doi-org.emedien.ub.uni-muenchen.de/10.1145/882262.882264

[17] N. Ashikhmin, "Fast texture transfer," *IEEE Computer Graphics and Applications*, vol. 23, no. 4, pp. 38–43, 2003.

[18] H. Lee, S. Seo, S. Ryoo, and K. Yoon, "Directional texture transfer," in *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering*, ser. NPAR '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 43–48. [Online]. Available: https://doi-org.emedien.ub.uni-muenchen.de/10.1145/1809939.1809945

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, pp. 84 – 90, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:195908774

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1409.1556

[21] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014. [Online]. Available: https://arxiv.org/abs/1406.2661

[22] W. R. Tan, C. S. Chan, H. Aguirre, and K. Tanaka, "Artgan: Artwork synthesis with conditional categorical gans," 2017. [Online]. Available: https://arxiv.org/abs/1702.03410

[23] B. Saleh and A. Elgammal, "Large-scale classification of fine-art paintings: Learning the right metric on the right feature," 2015. [Online]. Available: https://arxiv.org/abs/1505.00855

[24] R. Doon, T. Kumar Rawat, and S. Gautam, "Cifar-10 classification using deep convolutional neural network," in *2018 IEEE Punecon*, 2018, pp. 1–5.

[25] W. R. Tan, C. S. Chan, H. Aguirre, and K. Tanaka, "Improved artgan for conditional synthesis of natural image and artwork," 2018. [Online]. Available: https://arxiv.org/abs/1708.09533

[26] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," 2019. [Online]. Available: https://arxiv.org/abs/1812.04948

[27] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," 2018. [Online]. Available: https://arxiv.org/abs/1710.10196

[28] NVIDIA Labs, "Ffhq dataset," 2025, accessed: 6 March 2025. [Online]. Available: https://github.com/NVlabs/ffhq-dataset

[29] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," 2020. [Online]. Available: https://arxiv.org/abs/1912.04958

[30] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," 2021. [Online]. Available: https://arxiv.org/abs/2106.12423

[31] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2020. [Online]. Available: https://arxiv.org/abs/1703.10593

[32] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," 2022. [Online]. Available: https://arxiv.org/abs/2112.10752

[33] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Chen *et al.*, "Dall·e: Zero-shot text-to-image generation," *OpenAI Research*, 2021. [Online]. Available: https://arxiv.org/abs/2102.12092

[34] Victor Morin Artiste-Photo, "Vincent van gogh photo (cropped)," 1886, brussels clergymen circa 1886, with some unlikely claims that this is a portrait of Van Gogh. Discovered in the early 1990s, first exhibited 2004 in Seton Gallery, New Haven. Public domain, accessed: 5 July 25. [Online]. Available: https://commons.wikimedia.org/wiki/File:Vincent_van_Gogh_photo_cropped.jpg

[35] L. Zhang, A. Rao, and M. Agrawala, "Adding conditional control to text-to-image diffusion models," 2023. [Online]. Available: https://arxiv.org/abs/2302.05543

[36] J.-Y. Zhu and collaborators, "Cyclegan project page," 2025, accessed: 6 March 2025. [Online]. Available: https://junyanz.github.io/CycleGAN/

[37] Hugging Face, "Hugging face: The ai community building the future," 2025, accessed: 6 March 2025. [Online]. Available: https://huggingface.co/

[38] Fotor, "Goart - ai photo effects," 2025, accessed: 6 March 2025. [Online]. Available: https://goart.fotor.com/

[39] Stable Diffusion Web, "Stable diffusion online," 2025, accessed: 6 March 2025. [Online]. Available: https://stablediffusionweb.com/

[40] Midjourney, Inc., "Midjourney official website," 2025, accessed: 6 March 2025. [Online]. Available: https://www.midjourney.com/

[41] Adobe Inc., "Adobe products," 2025, accessed: 6 March 2025. [Online]. Available: https://www.adobe.com/products/

[42] Obvious (collective), "Edmond de Belamy," AI-generated; ink print on canvas 700x700 mm, Oct. 2018, public domain, accessed 5 July 2025. [Online]. Available: https://commons.wikimedia.org/wiki/File:Edmond_de_Belamy.png

[43] E. Tekateki, "Promt: "1960's art of cow getting abducted by ufo in midwest"," Aug. 2022, public domain, accessed 5 July 2025. [Online]. Available: https://commons.wikimedia.org/wiki/File:1960%27s_art_of_cow_getting_abducted_by_UFO_in_midwest.jpg

[44] Fährtenleser, "Incorrect prompt conversion: Three arms," Jan. 2024, prompt: "japanese woman with long hair and red lips, dress is kimono, background mount fuji", public domain, accessed 5 July 2025. [Online]. Available: https://commons.wikimedia.org/wiki/File:Incorrect_prompt_conversion.png

[45] T. Paglen, "Hallucinations," 2020, accessed: 6 March 2025. [Online]. Available: https://paglen.studio/2020/04/09/hallucinations/

[46] D. XYZ, "Exhibition 10074," 2025, accessed: 6 March 2025. [Online]. Available: https://daily.xyz/exhibition/10074

[47] A. Mohné, "3d google earth model 7," 2025, accessed: 6 March 2025. [Online]. Available: https://www.xn--achimmohn-j4a.net/3d-google-earth-model-7-die-erdkugel-als-koffer-news.html

[48] T. W. Kuhn, "Mary-audrey ramirez: Forced amnesia," in *KUNSTFORUM International, Band 289: Cuteness*, P. Funken, Ed. Mainz, Germany: Kunstforum International, Apr. 2023, pp. 230–232, exhibition at Kunsthalle Gießen, 15 April – 30 June 2023. [Online]. Available: https://www.kunstforum.de/artikel/mary-audrey-ramirez/

[49] D. N. Ismail and K. Muhlen, "Mary-audrey ramirez: Forced amnesia 15.04.2023 – 30.06.2023," in *Pressinformation*. Kunsthalle Gießen, Apr. 2023, Online pdf, accessed: 3 July 2025. [Online]. Available: https://kunsthalle-giessen.de/wp-content/uploads/2023/05/mary-audrey-ramirez.forced-amnesia_kunsthalle-giessen-_en-1.pdf

[50] A. Mohné, "Low poly tree," 2020, accessed: 3 July 2025. [Online]. Available: https://www.achimmohn.net/low_poly-_tree-insel-grafenwerth-2020-23.html

[51] D. Kotovenko, O. Grebenkova, N. Sarafianos, A. Paliwal, P. Ma, O. Poursaeed, S. Mohan, Y. Fan, Y. Li, R. Ranjan, and B. Ommer, "Wast-3d: Wasserstein-2 distance for scene-to-scene stylization on 3d gaussians," 2024. [Online]. Available: https://arxiv.org/abs/2409.17917

[52] H. Baatz, J. Granskog, M. Papas, F. Rousselle, and J. Novák, "Nerf-tex: Neural reflectance field textures," *Computer Graphics Forum*, vol. 41, no. 6, pp. 287–301, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.14449

[53] T. Thonat, F. Beaune, X. Sun, N. A. Carr, and T. Boubekeur, "Tessellation-free displacement mapping for ray tracing," *ACM Transactions on Graphics (TOG)*, vol. 40, pp. 1 – 16, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:245015631

[54] H. Jung, S. Nam, N. Sarafianos, S. Yoo, A. Sorkine-Hornung, and R. Ranjan, "Geometry transfer for stylizing radiance fields," *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8565–8575, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:267365398

[55] M. Segu, M. Grinvald, R. Y. Siegwart, and F. Tombari, "3dsnet: Unsupervised shape-to-shape 3d style transfer," *ArXiv*, vol. abs/2011.13388, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:227208689

[56] R. Wang, G. Que, S. Chen, X. Li, J. Y. Li, and J. Yang, "Creative birds: Self-supervised single-view 3d style transfer," *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8741–8750, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:260164875

[57] H.-P. Huang, H.-Y. Tseng, S. Saini, M. K. Singh, and M.-H. Yang, "Learning to stylize novel views," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13 849–13 858, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:235248243

[58] Y. Li, H. yu Chen, E. Larionov, N. Sarafianos, W. Matusik, and T. Stuyck, "Diffavatar: Simulation-ready garment optimization with differentiable simulation," 2024. [Online]. Available: https://arxiv.org/abs/2311.12194

[59] N. Sarafianos, T. Stuyck, X. Xiang, Y. Li, J. Popovic, and R. Ranjan, "Garment3dgen: 3d garment stylization and texture generation," 2024. [Online]. Available: https://arxiv.org/abs/2403.18816

[60] M. Ye, M. Danelljan, F. Yu, and L. Ke, "Gaussian grouping: Segment and edit anything in 3d scenes," 2024. [Online]. Available: https://arxiv.org/abs/2312.00732

[61] Y. Zhang, Z. He, J. Xing, X. Yao, and J. Jia, "Ref-npr: Reference-based non-photorealistic radiance fields for controllable scene stylization," 2023. [Online]. Available: https://arxiv.org/abs/2212.02766

[62] J. Chen, B. Ji, Z. Zhang, T. Chu, Z. Zuo, L. Zhao, W. Xing, and D. Lu, "Testnerf: Text-driven 3d style transfer via cross-modal learning," in *International Joint Conference on Artificial Intelligence*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:260858361