ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ARTIFICIAL INTELLIGENCE

MASTER THESIS

in

Deep Learning

GENERATIVE AI IN ARTISTIC STYLE TRANSFER: PERFORMANCE, PERCEPTION, AND EVALUATION

CANDIDATE

Razvan Ciprian Stricescu

SUPERVISOR

Prof. Andrea Asperti

Academic year 2024-2025

Session 5th

”Life before death. Strength before weakness. Journey before destination. That was their motto, and was the First Ideal of the Immortal Words.”

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Generative Artificial Intelligence (AI) has profoundly transformed the landscape of digital creativity, enabling the automated generation of art, images, and visual compositions. Advances in deep learning have led to the development of sophisticated tools that can replicate and transfer artistic styles with remarkable fidelity [13, 39]. Techniques such as Generative Adversarial Networks (GANs) [22], diffusion models [17], and transformer-based architectures [11] have significantly expanded the boundaries of artistic synthesis.

Among these innovations, style transfer has emerged as a pivotal technique for altering the aesthetic characteristics of an image while preserving its core structure. Early work by Gatys et al. [13] introduced convolutional neural networks (CNNs) capable of extracting and reapplying artistic textures. More recently, models like **StyleGAN** [22], **Stable Diffusion** [40], and **DALL·E** [39] have enhanced the precision and versatility of style transfer through advances in latent-space manipulation and multimodal conditioning.

At the same time, the growing capabilities of Large Language Models (LLMs) have sparked significant research interest regarding their emergent abilities. LLMs exhibit proficiency in natural language understanding and generation, leading to discussions on their potential role in evaluating and producing creative works. Recent studies suggest that beyond text processing, these models can engage with aesthetic criteria, analyze artistic content, and

generate novel interpretations of creative artifacts [46, 21]. A key challenge in this area is the inherent sensorial limitation of LLMs. Unlike humans, who experience art through direct sensory perception, AI models process text-based descriptions and encoded visual information [7]. This raises critical questions about the depth and authenticity of their aesthetic judgments. While AI can replicate stylistic elements with high fidelity, its ability to internalize and critically evaluate artistic principles remains underexplored.

Additionally, the multimodal capabilities of AI systems play a crucial role in aesthetic assessment. Current generative models integrate textual prompts with image generation, yet the extent to which they can self-assess and refine their outputs in alignment with aesthetic principles remains largely unknown [50]. Future advancements in AI aesthetics will likely hinge on improved multimodal integration, allowing models to process and critique visual, auditory, and textual content holistically.

Another major consideration is the role of biases, hallucinations, and deep fakes in AI-generated art. Since generative models learn from extensive datasets, they inherently reflect the biases present in their training data [4]. This raises ethical concerns regarding content authenticity and misinformation. At the same time, LLMs occasionally produce hallucinatory outputs statements or images that appear plausible but lack grounding in real-world artistic traditions. Addressing these challenges requires balancing ethical oversight with the preservation of AI's generative creativity. Policy interventions, digital traceability, and industry responsibility will be essential in mitigating these risks while fostering a diverse and pluralistic AI-driven creative landscape.

This thesis critically examines the intersection of generative AI and artistic style transfer, assessing its technological foundations, aesthetic implications, and avenues for further refinement. Building on our previous work [3], which explored the capabilities and limitations of generative models in replicating artistic styles, this study further evaluates leading models' ability to adhere to artistic conventions, interpret stylistic nuances, and maintain authenticity. By

doing so, it aims to contribute to the broader discourse on AI's role in creative industries and its potential to enhance digital artistry.

1.1 Problem statement

The rapid advancements in generative artificial AI have significantly transformed digital creativity, particularly in the domain of artistic style transfer. State-of-the-art generative models, including diffusion-based architectures and GANs, have demonstrated remarkable capabilities in synthesizing images that emulate traditional artistic styles. However, these models face persistent challenges concerning authenticity, fidelity, and compositional balance.

Despite achieving high aesthetic quality, AI-generated artworks often suffer from limitations such as hyperrealistic distortions, misinterpretation of stylistic elements, and an inability to maintain structural coherence in complex compositions. Public perception studies indicate that while AI can effectively reproduce impressionist and abstract styles, it struggles with historically intricate styles such as baroque or high renaissance, where fine-grained detail, anatomical precision, and depth perception are crucial. Moreover, the extent to which current models adhere to user prompts and preserve the integrity of artistic conventions remains an open research question.

The purpose of this project is to investigate the extent to which these models can authentically replicate artistic styles while maintaining compositional and structural integrity. Furthermore, it explores the role of user perception studies and survey-based evaluations in determining the authenticity and effectiveness of AI-generated art. A key contribution of this research is the creation of a supervised dataset of AI-generated artworks, systematically labeled and analyzed to assess model performance across different artistic styles and compositions. By systematically evaluating these aspects, this research seeks to contribute to the broader discourse on AI's role in digital artistry and its

potential for creative enhancement.

1.2 Objective

The primary objective of this thesis is to conduct a comprehensive evaluation of generative AI models for artistic style transfer, focusing on both technical and aesthetic dimensions. Specifically, the study aims to:

- **Assess the fidelity of AI-generated artworks:** Evaluate how well generative models adhere to traditional artistic styles, considering various details such as structural coherence, color palette accuracy, and brushstroke emulation.
- **Develop and analyze an AI-generated artwork dataset:** Construct a structured dataset containing AI-generated images labeled by style, period and subjects, providing a foundation for further research in AI-assisted artistry.
- **Analyze prompt adherence and interpretability:** Investigate how effectively AI models translate textual prompts into visually faithful artistic compositions, ensuring alignment with user-specified stylistic requirements.
- **Identify common distortions and artifacts:** Examine the recurring inaccuracies in AI-generated images, including anatomical inconsistencies, hyperrealism, and contextual misinterpretations, which hinder their authenticity.
- **Compare leading generative models:** Conduct a comparative study of different generative models, both text-to-image and image-to-image to determine their relative strengths and weaknesses in artistic style transfer.

- **Evaluate public perception through surveys:** Utilize survey-based assessments to measure how convincingly AI-generated images replicate human-created art and how audiences perceive their authenticity and aesthetic value.
- **Contribute to the discourse on AI in creative industries:** Provide insights into the practical applications of AI-assisted artistry, addressing its potential benefits, ethical implications, and limitations in professional creative workflows.

1.3 Motivation and Significance of the study

The intersection of artificial intelligence and digital art has introduced unprecedented possibilities in creative expression, enabling both professional artists and hobbyists to explore new artistic frontiers. Generative models have not only automated style transfer but have also enhanced artistic workflows, reducing the time and effort required for traditional manual processes [36]. However, the use of AI in artistic creation remains a subject of debate due to concerns about authenticity, originality, and the ethical implications of AI-generated art.

Despite their advancements, current generative AI models still exhibit challenges in maintaining fidelity to artistic conventions. Studies have shown that while AI can replicate certain art styles with high accuracy, it struggles with complex compositions requiring a deep understanding of artistic techniques [55]. Furthermore, emerging research suggests that AI-generated images often fail to preserve fine stylistic details and introduce distortions that make them distinguishable from human-made artwork [54].

Another crucial aspect is the public perception of AI-generated art. Previous studies have demonstrated that viewers can often distinguish between AI-generated and human-created artwork based on stylistic inconsistencies and unnatural patterns [48]. This raises questions about whether generative

AI can truly be considered a tool for artistic augmentation or if it risks diminishing the value of traditional human artistry. By conducting perception-based surveys, this thesis seeks to address these concerns and evaluate the extent to which AI-generated artwork aligns with human artistic expectations.

In addition to perception studies, a major contribution of this research is the development of a structured dataset of AI-generated images labeled across multiple criteria. This dataset serves as a critical tool for evaluating AI-generated artwork in a standardized manner, allowing for a deeper analysis of model performance in relation to artistic fidelity and compositional accuracy. By providing a high-quality dataset of AI-generated images, this study contributes to the broader field of AI research, supporting further investigations into deepfake detection, style adaptation, and automated art critique.

Furthermore, the increasing use of AI in creative industries necessitates a thorough understanding of its capabilities and limitations. AI-generated art is being integrated into advertising, gaming, and digital media, making it essential to assess its reliability and impact on human artists [8]. Understanding the technical and aesthetic challenges of AI-driven style transfer will contribute to developing improved models that better align with artistic principles and user expectations.

This study is motivated by the need to bridge the gap between AI's computational efficiency and the artistic intricacies of traditional styles. By critically evaluating state-of-the-art generative models and their effectiveness in style transfer, this research aims to provide valuable insights into how AI can be refined to support and enhance creative practices while maintaining artistic authenticity.

1.4 Methodological Approach

This study employs a multi-faceted methodological framework to analyze the capabilities and limitations of generative AI in artistic style transfer. The

methodology consists of five primary components:

1.4.1 AI-Pastiche Dataset: Overview and Structure

The **AI-Pastiche Dataset** is a structured collection of **953 AI-generated artworks**, produced using **twelve different models**: DALL-E 3, Stable Diffusion 1.5, Stable Diffusion 3.5 Large, Flux 1.1 Pro, Flux 1 Schnell, OmniGen, Ideogram, Kolors 1.5, Firefly Image 3, Leonardo Phoenix, Midjourney V6.1, and Auto-Aesthetics v1. This dataset provides a diverse range of AI-generated

images, offering a foundation for evaluating the performance of different generative models in artistic style replication.

Each image in the dataset was generated based on **73 structured prompts**, designed to cover a wide range of artistic styles and historical periods. The dataset is intended to facilitate a comprehensive evaluation of AI-generated art, focusing on aspects such as stylistic fidelity, compositional integrity, and

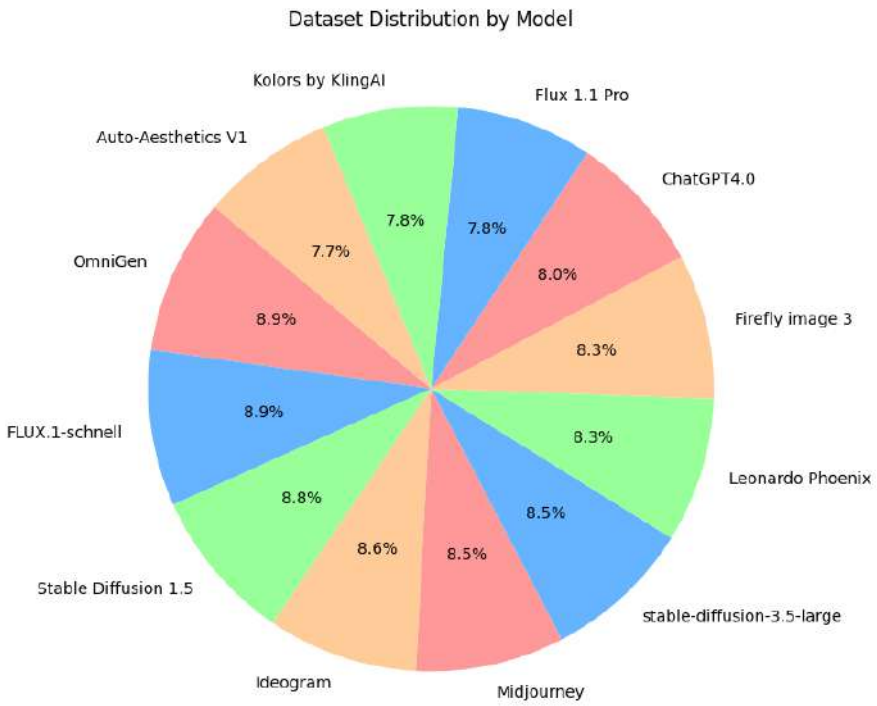


Figure 1.1: Dataset Distribution by Model. This pie chart illustrates the proportion of images generated by each AI model.

adherence to user prompts. The key metadata attributes associated with each image are detailed in Table 1.1.

(a) Stable-Diffusion-3.5 (b) Flux 1.1 -Schnell (c) Midjourney

Figure 1.2: Examples of images generated with the same prompt on different models.

The dataset includes a variety of subjects, such as landscapes, portraits, abstract compositions, and human figures. Understanding the distribution of these subjects helps in identifying biases and trends in AI-generated artworks (Figure 1.3).

Attribute	Description
Generative Model Used	Identifying which AI model produced the image.
Subject Matter	Categories such as landscape, cityscape, animals, abstract compositions, and human figures.
Artistic Style	The specific artistic movement or tradition the image aims to replicate (e.g., Impressionism, Surrealism, Baroque).
Historical Period	Categorizing the style's origin (e.g., Renaissance, 19th-century realism, modernist movements).
Prompt Structure	The textual prompt used to generate the image.
Generated Image Filename	A unique identifier for each image.

Table 1.1: Dataset Metadata Attributes

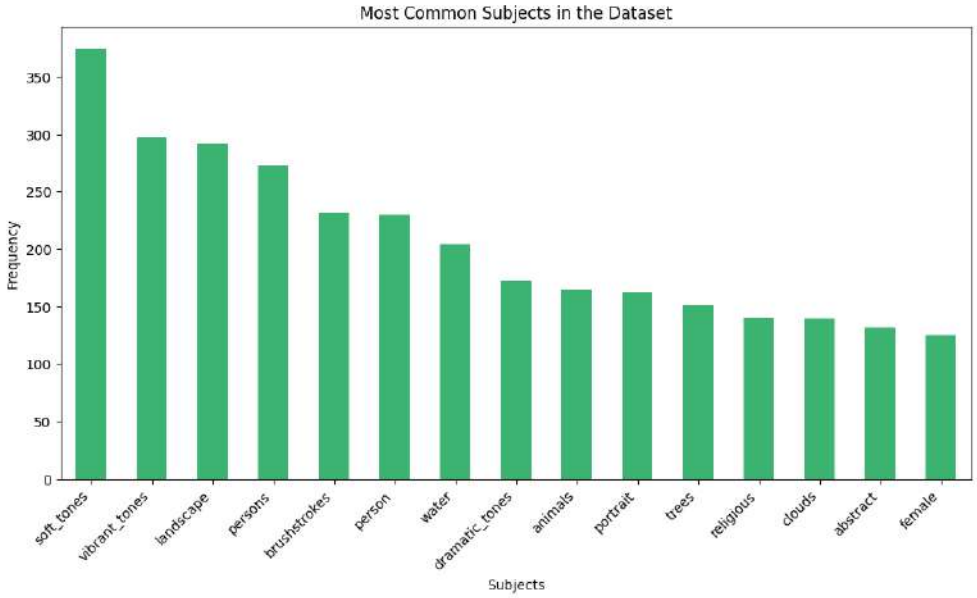


Figure 1.3: Most Common Subjects in the Dataset. The bar chart highlights the frequency of subjects such as landscapes, brushstrokes, and persons.

Each image is also labeled according to its intended artistic movement, ranging from Impressionism to Baroque, as depicted in Figure 1.4.

The AI-Pastiche Dataset serves as a structured benchmark for evaluating generative models in artistic style transfer. By analyzing the dataset, it is possible to assess not only the visual quality of AI-generated images but also their alignment with stylistic conventions and historical references. The dataset provides a basis for identifying common challenges in AI-generated art, such as over-polished textures, inconsistencies in human figures, and deviations from historical accuracy. These insights contribute to a deeper understanding

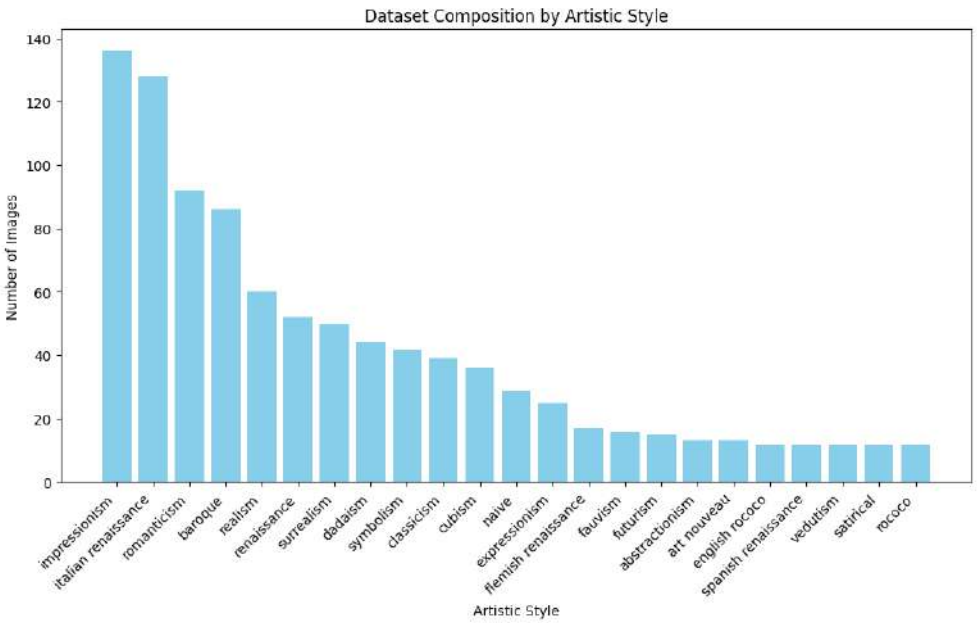


Figure 1.4: Dataset Composition by Artistic Style. The histogram displays the number of images corresponding to different artistic movements.

(a) Surrealism (b) Vedutism (c) Fauvism

Figure 1.5: Examples of images generated with different artistic movements

of the capabilities and limitations of current generative AI models.

1.4.2 Comparative Model Analysis

The study systematically compares the performance of multiple generative models. These models vary in architecture, training approach, and output control mechanisms. To ensure consistency, the same set of prompts was used across all models.

Performance was evaluated using a combination of quantitative and qualitative measures, including:

- **Style Accuracy** – The extent to which the generated image matches the intended artistic style.
- **Structural Integrity** – The ability of the AI model to maintain compositional balance and avoid distortions.
- **Prompt Fidelity** – How accurately the model interprets and executes the given prompt.

1.4.3 Survey-Based User Perception Study

To assess how well AI-generated images align with human artistic expectations, a large-scale public perception survey was conducted with **approximately 600 participants**. The survey was structured as follows:

- Participants were shown a mix of **AI-generated and human-created artworks** in a randomized order.
- They were asked to classify each image as either **AI-generated or human-made**.
- Additionally, a restricted number of people rated the artistic fidelity of each image in relation to its generative prompt.

The results from this survey provided insights into:

1. The detectability of AI-generated images.
2. Which styles are more convincingly replicated by AI.
3. The level of acceptance or skepticism toward AI-generated artworks.

1.4.4 Artifact Detection and Qualitative Assessment

A qualitative review was conducted to identify common distortions and artifacts in AI-generated images. The main issues observed included:

- **Hyperrealism or unnatural textures** – Particularly in impressionistic and surrealist styles.
- **Inconsistent anatomy and proportions** – Notably in human and animal representations.
- **Loss of stylistic coherence** – Where elements of different styles merged unintentionally.
- **Color inconsistencies** – Some models introduced unrealistic saturation or hues not representative of the intended style.

1.4.5 Performance Metrics and Statistical Analysis

To provide a robust evaluation, several quantitative performance metrics were used:

- **Perceptual Error Rate** – The percentage of AI-generated images mistaken for human-created works.
- **Prompt Fidelity Score** – A rating scale to quantify how well an image aligns with the provided prompt.

Statistical tests, were conducted to determine significant differences between the models' outputs and the survey results.

1.4.6 Enhancements Based on Dataset Metadata

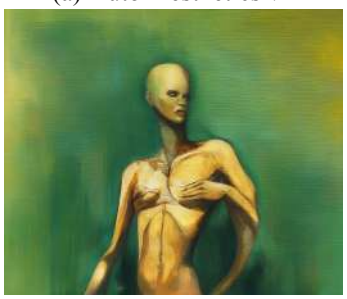
The dataset includes a broad spectrum of artistic subjects, covering landscapes, cityscapes, abstract forms, and portraiture. Specific insights derived from the dataset include:

- **Historical Styles Covered:** Flemish Renaissance, Vedutism, Dadaism, Impressionism, Realism, Symbolism, and more.
- **Diversity in Subject Matter:** The dataset contains a balanced mix of elements such as **human figures, nature, urban environments, and conceptual themes**.
- **AI Model Biases:** Some models consistently performed better in specific styles (e.g., Midjourney excelled in surrealism, while Stable Diffusion performed well in impressionistic renderings).

(a) Auto-Aesthetics v1

(b) Midjourney

(c) Kolors by KlingAI



(d) OmniGen

(e) Flux 1.1 - Pro

(f) ChatGPT

Figure 1.6: Examples of diversity in Subject Matter.

Chapter 2

Generative Models

2.1 Comparison of techniques

Generative models have significantly evolved in recent years, leading to remarkable advancements in artistic style transfer and image synthesis. This section provides a comparative analysis of the primary generative modeling approaches used in artistic applications, with a focus on text-to-image and image-to-image models.

2.1.1 Text-to-Image Models

Text-to-image models generate images based on natural language descriptions. These models employ deep neural networks that learn mappings between textual representations and visual data. The most prominent architectures include diffusion models, Generative Adversarial Networks (GANs), and transformer-based models.

Diffusion-Based Text-to-Image Models

Diffusion models operate by iteratively refining noise to generate an image. The process is based on a denoising score matching (DSM) approach, where

an image $\mathbf{x}_0$ is gradually transformed into Gaussian noise $\mathbf{x}_T$, and then reconstructed using learned transition probabilities [17, 42].

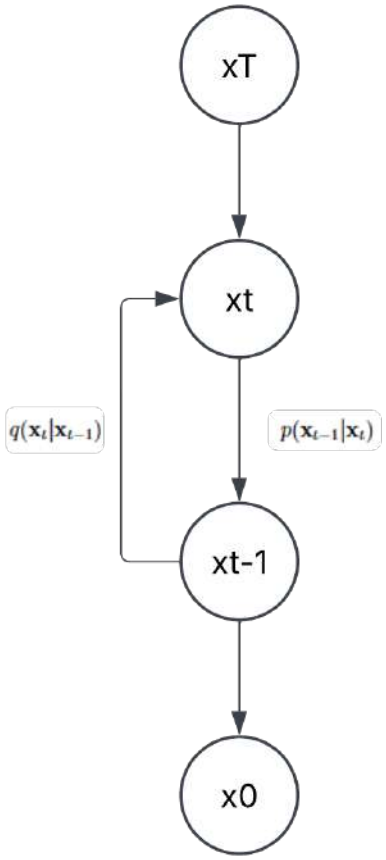


Figure 2.1: Simple overview of diffusion models structure

Formally, the forward diffusion process is defined as:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)\mathbf{I}) \tag{2.1}$$

where α_t is the noise schedule parameter controlling the amount of added noise at each timestep t. The model then learns a reverse process:[9]

$$p(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)) \tag{2.2}$$

where μ_θ and Σ_θ are the learned mean and variance parameters. Text conditioning is achieved through classifier-free guidance (CFG), in which the denoising function [32] is modified as:

$$\hat{\epsilon}_\theta(\mathbf{x}_t, y) = \epsilon_\theta(\mathbf{x}_t) + w(\epsilon_\theta(\mathbf{x}_t, y) - \epsilon_\theta(\mathbf{x}_t)) \tag{2.3}$$

where w is a guidance scale hyperparameter and y represents the textual prompt.

GAN-Based Text-to-Image Models

GAN-based text-to-image models employ a generator G that synthesizes an image from noise and a discriminator D that evaluates its realism [15]. The generator takes a latent noise vector z from a prior distribution p_z and produces an image, while the discriminator distinguishes real from generated images.

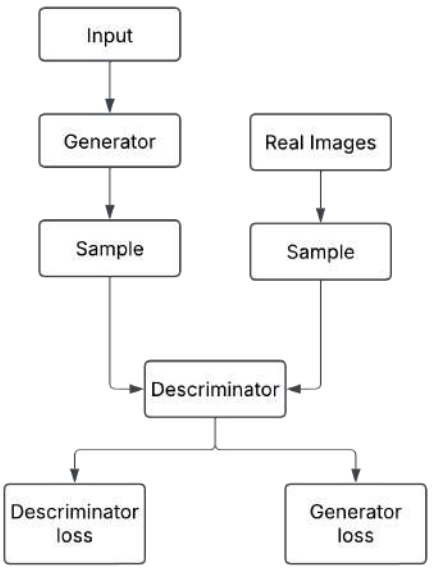


Figure 2.2: Simple overview of GAN-based models

The adversarial objective is:

$$\min_G \max_D E_{x \sim p_{data}}[\log D(x)] + E_{z \sim p_z}[\log(1 - D(G(z)))] \tag{2.4}$$

Conditional GANs (cGANs) extend this by conditioning both G and D on additional information y [30].:

$$G(z, y) \rightarrow \text{generated image}, \quad D(x, y) \rightarrow \text{probability of real vs. fake} \quad (2.5)$$

Training progresses as G improves in generating realistic images, while D refines its discrimination ability.

Transformer-Based Text-to-Image Models

Transformer-based models use attention mechanisms to directly map text to images. They employ Vision-Language Pretraining (VLP) techniques, such as Contrastive Language-Image Pretraining (CLIP), which learn to align text embeddings $E(y)$ with visual embeddings $E(x)$ [38].

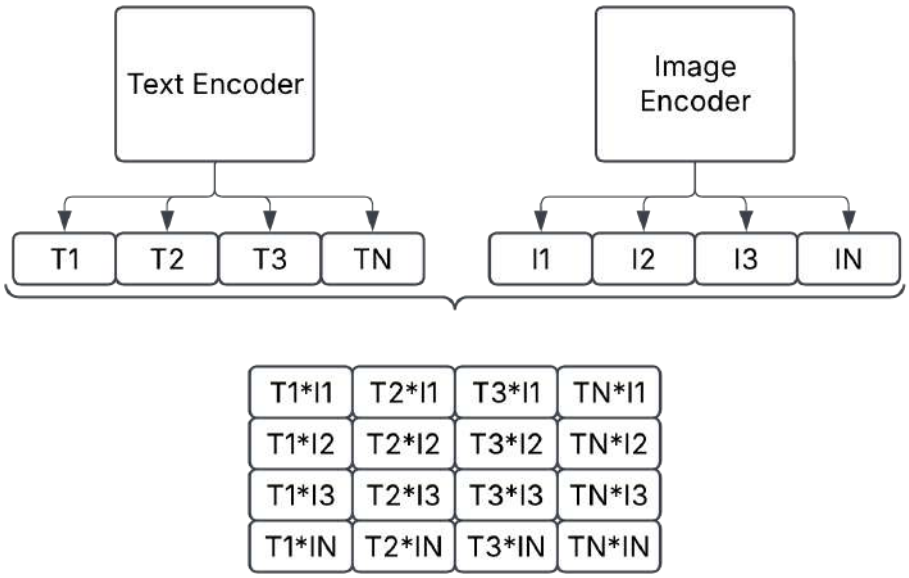


Figure 2.3: Simple overview of Transformer-based models

The optimization objective is:

$$\mathcal{L} = -\sum_{i} \log \frac{\exp(\cos(E(x_i), E(y_i)))}{\sum_j \exp(\cos(E(x_i), E(y_j)))} \tag{2.6}$$

where cos is the cosine similarity between embeddings.

2.1.2 Comparison of Text-to-Image Models

Table 2.1 presents a comparative overview of these three major types of generative models: Generative Adversarial Networks (GANs), Diffusion Models, and Transformer-Based Models. These models differ in their underlying mechanisms, training complexity, speed, and interpretability, making them suitable for different applications.

Feature	GANs	Diffusion Models	Transformer-Based Models
Main Concept	Adversarial Training	Iterative Noise Reduction	Attention Mechanism
Training Complexity	High	Very High	High
Generation Speed	Fast	Slow	Moderate
Image Quality	Sharp but can have artifacts	High Fidelity	Contextually Strong
Interpretability	Low	Moderate	High
Common Models	StyleGAN, BigGAN	Stable Diffusion, Imagen	DALL·E, Parti, Muse
Best Use Cases	Realistic Faces, Art	High-quality Images	Text-Image Mapping

Table 2.1: Comparison of Generative Models for Text-to-Image Models

2.1.3 Image-to-Image Models

Image-to-image models take an input image and generate a modified version, either transferring style, enhancing details, or generating variations [19]. These models include CycleGANs, style transfer networks, and diffusion-based architectures.

CycleGANs for Image-to-Image Translation

CycleGANs enable unpaired image-to-image translation by learning a mapping between two domains X and Y without requiring direct pixel-wise correspondence [56].

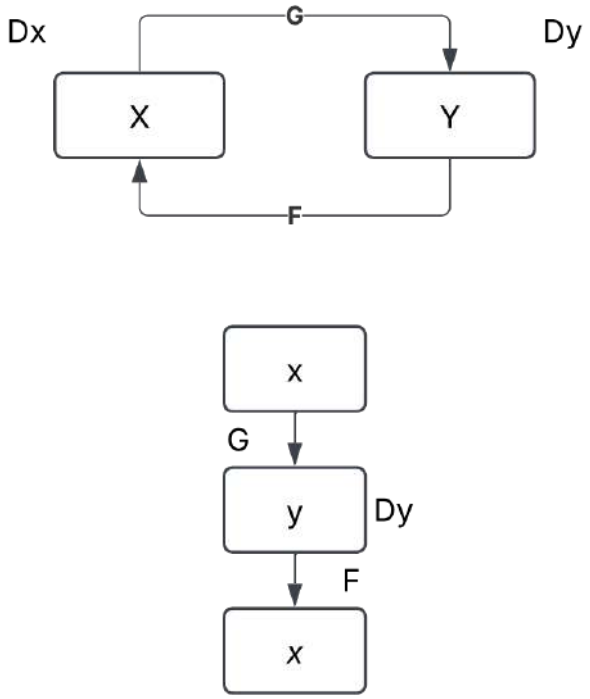


Figure 2.4: Simple overview of CycleGAN-based models

The objective consists of two adversarial losses:

$$\mathcal{L}_{\text{GAN}}(G, D_Y) = E_{y \sim p_Y}[\log D_Y(y)] + E_{x \sim p_X}[\log(1 - D_Y(G(x)))] \quad (2.7)$$

and a cycle-consistency loss that enforces $G(F(y)) \approx y$:

$$\mathcal{L}_{\text{cycle}}(G, F) = E_{x \sim p_X}[||F(G(x)) - x||_1] + E_{y \sim p_Y}[||G(F(y)) - y||_1] \quad (2.8)$$

where $G : X \to Y$ and $F : Y \to X$ are learned mappings, and D_X, D_Y are domain-specific discriminators.

Neural Style Transfer

Neural Style Transfer (NST) applies the artistic features of one image to another by optimizing a content image I_c to match the style of a reference image I_s [14].

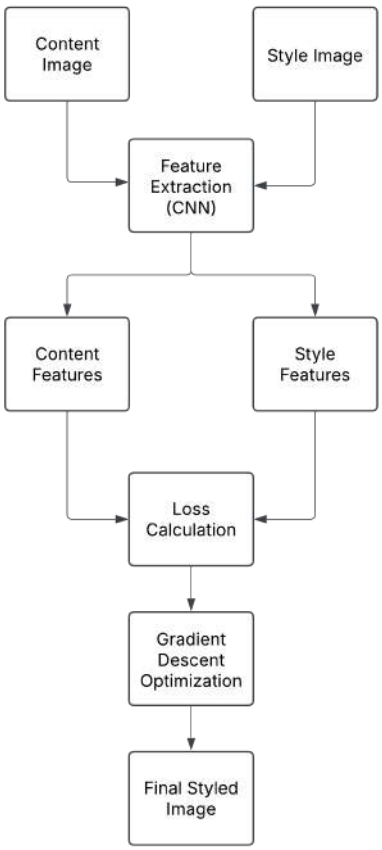


Figure 2.5: Simple overview of Neural Style Transfer-based models

The objective function is:

$$\mathcal{L} = \alpha \mathcal{L}_{\text{content}}(I_c, I_g) + \beta \mathcal{L}_{\text{style}}(I_s, I_g) \tag{2.9}$$

where I_g is the generated image, and the losses are computed as:

$$\mathcal{L}_{\text{content}} = ||\phi_l(I_c) - \phi_l(I_g)||_2^2 \tag{2.10}$$

$$\mathcal{L}_{\text{style}} = \sum_{l} ||G^l_\phi(I_s) - G^l_\phi(I_g)||^2_2 \tag{2.11}$$

where ϕ_l is the feature map at layer l, and G^l_ϕ represents the Gram matrix of features [20].

2.1.4 Comparison of Image-to-Image Models

Image-to-image models enable the transformation of one image into another, allowing for tasks such as style transfer, super-resolution, and domain adaptation. Unlike text-to-image models, which generate images from textual prompts, image-to-image models modify an existing image while preserving its structure. Table 2.2 provides a comparative overview of key image-to-image generative models, focusing on their underlying mechanisms, strengths, and best use cases.

Feature	CycleGAN	Neural Style Transfer (NST)	Diffusion-Based Models
Main Concept	Domain Translation	Artistic Style Transfer	Iterative Denoising
Training Complexity	High	Low	Very High
Generation Speed	Fast	Real-time	Slow
Image Fidelity	Moderate	High	Very High
Need for Paired Data	No	Yes (Style Reference)	No
Common Models	CycleGAN, pix2pix	Gatys et al. CNNs	Stable Diffusion, Imagen
Best Use Cases	Photo-to-Painting, Domain Adaptation	Artistic Filters	Image Inpainting, High-Quality Edits

Table 2.2: Comparison of Image-to-Image Generative Models

This comparison highlights key differences in how these models function. CycleGAN is well-suited for unpaired image-to-image translation, making it useful for tasks like converting photographs into paintings. Neural Style Transfer (NST) provides fast and artistic transformations but relies on a predefined style reference. Meanwhile, diffusion-based models deliver exceptional detail and fidelity at the cost of high computational requirements. Understanding these differences helps in selecting the most appropriate model for various artistic and real-world applications.

2.2 Deep dive into specific models

2.2.1 Omnigen: Unified Image Generation

Omnigen is a multimodal generative model based on a latent diffusion framework. Unlike traditional diffusion models, Omnigen integrates cross-modal attention mechanisms that allow it to process text, image, and style embeddings simultaneously. The model operates in a structured latent space, where the denoising process is guided by semantic embeddings rather than raw pixel-based information [47].

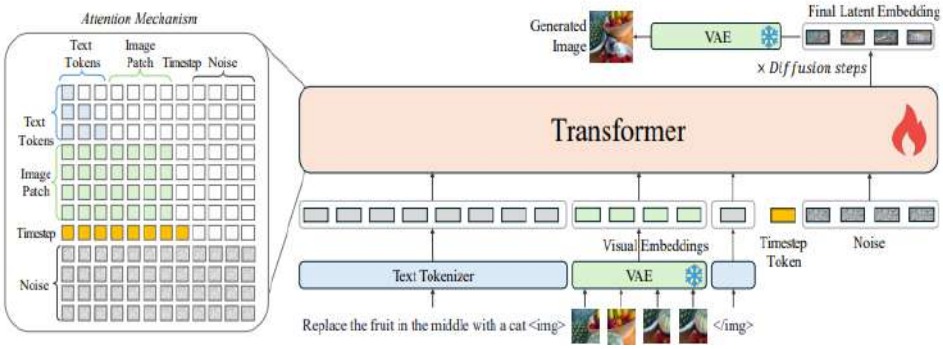


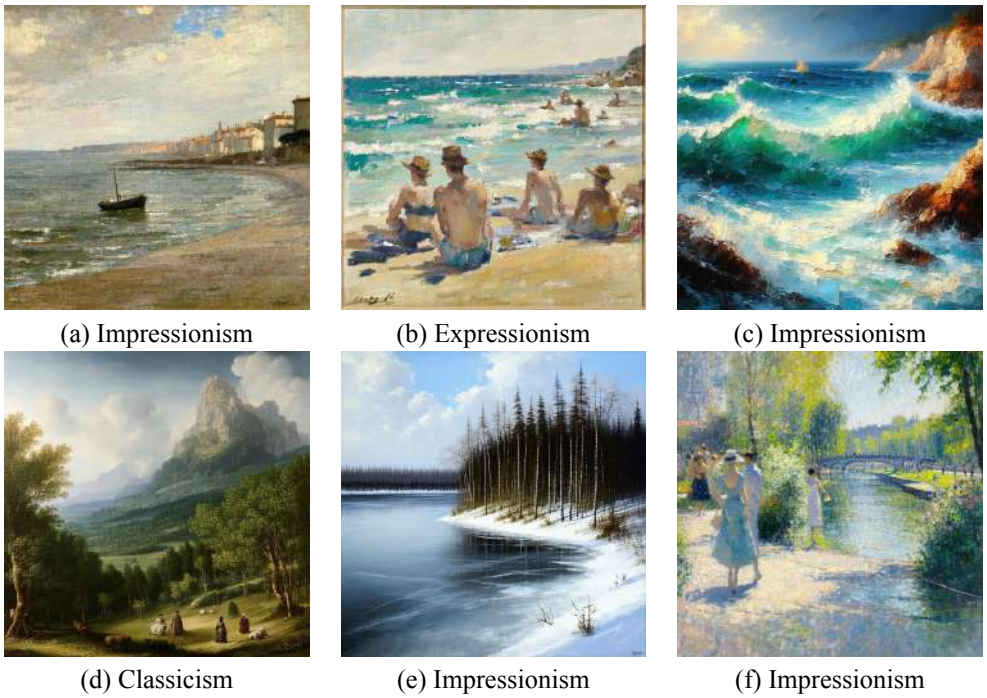
Figure 2.6: Omnigen model framework. Image adapted from [47].

The architecture is built upon a Transformer-ViT hybrid backbone, which enables improved context retention across multiple conditioning inputs. Additionally, it leverages a contrastive embedding alignment strategy that ensures text and image representations align efficiently, reducing semantic drift in output generation. The introduction of adaptive denoising schedules in Omnigen significantly improves convergence efficiency while maintaining image integrity during progressive noise removal.

(a) Flemish Renaissance (b) Dadaism (c) Romanticism

(d) Expressionism (e) Art Nouveau (f) Romanticism

Figure 2.7: Examples of artworks created with OmniGen.



(a) Impressionism (b) Expressionism (c) Impressionism

(d) Classicism (e) Impressionism (f) Impressionism

Figure 2.8: Examples of artworks created with Stable Diffusion 3.5 Large.

2.2.2 Stable Diffusion 3.5 Large: Rectified Flow Diffusion with Multi-Scale Conditioning

Stable Diffusion 3.5 Large refines the traditional diffusion model by incorporating Rectified Flow Transformers, an innovation that allows diffusion processes to be smoothly interpolated along a learned trajectory rather than iteratively removing noise through stochastic denoising steps [12]. This approach significantly reduces computational overhead while improving sample consistency.

The model utilizes Multi-Scale Hierarchical Conditioning, where high-level semantic information (e.g., scene composition) is introduced at earlier denoising steps, while low-level details (e.g., texture fidelity) are reinforced at later stages [55]. Unlike prior versions, Stable Diffusion 3.5 Large applies Query-Key Normalization (QK Norm) to its self-attention layers, mitigating over-amplification of dominant visual features, thus improving text-image coherence in complex prompts [16].

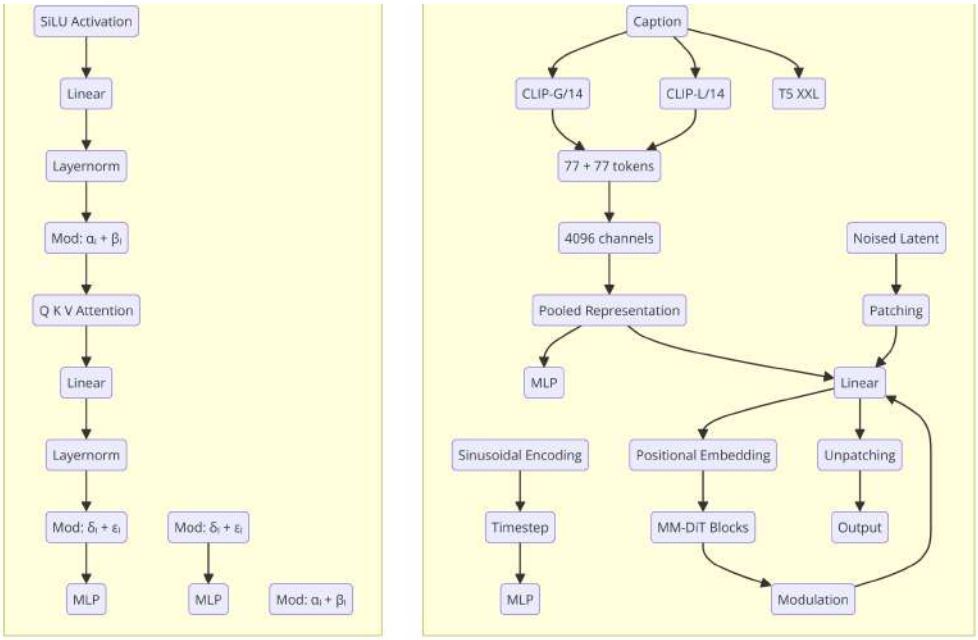
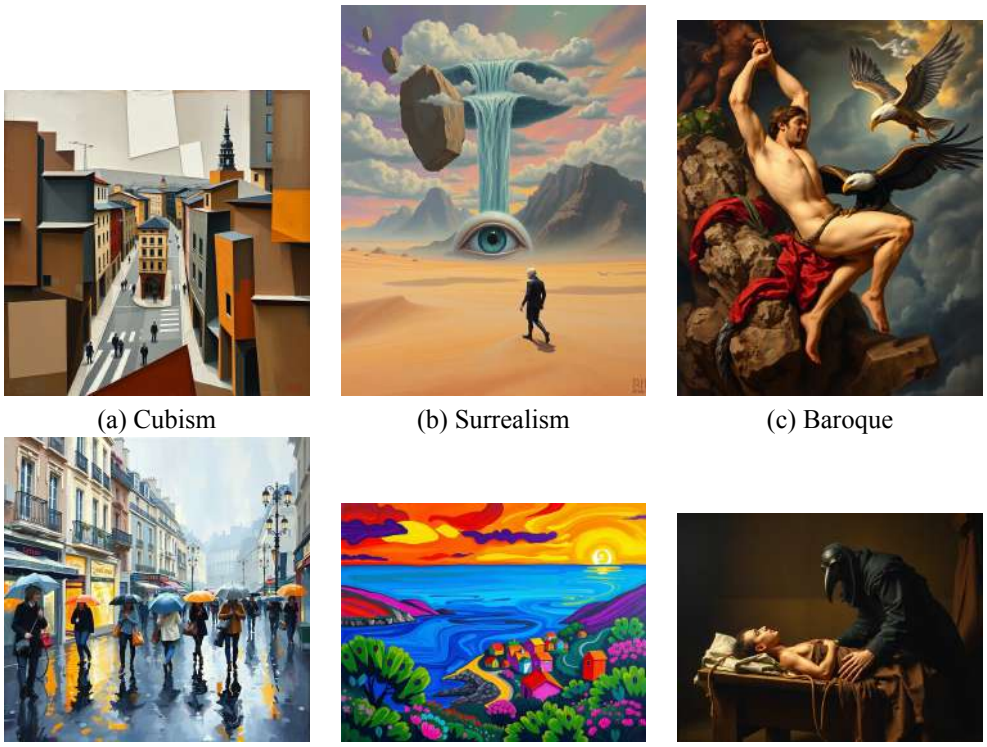


Figure 2.9: Stable Diffusion 3.5 Large simplified model framework. Image re-adapted from [12].

Another key advancement in this model is the integration of LoRA-adapted fine-tuning, which allows for efficient adaptation to new artistic styles without full model retraining [54]. By compressing style-based learnable parameters, the model enhances domain adaptability without requiring massive computational resources.

2.2.3 Flux Schnell: Adversarially Guided Latent Diffusion



(a) Cubism (b) Surrealism (c) Baroque

(d) Impressionism (e) Fauvism (f) Baroque

Figure 2.10: Examples of artworks created with Flux Schnell.

Flux Schnell differs from conventional diffusion models by incorporating an adversarial latent refinement module. Instead of solely relying on a Markovian denoising process, it employs a latent adversarial discriminator that evaluates intermediate diffusion steps for realism, consistency, and perceptual fidelity [24].

A distinctive feature of Flux Schnell is its Turbo Image-to-Image Translation Pipeline, which implements a one-step latent warping technique rather

than performing gradual denoising over multiple iterations. This method is optimized for real-time applications, particularly in scenarios requiring rapid artistic transformations [35].

Furthermore, Flux Schnell utilizes a conformal latent mapping function, ensuring that stylistic adjustments do not distort underlying structural elements. The incorporation of a self-normalizing adversarial loss function allows for improved generalization across varied artistic domains while reducing high-frequency noise artifacts.

2.2.4 Pix2Pix: Conditional GAN for Image Translation

Pix2Pix follows a paired adversarial learning paradigm, where a U-Net-based generator translates input images into a desired style while a PatchGAN discriminator evaluates local-level coherence [5]. Unlike diffusion models, which progressively refine images through noise removal, Pix2Pix operates via a direct mapping approach based on pixel-level transformations.

A crucial improvement in modern Pix2Pix adaptations is the integration of self-attention layers, which enhance context retention in complex transformations. Additionally, style injection modules have been introduced, allowing the generator to modulate feature maps dynamically based on external style conditioning signals [8].

Despite its speed advantages over diffusion-based approaches, Pix2Pix remains constrained by paired training dependencies, requiring extensive manually labeled datasets to achieve high-fidelity style replication. However, recent advancements in semi-supervised adversarial training have begun addressing this limitation.

2.2.5 Stable Diffusion Variations: Fine-Tuned Adaptations and Specialized Architectures

Stable Diffusion Variations expand on the foundational principles of latent diffusion models by introducing adaptive fine-tuning strategies that improve stylistic consistency across multiple artistic domains. These variations primarily employ LoRA-based parameter adaptation, where lightweight fine-tuning layers adjust internal representations while preserving core diffusion dynamics [54].

Notable enhancements in these variations include:

- Feature-guided Style Transfer (FeaST), enabling finer control over local stylistic attributes without compromising content structure [36].
- Step-aware Prompt Conditioning, where the generative process dynamically adjusts text-based embeddings across different denoising stages [55].
- Extended Resolution Support, allowing high-resolution generations beyond the conventional limitations of earlier diffusion models.

2.3 Definition of Technical capabilities

The technical capabilities of modern generative models in artistic style transfer have significantly evolved due to advancements in deep learning architectures and optimization techniques. This section assesses these capabilities across different dimensions, including fidelity to source styles, structural integrity, prompt adherence, and computational efficiency.

2.3.1 Fidelity to Source Styles

Fidelity to source styles is a crucial measure of how well a generative model replicates the defining characteristics of a given artistic movement. State-of-the-art models such as Stable Diffusion 3.5 Large, Omnigen, and Flux Schnell have demonstrated remarkable proficiency in mimicking impressionist, cubist, and surrealist styles [36]. However, more intricate styles, such as Baroque or High Renaissance, often pose challenges due to their fine-grained details and anatomical precision. Diffusion models, leveraging iterative noise reduction, generally outperform GAN-based models in producing high-fidelity style transfer, particularly for abstract and expressionist forms [55].

2.3.2 Structural Integrity and Compositional Balance

Generative models must maintain structural integrity when applying artistic transformations. Issues such as distorted anatomy, unnatural lighting, and inconsistencies in brushstroke application frequently emerge, especially in complex scenes [54]. The assessment study indicates that models perform well in rendering simple compositions but struggle with preserving depth perception and object coherence in multi-subject artworks. Despite the improvements brought by transformer-based conditioning methods like CLIP, occasional artifacts, such as extra limbs or warped figures, persist in AI-generated images [8].

2.3.3 Control Over Transfer Intensity

A key feature in modern style transfer models is their ability to modulate the intensity of artistic transformation. Methods such as Low-Rank Adaptation (LoRA) and QA-LoRA offer enhanced fine-tuning capabilities, allowing users to dictate the extent to which a generated image adheres to a given style [49]. While some models provide extensive control over style blending, others, like DALL·E 3, tend to prioritize aesthetic coherence over strict adherence to the

provided reference style, occasionally leading to an over-saturation of certain stylistic elements [21].

2.3.4 Prompt Adherence and Semantic Understanding

The effectiveness of generative models is also evaluated based on their ability to accurately interpret and execute textual prompts. Transformer-based architectures, such as those incorporating vision-language pretraining (e.g., CLIP-guided diffusion), offer a superior understanding of nuanced stylistic directives [38]. However, despite advancements, inconsistencies remain in translating textual descriptions into visual outputs, particularly in cases where prompts specify subtle texture variations or complex lighting scenarios. Comparative results indicate that while models like Midjourney and Stable Diffusion excel in artistic adaptation, they sometimes diverge from the exact phrasing of the prompt, requiring multiple iterations for satisfactory results [50].

2.3.5 Performance and Computational Efficiency

The computational demands of generative models directly impact their accessibility and usability. Diffusion models, which require iterative denoising steps, generally entail higher computational costs compared to GAN-based models. However, optimizations such as latent diffusion and rectified flow transformers have mitigated some of these constraints, enabling high-resolution image synthesis with reduced inference time [18]. Additionally, proprietary models like Midjourney and Firefly Image 3 provide streamlined generation processes but often at the expense of transparency regarding their training methodologies [1].

2.3.6 Common Artifacts and Limitations

Despite their impressive capabilities, generative models exhibit recurring artifacts that hinder their authenticity. Key issues include:

- **Hyperrealism in Historical Styles**: AI-generated artworks occasionally introduce overly sharp details in styles that traditionally favor softer textures [44].

  Case Study: Figure 2.11 describes a Romanticism-style painting featuring three elegantly dressed figures in a pastoral setting. While the image successfully captures the essence of 19th-century portraiture, several hyperrealistic artifacts emerge, such as the facial features of the subjects are rendered with an excessive level of sharpness and smoothness, deviating from the characteristic soft brushstrokes of Romanticist paintings. Furthermore the skin tones appear unnaturally perfect, lacking the subtle textural variations typically seen in historical oil paintings. The lightning and shading are overly refined, giving the figures a photographic quality that contrasts starkly with the intended painterly aesthetic.

Figure 2.11: AI-generated image with Stable Diffusion 3.5 large in Romantic style.

- **Anatomical Distortions**: Hands, limbs, and facial features often suffer from irregular proportions, particularly in multi-figure compositions [51].

  Case Study: In figure 2.12, an Italian Renaissance-style painting generated by Flux Schnell, we can observe a seated Madonna and Child. While the image captures the characteristic color palette and composition associated with Renaissance masters, closer inspection reveals

multiple distortions: The Madonna's facial features appear asymmetrical and unnaturally smooth, lacking the subtle anatomical details that define human expression. The infant's limbs, particularly the arms and fingers, are disproportionate, with some fingers appearing fused or unnaturally bent. The Madonna's hands, particularly the way she holds the infant, exhibit awkward positioning and an unrealistic grasp, disrupting the natural interaction between figures.

Figure 2.12: AI-generated image with Omnigen in Italian Renaissance style.

- **Color and Lighting Inconsistencies**: Some models struggle with maintaining color harmony, leading to unnatural saturation or lighting mismatches [52].

  Case Study: A compelling example is a Romanticism-style landscape, generated by Flux Schnell, painting 2.13 featuring a lakeside cabin under a dramatic moonlit sky. We can observe that the light sources in the image appear conflicting. The moon emits an unnatural glow, creating an exaggerated halo effect, while the illuminated clouds suggest

an inconsistent light direction. The interior lighting of the cabin appears overly saturated and detached from the natural ambiance, disrupting the cohesion of the scene. And furthermore, the reflection of the moonlight on the water is unnaturally sharp and overly bright, lacking the organic diffusion expected in real-world reflections.

Figure 2.13: AI-generated image with Flux Schnell in Romantic style.

- **Anachronistic Elements**: Models sometimes insert objects or features inconsistent with the specified historical period, impacting authenticity [26].

  Case Study: A prime example of anachronistic errors can be observed in the image generated by Flux Schnell, visible in figure 2.14. The image presents a classical religious composition reminiscent of Renaissance paintings, featuring a woman holding an infant while traveling with a man on horseback. While the composition, attire, and lighting largely adhere to historical artistic conventions, a critical flaw emerges: the male figure wears a flat cap, a headpiece commonly associated with 19th and 20th-century European working-class attire. Such an element is historically inconsistent with the period implied by the rest of the image and disrupts the authenticity of the generated artwork.

Figure 2.14: AI-generated image with Flux Schnell in Italian Renaissance style.

2.4 Methodology

This study employs a comparative methodological approach to evaluate the performance and design of generative models systematically. By placing different models under controlled conditions and using uniform prompts, we enable a structured comparison across architectural differences and generative capabilities.

Examples of such prompts:

Prompt: ”A serene painting of an alpine pasture in the impressionism style of the second half of the 19th century. Brushstrokes must be rough and clearly visible. The painting features a small flock of sheep grazing, grouped together, in a wide, open landscape. The perspective is expansive, emphasizing the depth of the landscape. The horizon line is set high, with a thin line of distant snow-capped mountains, allowing for a broad view of the foreground and middle ground. The foreground consists of rocky terrain and sheep, while the middle ground stretches into rolling

fields. On the right, almost at the foreground, a shepherd sits calmly, observing the scene. The color palette is composed of soft, earthy tones—light greens, browns. There are no light-and-shadow effects. ”

Prompt ”A dramatic and vivid painting depicting Saint Jerome in his study, rendered in the style of Titian. Saint Jerome is portrayed as an elderly figure with a long white beard, seated at a desk in a contemplative pose. His surroundings are spacious, with a sense of depth and distance in the composition. He is dressed in rich, flowing robes of earthy reds and deep browns, evoking a sense of gravitas and humility. The desk is adorned with books, manuscripts, and an open Bible, while a prominent skull rests nearby, symbolizing mortality and the fleeting nature of life. The room is dimly lit, with light falling on Jerome’s aged face and hands, emphasizing his wisdom and spiritual intensity. In the background, shelves of books and a faint crucifix add depth, while the surrounding darkness focuses attention on the central figure. The composition highlights Titian’s mastery of dynamic poses, warm tones, and painterly textures, creating an atmosphere of profound reflection and divine connection. ”

We utilized the AI-Pastiche dataset to ensure a controlled evaluation environment. For each generative model, we generated images using predefined prompts spanning various historical artistic styles. Each prompt was tested across multiple runs to assess intra-model consistency and variation in outputs.

Secondly, we implemented a human-in-the-loop evaluation framework. A panel of people reviewed the generated images, scoring them based on stylistic fidelity, compositional coherence, and perceptual authenticity. The evaluation followed a double-blind protocol, where evaluators were unaware of the model responsible for generating each image, ensuring unbiased assessments.

By employing this comparative methodology, we highlight not only the technical strengths and limitations of each generative model but also their stylistic tendencies and prompt adherence, offering a holistic understanding of generative AI performance in artistic replication.

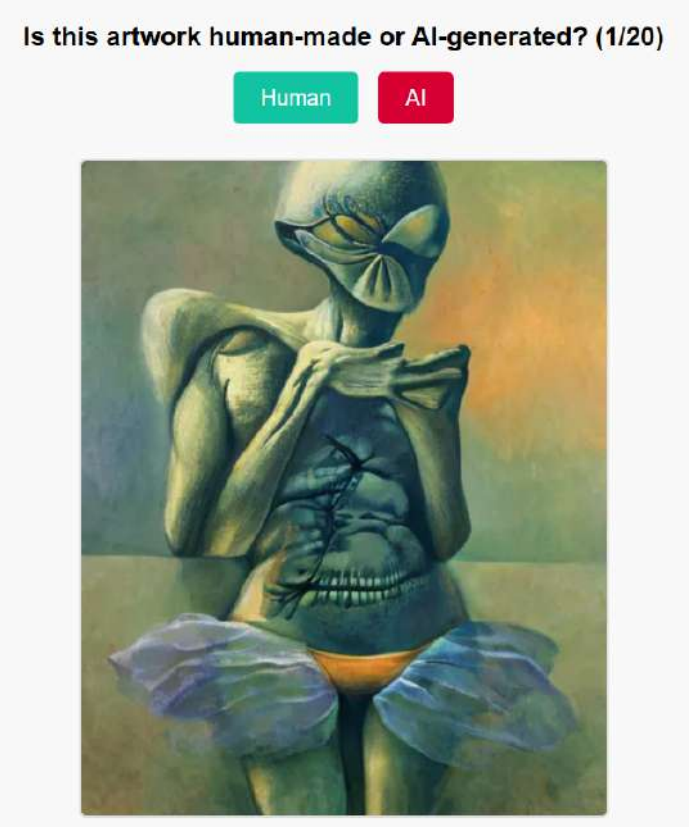


Figure 2.15: Example of a question posed in the survey.

Chapter 3

Results

3.1 Overview

This chapter presents the findings of our comparative evaluation of generative AI models for artistic style transfer. The key focus is on assessing the fidelity of AI-generated artworks, adherence to prompts, and public perception of authenticity. The findings are structured into quantitative metrics, survey results, and qualitative observations.

3.2 Evaluation of Model Performance

We evaluated AI-generated images based on the following criteria:

- **Style Accuracy**: Faithfulness of generated images to predefined artistic movements.
- **Structural Integrity**: Preservation of proper perspective, proportion, and balance.
- **Prompt Adherence**: The extent to which the generated image aligns with the provided textual prompt.

.

The ranking of generative models by their average classification score provides an overview of their relative performance. As depicted in Figure 3.1, ChatGPT4.0, Leonardo Phoenix, and Midjourney achieved the highest classification scores, suggesting that these models produce images that closely align with human artistic expectations. These models demonstrated superior adherence to stylistic prompts, rendering artworks with a balance of structure, depth, and texture. Conversely, models such as Auto-Aesthetics V1 and Stable Diffusion 1.5 exhibited the lowest classification scores, indicating that their outputs were more frequently identified as AI-generated. This suggests that their generated images may lack subtle stylistic details or introduce anomalies that distinguish them from human-created works.

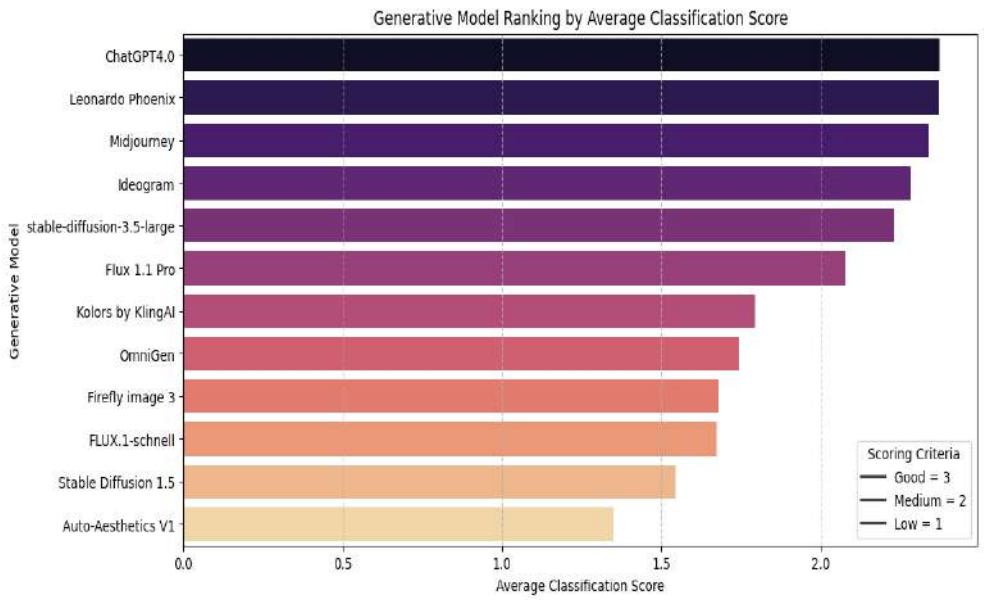


Figure 3.1: Generative model ranking by average classification score.

Further analysis of model performance is shown in Figure 3.3, where classification distributions for each generative model are examined. The figure highlights variations in the proportion of images classified as ”Good,” ”Medium,” or ”Low” quality. Midjourney, Leonardo Phoenix, and ChatGPT4.0 demonstrated a higher proportion of ”Good” classifications, reinforcing their ability

to produce convincing artistic outputs. On the other hand, models like Auto-Aesthetics V1 and FLUX.1-schnell exhibited a higher percentage of ”Low” classifications, reflecting limitations in their ability to convincingly mimic traditional artistic styles.

(a) Italian Renaissance (b) Rococo (c) Baroque

(d) Italian Renaissance (e) Rococo (f) Baroque

Figure 3.2: The first row displays artworks generated by ChatGPT-4.0, while the second row shows artworks generated by Auto-Aesthetics V1, representing the best and worst generative models based on the ranking by average classification score.

A broader perspective on classification distributions is provided in Figure 3.4, which aggregates classification scores across all models. The results indicate that a substantial percentage of images rendered fell in the ’Medium’ category for classification, suggesting that although AI-generated art tends to exhibit a good level of stylistic fidelity, there exists a substantial difference between high-scoring and low-scoring models. The high rate of ’Low’ classification scores also indicates inconsistencies in some generation outputs, suggesting a shortcoming in model performance in fine details in art, anatomical correctness, and compositional harmony.

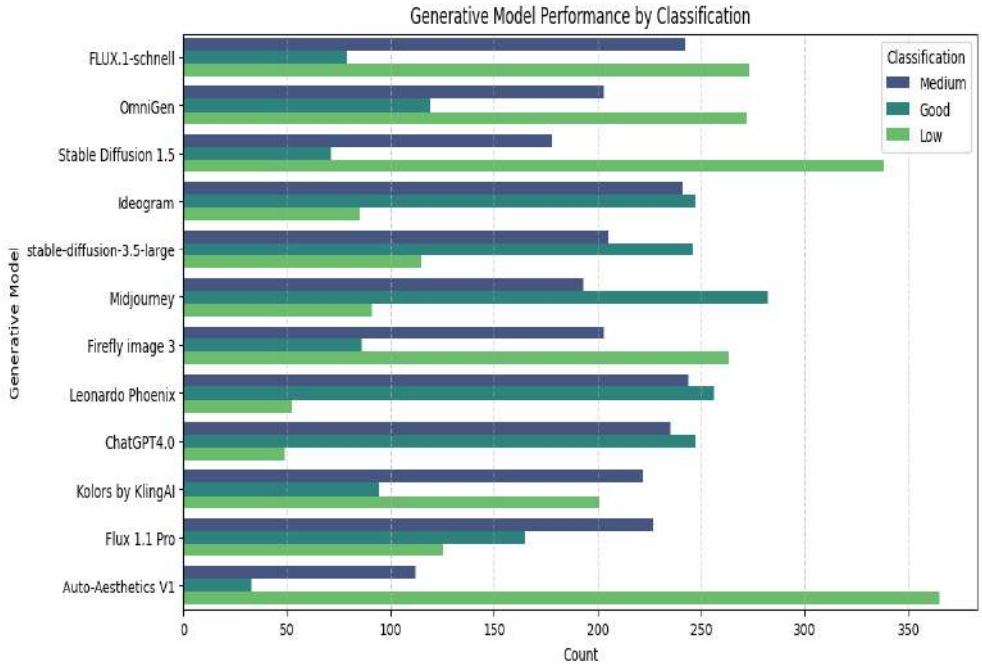


Figure 3.3: Generative model performance by classification category.

The combined examination of these statistics indicates differences in success rates in different generation models. While some models have a consistent record of generating high-quality outputs with high levels of adherence to conventional styles, others struggle with prompt fidelity, realism, and similarity in style. These findings suggest a scope for future developments in generation AI, particularly in those that require more emphasis on fine artwork details and situational awareness.

3.3 Survey Results

A public perception survey was conducted to assess how convincingly AI-generated images replicated human-created artwork. Participants were asked to classify images.

The analysis of misclassification rates reveals that certain prompts led to AI-generated images being frequently mistaken for real artwork. The highest

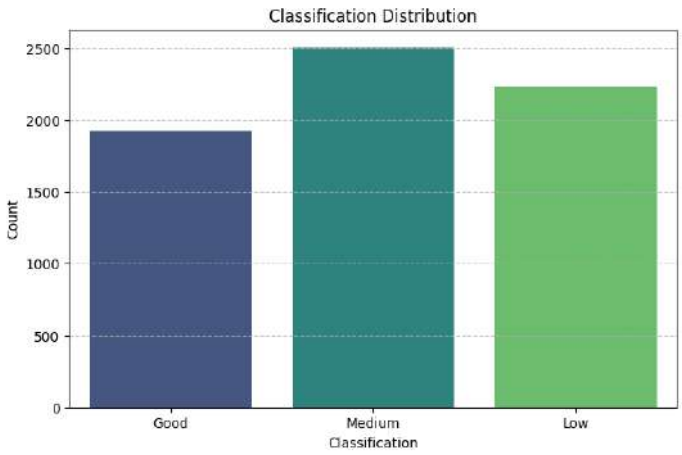


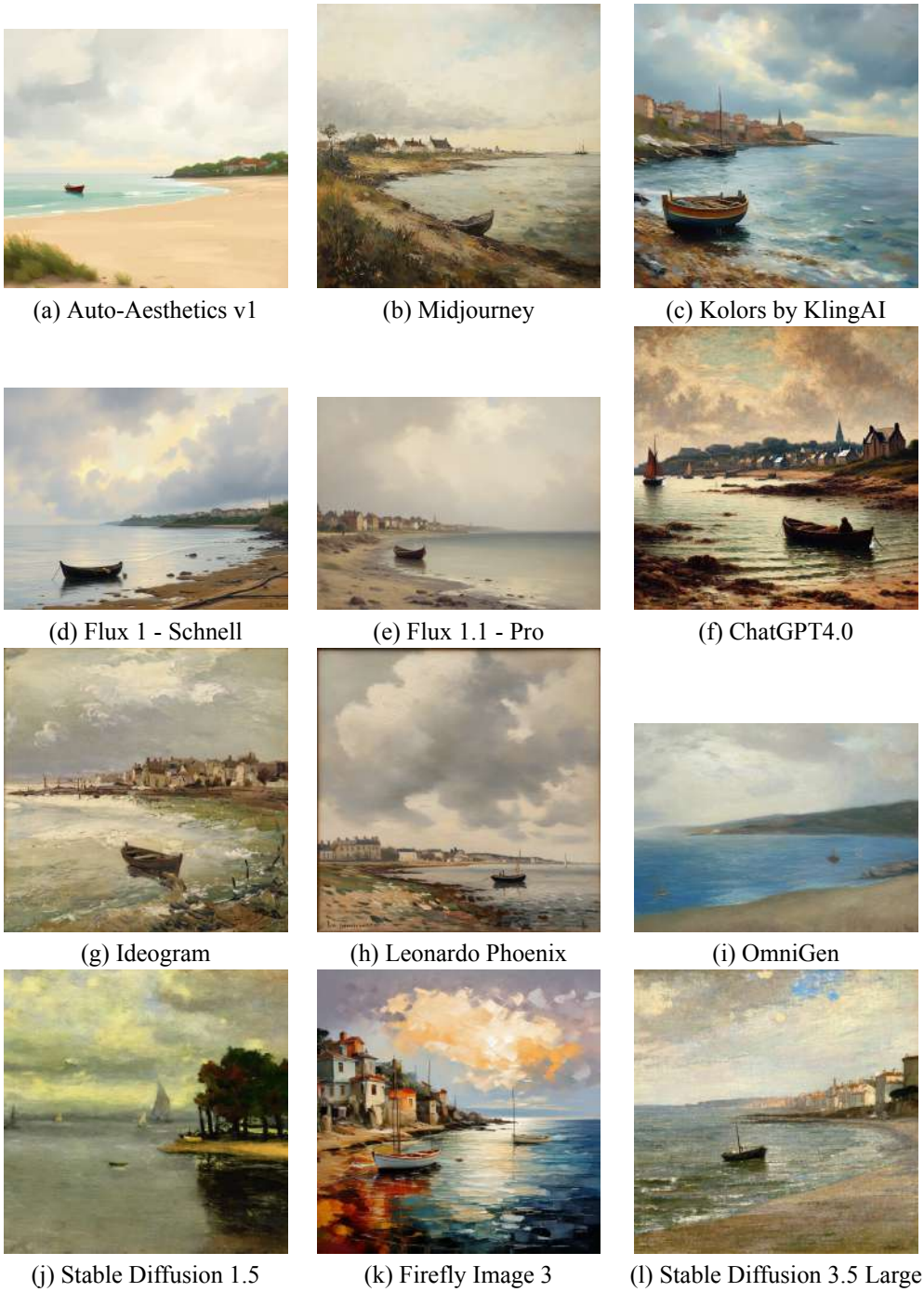
Figure 3.4: Overall classification distribution of AI-generated artworks.

misclassification rates were observed for prompts 23 and 13, where participants failed to correctly identify AI-generated images in more than 60% of instances. This suggests that specific textual prompts yield outputs that closely align with human artistic tendencies, possibly due to their emphasis on impressionistic or abstract elements, which AI models seem to replicate with notable accuracy.

Prompt 23: ”Generate a detailed coastal landscape painting in the Impressionist style movement of the 19th century. The painting depicts a coastal scene with a small boat near the water’s edge and distant buildings on the shore. The skyline is dominated by a cloudy sky, suggesting an overcast day. The colors are muted and earthy, conveying a serene yet slightly melancholic atmosphere. The technique used appears to be Impressionism, characterized by short, visible brushstrokes that capture the essence of the scene rather than detailed realism. The focus is on the play of light and color, creating a sense of movement and the transient nature of the landscape.”

As regards prompt 23, the effectiveness of this prompt in generating a believable Impressionist-style painting lies in its deliberate design of visual, technical, and emotive elements in accord with 19th-century Impressionist

technique. By specifying that it must use 'short, visible brushstrokes' and stress the 'play of light and color,' the prompt tells AI to replicate the painterly character typical of such figures as Claude Monet and Camille Pissarro. Artworks generated with this prompt are presented in Figure 3.5.



(a) Auto-Aesthetics v1 (b) Midjourney (c) Kolors by KlingAI

(d) Flux 1 - Schnell (e) Flux 1.1 - Pro (f) ChatGPT4.0

(g) Ideogram (h) Leonardo Phoenix (i) OmniGen

(j) Stable Diffusion 1.5 (k) Firefly Image 3 (l) Stable Diffusion 3.5 Large

Figure 3.5: Artworks generated using prompt number 23.

Furthermore, the subject matter, a coastal scene with a small boat at the water's edge and a distant set of houses along the shore, is thematically in accord with paintings from the period, as Impressionists tended to depict marine scenes to capture changing lighting and atmospheric effects. The use of a cloudy, overcast sky further reinforces the Impressionist preference for capturing transient atmospheric states and a sense of ephemerality that was a hallmark of the school.

The use of 'muted and earthy colors' also discourages AI from using highly saturated colors, ensuring that the resulting picture has a soft, blended tonality typical of 19th-century oil paintings. By using specific instructions in conjunction with open-ended language, the prompt permits a natural-looking arrangement without relying on contemporary digital exactness while preserving the spontaneity and movement integral to Impressionist paintings.

Through a combination of deliberate design and subtle suggestion, this prompt optimizes AI's ability to create an image that is not only visually accurate but also stylistically and historically authentic.

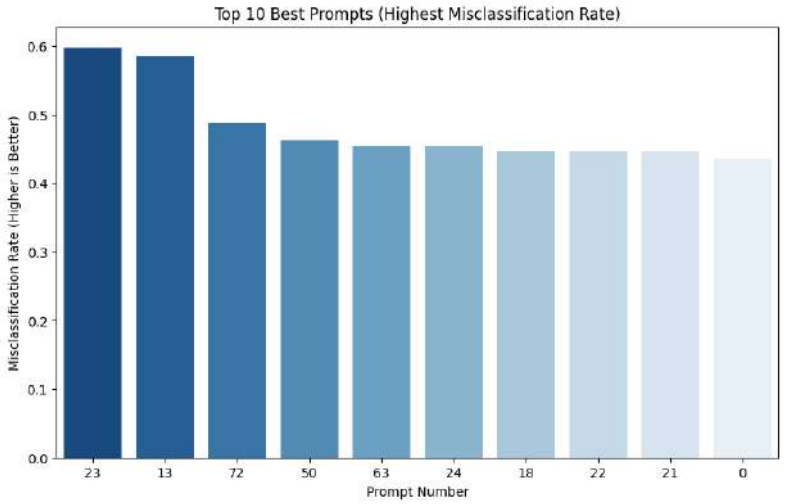


Figure 3.6: Top 10 prompts with the highest misclassification rates.

A deeper understanding of the user classification performance is provided

through the confusion matrix, which highlights the instances of correct and incorrect classifications. The data indicates that real images were correctly identified in the majority of cases, with 4726 correctly classified as real. However, a substantial portion, 1889 real images, were misclassified as AI-generated. Conversely, 4978 AI-generated images were correctly identified, while 1215 were mistakenly assumed to be human-created. These findings suggest a perceptual bias among participants, wherein some real images were considered AI-generated due to hyperrealistic features, while certain AI-generated images successfully deceived viewers by adhering to stylistic conventions that closely resemble human artistry.

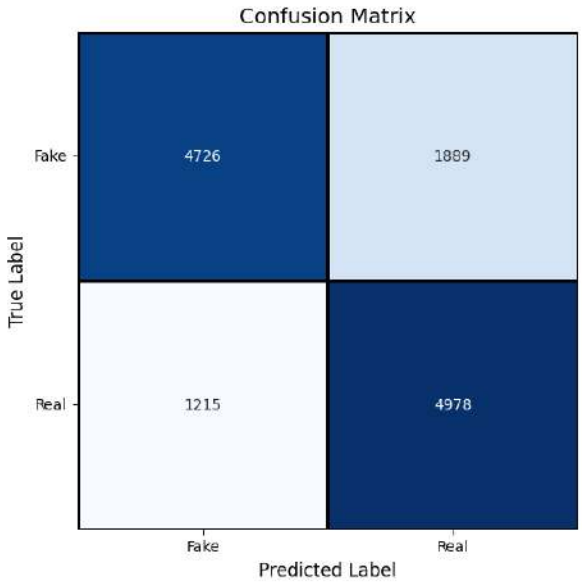


Figure 3.7: Confusion matrix showing classification performance.

Further examination of user responses against ground truth labels underscores the complexity of distinguishing AI-generated works from human-made ones. The classification of fake images reveals that, although many were correctly identified as AI-generated, a significant number were still perceived as authentic human artwork. Similarly, the misclassifications of real images indicates that viewers occasionally assumed genuine artistic pieces to be AI-generated, reinforcing the notion that modern generative models are capable

of producing outputs that challenge traditional distinctions between artificial and human creativity.

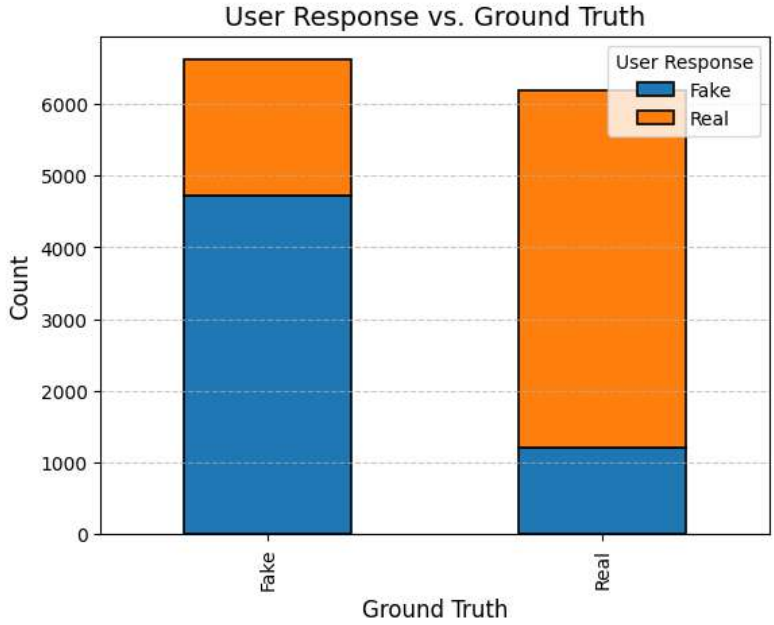


Figure 3.8: User response comparison against ground truth labels.

A comparative analysis of misclassification rates across different historical periods further highlights the strengths and limitations of AI-generated art. The results indicate that AI-generated images that mimicc 20th-century artistic movements, including surrealism and abstraction, exhibited the highestmisclassification rate. This suggests that contemporary styles, which often embrace abstract and experimental forms, are more readily synthesized by AI in a manner that aligns with human expectations. In contrast, styles that require meticulous detail, such as those from the Baroque and Renaissance periods, were more accurately classified, as AI models often struggle with fine-grained precision and depth perception. The 18th century category had the lowest misclassification rate, reinforcing the notion that AI models face challenges when replicating more intricate brushwork and historical accuracy.

Model-wise misclassification rates also reflect differences in generative capability. Ideogram and Midjourney had the highest misclassification rates,

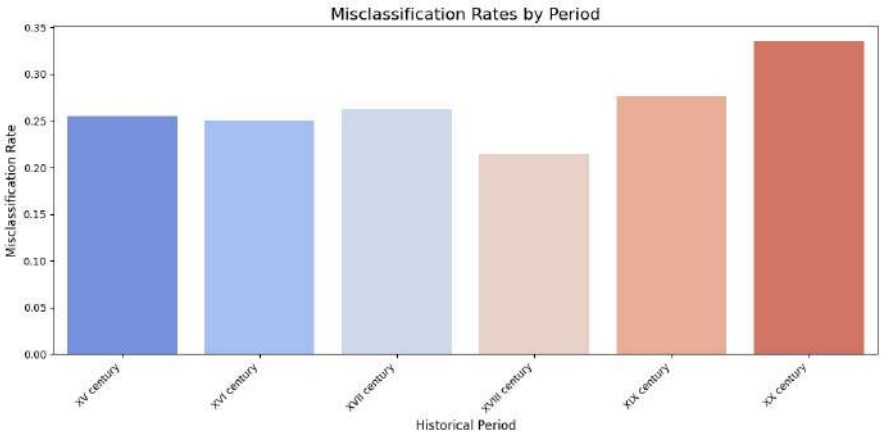


Figure 3.9: Misclassification rates across historical periods.

meaning their outputs were frequently mistaken for human-created art. In contrast, models like Auto-Aesthetics V1 showed lower misclassification rates, suggesting that their outputs lacked the subtle details needed to convincingly mimic traditional artistic styles. These findings highlight the varying strengths of different generative models and their ability to create images that blur the distinction between human and AI-generated art.

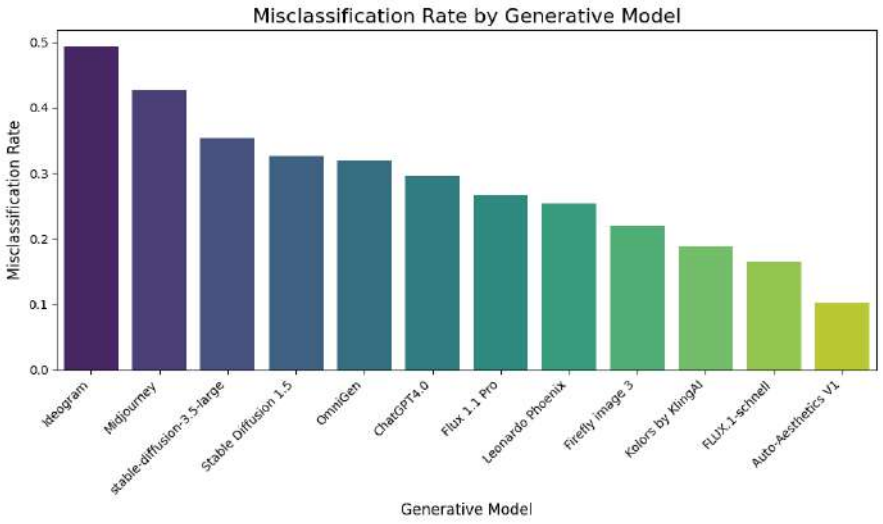


Figure 3.10: Misclassification rate per generative model.

The analysis of the least deceptive prompts, or those with the lowest misclassifications rate, further elucidates the limitations of AI-generated art. Prompts 26 and 42 consistently produced images that participants accurately identified

as AI-generated, suggesting that specific stylistic choices expose the underlying constraints of AI models. These prompts may involve intricate structural compositions, challenging perspective work, or stylistic nuances that remain difficult for AI to replicate convincingly.

This pattern is particularly evident in Prompt 26, which attempted to generate a 20th-century Surrealist painting but ultimately failed to deceive participants. The shortcomings of this prompt highlight key weaknesses in AI-generated art, particularly when it comes to capturing the psychological depth and unpredictability that define Surrealism. Unlike genuine Surrealist works, which rely on subconscious associations, dream logic, and paradoxical spatial relationships, the AI-generated output followed a structured, almost formulaic composition. The elements, a barren landscape, a meditative woman in a flowing gown, a transparent orb reflecting a historical scene, and a distant volcano, felt more like a checklist of surrealist motifs rather than an organically conceived composition. This predictability, combined with an overemphasis on ”precise, realistic detail,” resulted in an image that appeared too polished and logically arranged, lacking the painterly distortions, irrational juxtapositions, and enigmatic atmosphere characteristic of true Surrealist works. Artworks generated with this prompt are presented in Figure 3.12.

Prompt 26: ”Generate a detailed surrealist painting in the style of the 20th century Surrealism. The painting portrays a woman in a long, flowing gown standing on a cracked, barren landscape. She is looking upward with her hands in a prayer or meditative pose. To her left hangs a large, transparent orb reflecting a historical or symbolic scene. The background features a cloudy sky and a distant volcano emitting smoke and fire. The technique used appears to be surrealism, characterized by its dream-like, fantastical elements and precise, realistic detail. The composition blends realistic portrayals with abstract concepts, creating a sense of mystery and contemplation.”

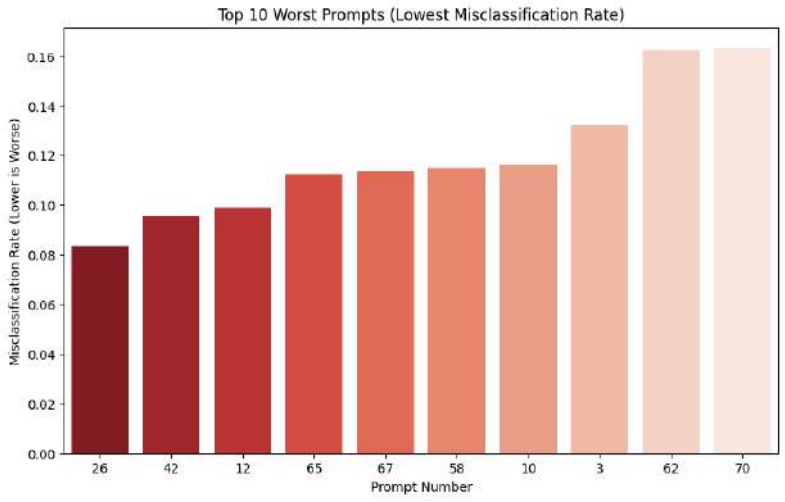


Figure 3.11: Top 10 prompts with the lowest misclassification rates.

Overall, the survey results illustrate the evolving capability of generative AI in artistic applications. While certain prompts and models achieve highly deceptive outputs, others reveal the persistent challenges AI faces in adhering to historical and compositional fidelity. The findings indicate that impressionistic and abstract styles are among the most convincingly replicated, whereas highly detailed classical styles remain a challenge. Moreover, the perceptual biases exhibited by participants, particularly the tendency to misclassify hyperrealistic real images as AI-generated, suggest that the integration of AI in artistic workflows is shifting traditional notions of artistic authenticity. Future research may explore adaptive AI training strategies to enhance historical accuracy and mitigate biases in generated images, further advancing the field of AI-assisted artistry.

3.4 Summary

The evaluation of generative AI models for artistic style transfer reveals significant variations in performance across different models and historical periods. The analysis of classification scores, misclassifications rate, and user perception data demonstrates that certain AI-generated artworks closely align with

human-created pieces, while others exhibit clear limitations in compositional coherence and stylistic fidelity.

The ranking of generative models highlights that Midjourney, Leonardo Phoenix, and ChatGPT4.0 consistently produce images that align well with artistic expectations, achieving high classification scores and a greater proportion of ”Good” ratings. These models effectively capture stylistic nuances, rendering convincing textures, lighting effects, and brushstroke details. In contrast, models such as Auto-Aesthetics V1 and Stable Diffusion 1.5 exhibit lower classification scores, suggesting a need for improvement in fine-grained artistic elements and structural integrity.

The misclassification rates provide further insights into the difficulty of distinguishing AI-generated artworks from human-made pieces. The findings suggest that impressionism, surrealism, and abstract styles are more readily synthesized by AI, often leading to higher misclassification rates. Conversely, styles with intricate details, such as those from the Baroque and Renaissance periods, pose greater challenges for AI models, resulting in lower misclassification rates and a higher likelihood of correct identification by participants.

Survey results reinforce these findings, showing that participants frequently misidentified certain AI-generated artworks as human-created, particularly those produced by models with high classification scores. This suggests that advances in generative AI are gradually blurring the line between artificial and human creativity, raising important questions about the role of AI in artistic production. At the same time, the survey highlights the persistence of perceptual biases, with some real images being mistakenly classified as AI-generated, indicating that hyperrealism and stylized elements influence human judgment.

The overall distribution of classification scores suggests that a significant portion of AI-generated artworks fall within the ”Medium” category, implying that while AI can replicate artistic elements with some degree of accuracy, inconsistencies and artifacts remain prevalent. Common distortions include

anatomical inaccuracies, hyperrealistic details, and unintended stylistic deviations, which impact the overall authenticity of the generated images.

In conclusion, while generative AI has made significant strides in artistic style transfer, there remains substantial room for improvement. The findings suggest that future advancements should focus on enhancing prompt adherence, refining structural integrity, and reducing stylistic inconsistencies. Additionally, the growing ability of AI to produce deceptive artistic outputs calls for further exploration into ethical considerations, including transparency in AI-assisted artistic creation and its impact on traditional artistic practices. As AI-generated art continues to evolve, understanding its strengths and limitations will be crucial for its responsible integration into the creative landscape.

(a) Auto-Aesthetics v1

(b) Midjourney

(c) Kolors by KlingAI

(d) Flux 1 - Schnell

(e) Flux 1.1 - Pro

(f) ChatGPT4.0

(g) Ideogram

(h) Leonardo Phoenix

(i) OmniGen

(j) Stable Diffusion 1.5

(k) Firefly Image 3

(l) Stable Diffusion 3.5 Large

Figure 3.12: Artworks generated using prompt number 26.

Chapter 4

Conclusions

The exploration of AI-generated art and its perception among users provides valuable insights into the evolving landscape of generative models in artistic creation. This thesis examined the effectiveness of various generative models, their ability to mimic traditional artistic styles, and the challenges associated with distinguishing AI-generated artworks from human-created pieces. Through a comprehensive evaluation incorporating user surveys, classification analysis, and misclassification rates, several key findings emerged that contribute to a deeper understanding of AI-assisted artistry.

4.1 Key Findings and Contributions

One of the study's key findings is the significant variation in performance across generative models. Models such as Midjourney, Leonardo Phoenix, and ChatGPT4.0 consistently produced high-quality outputs that closely resembled human-created art, achieving high classification scores and misclassification rates [28, 25, 33]. These models excelled in capturing stylistic nuances, realistic textures, and compositional depth. In contrast, models like Auto-Aesthetics V1 and Stable Diffusion 1.5 performed less effectively, as their outputs were more easily identified as AI-generated due to visible artifacts, inconsistencies, or a lack of fine details [31, 40].

A historical period analysis revealed that AI models are particularly proficient at synthesizing modern and abstract styles, leading to higher misclassification rates for these artistic movements. On the other hand, classical and highly detailed styles, such as those from the Renaissance and Baroque periods, were more accurately classified [26]. These findings indicate that while AI can effectively replicate broader stylistic elements, challenges remain in capturing intricate brushwork, perspective accuracy, and historically informed artistic techniques.

User perception studies further demonstrated the increasing sophistication of AI-generated art. Survey results showed that participants frequently misidentified AI-generated images as real, particularly when created by top-performing models. At the same time, some real images were mistaken for AI-generated artwork, revealing potential biases in human perception, where hyper-realistic details or stylized elements can lead to assumptions about artificiality. These findings emphasize that as generative models continue to advance, their outputs will increasingly challenge traditional ideas of authenticity and artistic authorship [6].

Another key aspect explored was the role of textual prompts in shaping AI-generated outputs. The analysis of the best and worst performing prompts highlighted how specific phrasing can significantly influence image quality and misclassification rates. Some prompts guided the AI to create outputs that closely adhered to artistic conventions, making them nearly indistinguishable from human-created artwork. Others exposed AI's limitations, leading to structural inconsistencies and lower classification scores. This suggests that optimizing prompt engineering can be a crucial factor in improving AI-generated artistic outputs [37, 34, 41].

The overall classification distribution analysis revealed that while many AI-generated artworks fall into the 'Medium' category, indicating partial success in mimicking human artistry, there remains a significant number of images that are either convincingly real or clearly artificial. This distribution

highlights the ongoing need for improvements in AI training methodologies to enhance realism, contextual accuracy, and stylistic coherence [44].

4.2 Insights on Model Structure and Limitations

The architectural structure of generative models plays a crucial role in determining their effectiveness in artistic replication. Most state-of-the-art models, such as Midjourney and ChatGPT4.0, leverage transformer-based architectures that enable a more nuanced understanding of artistic styles and composition [29]. These models incorporate large-scale datasets of historical and contemporary artworks, allowing them to generate outputs with high degrees of realism. However, they also face limitations in areas such as anatomical accuracy, fine detail rendering, and maintaining stylistic consistency across variations of the same prompt [53].

Diffusion-based models, including Stable Diffusion 1.5 and Stable-Diffusion-3.5-large, employ iterative refinement processes to create images progressively. While effective in producing high-resolution outputs, these models can sometimes introduce noise artifacts or struggle with maintaining logical consistency in complex compositions [43]. Furthermore, latent space manipulation techniques in generative models continue to evolve, but challenges remain in achieving greater control over fine artistic elements, brushwork emulation, and depth perception [23].

4.3 Future Directions

Despite the remarkable progress in generative AI, this research also raises important ethical and philosophical considerations. The increasing ability of AI to produce realistic artworks poses questions about originality, artistic intent, and the role of human creativity in an era of machine-assisted generation [10]. The findings suggest that while AI-generated art can complement traditional

artistic practices, transparency in AI-assisted creation and ethical considerations surrounding its usage must be carefully addressed.

Looking ahead, future research should explore adaptive AI training strategies that incorporate human feedback to refine generative outputs further. Enhancing AI's ability to replicate intricate artistic details, improving the contextual relevance of generated images, and mitigating perceptual biases will be essential in advancing AI's role in the creative industries [2]. Additionally, the intersection of AI-generated art and human artistic collaboration presents an exciting avenue for research, where AI can serve as a tool for augmenting human creativity rather than replacing it.

A promising direction for improvement is the integration of hybrid models that combine diffusion processes with transformer-based networks, potentially leading to more contextually aware and structurally coherent outputs [45]. Additionally, further research into fine-tuned datasets specific to historical art styles could improve AI's ability to replicate period-specific techniques with greater authenticity [27].

4.4 Final Remarks

In conclusion, this thesis has demonstrated that AI-generated art is rapidly advancing in its ability to emulate traditional artistic styles. While challenges remain, particularly in achieving fine-grained artistic authenticity, the findings indicate that AI is an increasingly powerful tool in artistic production. As AI continues to evolve, understanding its impact, refining its capabilities, and ensuring ethical integration into creative domains will be crucial for shaping the future of AI-assisted artistry. By addressing the identified challenges and focusing on ethical and technical improvements, generative AI can transition from an experimental novelty to a widely accepted and responsible tool for artistic expression. Further insights and a more profound analysis can be

found in [3], which critically assesses the style replication capabilities of contemporary generative models and presents the ”AI-pastiche” dataset, offering a comprehensive evaluation of AI-generated artistic imitations.

Acknowledgements

I want to sincerely thank Professor Andrea Asperti and my colleagues Tiberio, Fabio, and Franky for giving me the opportunity to be part of this group. Without the guidance of the professor and the support of my colleagues, all the effort put into this thesis and the article [3] would not have led to these results.

I also want to thank my family, who have always given me their unwavering support and acted as my anchor, even when I felt lost at sea. They are my most precious treasure. I am also deeply grateful to my friends and all the people who have played a significant role in my life. I will follow up with a list, though a truly complete one would be so long that printing it would probably bankrupt me.

Muda, I love all of you with all my heart. The Mugs (FreeViolino, Mugs, IALTRI, insert new name here), thank you for introducing me to a new passion that has added a whole new flavor to my life. Via Cairoli, for letting me bother you and welcoming me as one of your own, even though I live 9.5 km away. Maggio Esponenziale, for being the most insane, awesome, and creative people I have had the chance to stumble upon, and for also getting me addicted to Magic. Matteo for embarking on this journey with me and never abandoning it, no matter how dark it got outside.

I have explicitly chosen not to name anyone else individually, not because you are not important, but because I would love to thank each of you personally. If I am reading this, you are probably standing in front of me and I sincerely hope that I do not get too emotional.

Bibliography

[1] Adobe. *Adobe advances creative ideation with the new Firefly Image 3 Model*. Accessed: 2024-11-18. Apr. 2024. URL: https://blog.adobe.com/en/publish/2024/04/23/adobe-advances-creative-ideation-with-new-firefly-image-3-model.

[2] Nantheera Anantrasirichai, Fan Zhang, and David Bull. "Artificial Intelligence in Creative Industries: Advances Prior to 2025". In: *CoRR* abs/2501.02725 (2025). DOI: 10.48550/ARXIV.2501.02725. arXiv: 2501.02725. URL: https://doi.org/10.48550/arXiv.2501.02725.

[3] Andrea Asperti et al. *A Critical Assessment of Modern Generative Models' Ability to Replicate Artistic Styles*. 2025. arXiv: 2502.15856 [cs.CV]. URL: https://arxiv.org/abs/2502.15856.

[4] Emily M. Bender et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" In: *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event / Toronto, Canada, March 3-10, 2021*. 2021, pp. 610–623. DOI: 10.1145/3442188.3445922. URL: https://doi.org/10.1145/3442188.3445922.

[5] Tim Brooks, Aleksander Holynski, and Alexei A. Efros. "InstructPix2Pix: Learning to Follow Image Editing Instructions". In: *CoRR* abs/2211.09800 (2022). DOI: 10.48550/ARXIV.2211.09800. arXiv: 2211.09800. URL: https://doi.org/10.48550/arXiv.2211.09800.

[6] Emily Brown and Robert Davis. “The Influence Of AI Text-To-Image Technologies On Modern Art: A Survey”. In: *IOSR Journal of Computer Engineering* 26.5 (2024), pp. 1–6. URL: https://www.iosrjournals.org/iosr-jce/papers/Vol26-issue5/Ser-5/A2605050106.pdf.

[7] Anthony Chemero. “LLMs differ from human cognition because they are not embodied”. In: *Nature Human Behaviour* 7.11 (2023), pp. 1828–1829.

[8] Jiwoo Chung, Sangeek Hyun, and Jae-Pil Heo. “Style Injection in Diffusion: A Training-Free Approach for Adapting Large-Scale Diffusion Models for Style Transfer”. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*. 2024, pp. 8795–8805. DOI: 10.1109/CVPR52733.2024.00840. URL: https://doi.org/10.1109/CVPR52733.2024.00840.

[9] Prafulla Dhariwal and Alexander Quinn Nichol. “Diffusion Models Beat GANs on Image Synthesis”. In: *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*. Ed. by Marc’Aurelio Ranzato et al. 2021, pp. 8780–8794.

[10] Jane Doe. “Ethically dubious or a creative gift? How artists are grappling with AI in their work”. In: *The Guardian* (2024). URL: https://www.theguardian.com/artanddesign/article/2024/aug/30/xanthe-dobbie-futuer-sex-love-sounds-ai-video-celebrity-clones.

[11] Patrick Esser et al. “Scaling Rectified Flow Transformers for High-Resolution Image Synthesis”. In: *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. 2024. URL: https://openreview.net/forum?id=FPnUhsQJ5B.

[12] Patrick Esser et al. “Scaling Rectified Flow Transformers for High-Resolution Image Synthesis”. In: *CoRR* abs/2403.03206 (2024). DOI: 10.48550/ARXIV.2403.03206. arXiv: 2403.03206. URL: https://doi.org/10.48550/arXiv.2403.03206.

[13] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. “A Neural Algorithm of Artistic Style”. In: *CoRR* abs/1508.06576 (2015). arXiv: 1508.06576. URL: http://arxiv.org/abs/1508.06576.

[14] Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. “Image Style Transfer Using Convolutional Neural Networks”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*. IEEE Computer Society, 2016, pp. 2414–2423. DOI: 10.1109/CVPR.2016.265. URL: https://doi.org/10.1109/CVPR.2016.265.

[15] Ian J. Goodfellow et al. “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*. 2014, pp. 2672–2680.

[16] Alex Henry et al. “Query-Key Normalization for Transformers”. In: *Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020*. Ed. by Trevor Cohn, Yulan He, and Yang Liu. Vol. EMNLP 2020. Findings of ACL. Association for Computational Linguistics, 2020, pp. 4246–4253. DOI: 10.18653/V1/2020.FINDINGS-EMNLP.379. URL: https://doi.org/10.18653/v1/2020.findings-emnlp.379.

[17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising Diffusion Probabilistic Models”. In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*. Ed. by Hugo Larochelle et al. 2020. URL: https://proceedings.neurips.

cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

[18] Jonathan Ho et al. “Video Diffusion Models”. In: *NeurIPS*. 2022. URL: http://papers.nips.cc/paper%5C_files/paper/2022/hash/39235c56aef13fb05a6adc95eb9d8d66-Abstract-Conference.html.

[19] Phillip Isola et al. “Image-to-Image Translation with Conditional Adversarial Networks”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*. IEEE Computer Society, 2017, pp. 5967–5976. DOI: 10.1109/CVPR.2017.632. URL: https://doi.org/10.1109/CVPR.2017.632.

[20] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. “Perceptual Losses for Real-Time Style Transfer and Super-Resolution”. In: *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*. Ed. by Bastian Leibe et al. Vol. 9906. Lecture Notes in Computer Science. Springer, 2016, pp. 694–711. DOI: 10.1007/978-3-319-46475-6_43. URL: https://doi.org/10.1007/978-3-319-46475-6%5C_43.

[21] Subbarao Kambhampati. “Can Large Language Models Reason and Plan?” In: *CoRR* abs/2403.04121 (2024). DOI: 10.48550/ARXIV.2403.04121. arXiv: 2403.04121. URL: https://doi.org/10.48550/arXiv.2403.04121.

[22] Tero Karras et al. “Analyzing and Improving the Image Quality of StyleGAN”. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. Computer Vision Foundation / IEEE, 2020, pp. 8107–8116. DOI: 10.1109/CVPR42600.2020.00813. URL: https://openaccess.thecvf.com/content%5C_CVPR%5C_2020/html/Karras%5C_

Analyzing%5C_and%5C_Improving%5C_the%5C_Image%5C_Quality%5C_of%5C_StyleGAN%5C_CVPR%5C_2020%5C_paper.html.

[23] Mingi Kwon, Jaeseok Jeong, and Youngjung Uh. “Diffusion Models already have a Semantic Latent Space”. In: *CoRR* abs/2210.10960 (2022). DOI: 10.48550/arXiv.2210.10960. arXiv: 2210.10960. URL: https://doi.org/10.48550/arXiv.2210.10960.

[24] Black Forest Labs. *Announcing Black Forest Labs*. Accessed: 2025-1-11. 2024. URL: https://blackforestlabs.ai/announcing-black-forest-labs/.

[25] Leonardo.Ai. *Introducing Phoenix by Leonardo.Ai*. Accessed: 2024-11-18. Oct. 2024. URL: https://leonardo.ai/phoenix/.

[26] Peiyuan Liao et al. “The ArtBench Dataset: Benchmarking Generative Models with Artworks”. In: *CoRR* abs/2206.11404 (2022). DOI: 10.48550/ARXIV.2206.11404. arXiv: 2206.11404. URL: https://doi.org/10.48550/arXiv.2206.11404.

[27] Hui Mao, James She, and Ming Cheung. “Visual Arts Search on Mobile Devices”. In: *ACM Trans. Multim. Comput. Commun. Appl.* 15.2s (2019), 60:1–60:23. DOI: 10.1145/3326336. URL: https://doi.org/10.1145/3326336.

[28] Midjourney. *Version 6.1*. Accessed: 2024-11-18. July 2024. URL: https://updates.midjourney.com/version-6-1/.

[29] Sarah Miller and David Thompson. “The two models fueling generative AI products: Transformers and diffusion models”. In: *GPTechBlog* (2023). URL: https://www.gptechblog.com/generative-ai-models-transformers-diffusion-models/.

[30] Mehdi Mirza and Simon Osindero. “Conditional Generative Adversarial Nets”. In: *CoRR* abs/1411.1784 (2014). arXiv: 1411.1784. URL: http://arxiv.org/abs/1411.1784.

[31] Neural.love. *AI Art Revolution: Introducing Auto-Aesthetics for Personalized Gen AI experience*. Accessed: 2024-08-14. Aug. 2024. URL: https://neural.love/blog/auto-aesthetics-v1-ai-art-revolution.

[32] Alexander Quinn Nichol and Prafulla Dhariwal. “Improved denoising diffusion probabilistic models”. In: *International Conference on Machine Learning*. PMLR. 2021, pp. 8162–8171.

[33] OpenAI. “GPT-4 Technical Report”. In: *CoRR* abs/2303.08774 (2023). DOI: 10.48550/ARXIV.2303.08774. arXiv: 2303.08774. URL: https://doi.org/10.48550/arXiv.2303.08774.

[34] Jonas Oppenlaender. “The Creativity of Text-to-Image Generation”. In: *25th International Academic Mindtrek conference, Academic Mindtrek 2022, Tampere, Finland, November 16-18, 2022*. ACM, 2022, pp. 192–202. DOI: 10.1145/3569219.3569352. URL: https://doi.org/10.1145/3569219.3569352.

[35] Gaurav Parmar et al. “One-Step Image Translation with Text-to-Image Models”. In: *CoRR* abs/2403.12036 (2024). DOI: 10.48550/ARXIV.2403.12036. arXiv: 2403.12036. URL: https://doi.org/10.48550/arXiv.2403.12036.

[36] Wen Hao Png, Yichiet Aun, and Ming-Lee Gan. “FeaST: Feature-guided Style Transfer for high-fidelity art synthesis”. In: *Comput. Graph.* 122 (2024), p. 103975. DOI: 10.1016/J.CAG.2024.103975. URL: https://doi.org/10.1016/j.cag.2024.103975.

[37] Han Qiao, Vivian Liu, and Lydia B. Chilton. “Initial Images: Using Image Prompts to Improve Subject Representation in Multimodal AI Generated Art”. In: *C&C ’22: Creativity and Cognition, Venice, Italy,*

June 20 - 23, 2022. ACM, 2022, pp. 15–28. URL: https://doi.org/10.1145/3527927.3532792.

[38] Alec Radford et al. “Learning Transferable Visual Models From Natural Language Supervision”. In: *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, 2021, pp. 8748–8763. URL: http://proceedings.mlr.press/v139/radford21a.html.

[39] Aditya Ramesh et al. “Hierarchical Text-Conditional Image Generation with CLIP Latents”. In: *arXiv e-prints* (2022), arXiv–2204.

[40] Robin Rombach et al. “High-resolution image synthesis with latent diffusion models”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 10684–10695.

[41] Téo Sanchez. “Examining the Text-to-Image Community of Practice: Why and How do People Prompt Generative AIs?” In: *Creativity and Cognition, C&C 2023, Virtual Event, USA, June 19-21, 2023*. ACM, 2023, pp. 43–61. URL: https://doi.org/10.1145/3591196.3593051.

[42] Jiaming Song, Chenlin Meng, and Stefano Ermon. “Denoising Diffusion Implicit Models”. In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL: https://openreview.net/forum?id=St1giarCHLP.

[43] Bingyuan Wang, Qifeng Chen, and Zeyu Wang. “Diffusion-Based Visual Art Creation: A Survey and New Perspectives”. In: *CoRR* abs/2408.12128 (2024). DOI: 10.48550/ARXIV.2408.12128. arXiv: 2408.12128. URL: https://doi.org/10.48550/arXiv.2408.12128.

[44] Boheng Wang et al. “A study of the evaluation metrics for generative images containing combinational creativity”. In: *Artif. Intell. Eng. Des. Anal. Manuf.* 37 (2023). URL: https://doi.org/10.1017/s0890060423000069.

[45] Li Wang and Ming Zhao. “Diffusion Models in Generative AI: Principles, Applications, and Challenges”. In: *Preprints* (2025).

[46] Jason Wei et al. “Emergent Abilities of Large Language Models”. In: *Trans. Mach. Learn. Res.* 2022 (2022). URL: https://openreview.net/forum?id=yzkSU5zdwD.

[47] Shitao Xiao et al. “OmniGen: Unified Image Generation”. In: *CoRR* abs/2409.11340 (2024). DOI: 10.48550/ARXIV.2409.11340. arXiv: 2409.11340. URL: https://doi.org/10.48550/arXiv.2409.11340.

[48] Yifei Xu et al. “SGDM: An Adaptive Style-Guided Diffusion Model for Personalized Text to Image Generation”. In: *IEEE Trans. Multim.* 26 (2024), pp. 9804–9813. DOI: 10.1109/TMM.2024.3399075. URL: https://doi.org/10.1109/TMM.2024.3399075.

[49] Yuhui Xu et al. “QA-LoRA: Quantization-Aware Low-Rank Adaptation of Large Language Models”. In: *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL: https://openreview.net/forum?id=WvFoJccpo8.

[50] Shunyu Yao et al. “Tree of Thoughts: Deliberate Problem Solving with Large Language Models”. In: *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*. 2023. URL: http://papers.nips.cc/paper%5C_files/paper/2023/hash/271db9922b8d1f4dd7aaef84ed5ac703-Abstract-Conference.html.

[51] Lin Zhang et al. “FSIM: A Feature Similarity Index for Image Quality Assessment”. In: *IEEE Trans. Image Process.* 20.8 (2011), pp. 2378–2386. DOI: 10.1109/TIP.2011.2109730. URL: https://doi.org/10.1109/TIP.2011.2109730.

[52] Richard Zhang et al. “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric”. In: *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*. Computer Vision Foundation / IEEE Computer Society, 2018, pp. 586–595. DOI: 10.1109/CVPR.2018.00068. URL: http://openaccess.thecvf.com/content%5C_cvpr%5C_2018/html/Zhang%5C_The%5C_Unreasonable%5C_Effectiveness%5C_CVPR%5C_2018%5C_paper.html.

[53] Wei Zhang and Li Chen. “The Limitations and Ethical Considerations of ChatGPT”. In: *Digital Intelligence* 6.1 (2023), pp. 201–218. URL: https://direct.mit.edu/dint/article/6/1/201/118839/The-Limitations-and-Ethical-Considerations-of.

[54] Zhanjie Zhang et al. “ArtBank: Artistic Style Transfer with Pre-trained Diffusion Model and Implicit Style Prompt Bank”. In: *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada*. 2024, pp. 7396–7404. DOI: 10.1609/AAAI.V38I7.28570. URL: https://doi.org/10.1609/aaai.v38i7.28570.

[55] Zhanjie Zhang et al. “Towards Highly Realistic Artistic Style Transfer via Stable Diffusion with Step-aware and Layer-aware Prompt”. In: *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI 2024, Jeju, South Korea, August 3-9, 2024*.

2024, pp. 7814–7822. URL: https://www.ijcai.org/proceedings/2024/865.

[56] Jun-Yan Zhu et al. “Unpaired image-to-image translation using cycle-consistent adversarial networks”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2223–2232.